Segmentation en lignes de documents anciens : applications aux documents arabes

Nazih Ouwayed

▶ **To cite this version:**

Nazih Ouwayed. Segmentation en lignes de documents anciens : applications aux documents arabes. Autre [cs.OH]. Université Nancy 2, 2010. Français. NNT : 2010NAN23001 . tel-01752983

HAL Id: tel-01752983
https://hal.univ-lorraine.fr/tel-01752983

Submitted on 29 Mar 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITÉ DE LORRAINE

AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : ddoc-theses-contact@univ-lorraine.fr

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4
Code de la Propriété Intellectuelle. articles L 335.2- L 335.10
http://www.cfcopies.com/V2/leg/leg_droi.php
http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm

Nancy-Université
Université Nancy 2

École doctorale IAEM Lorraine

UFR mathématiques et informatique

Département de formation doctorale en informatique

Segmentation en lignes de documents anciens : application aux documents arabes

THÈSE

présentée et soutenue publiquement le 11 Juin 2010

pour l'obtention du

Doctorat de l'université Nancy 2

(spécialité informatique)

par

Nazih OUWAYED

Composition du jury

Président :	Salvatore-Antoine Tabbone	Professeur, Université Nancy 2
Rapporteurs :	Christian Olivier	Professeur, Université de Poitiers
	Jean-Yves Ramel	Professeur, Université de Tours
Examinateurs :	Salvatore-Antoine Tabbone	Professeur, Université Nancy 2
	François Auger	MCHC-HDR, Université de Nantes
Directeur de thèse :	Abdel Belaïd	Professeur, Université Nancy 2

Laboratoire Lorrain de Recherche en Informatique et ses Applications — UMR 7503

Loria

Mis en page avec la classe thloria.

Nancy-Université
Université Nancy 2

École doctorale IAEM Lorraine

UFR mathématiques et informatique **Département de formation doctorale en informatique**

Segmentation en lignes de documents anciens : application aux documents arabes

THÈSE

présentée et soutenue publiquement le 11 Juin 2010

pour l'obtention du

Doctorat de l'université Nancy 2

(spécialité informatique)

par

Nazih OUWAYED

Composition du jury

Président :	Salvatore-Antoine Tabbone	Professeur, Université Nancy 2
Rapporteurs :	Christian Olivier	Professeur, Université de Poitiers
	Jean-Yves Ramel	Professeur, Université de Tours
Examinateurs :	Salvatore-Antoine Tabbone	Professeur, Université Nancy 2
	François Auger	MCHC-HDR, Université de Nantes
Directeur de thèse :	Abdel Belaïd	Professeur, Université Nancy 2

Laboratoire Lorrain de Recherche en Informatique et ses Applications — UMR 7503

Loria

Mis en page avec la classe thloria.

Résumé

L'indexation de documents numérisés manuscrits pose le problème de la segmentation en lignes qui, si elle échoue, handicape les étapes suivantes d'extraction et de reconnaissance de mots. Dans les documents arabe anciens, s'ajoute à ce problème, la présence dans les marges, d'annotations souvent composées de lignes obliques. La détection de ces lignes est nécessaire et constitue un défi important pour l'indexation de ces documents. Ainsi, la segmentation visée dans ce travail de thèse concerne l'extraction de lignes multi-orientées. Pour ce problème, la bibliographie ne présente que des techniques rudimentaires basées essentiellement sur une projection directe de l'image du document suivant une seule direction et donc non applicable à du texte multi-orienté. Devant ce manque, nous avons proposé une approche adaptative permettant de localiser d'abord les zones d'orientation différentes, puis de s'appuyer sur chaque orientation locale pour extraire les lignes. Pendant ma thèse, j'ai développé les points suivants :

- Application d'un maillage automatique en utilisant le modèle de contour actif (snake).
- Préparation du signal de profil de projection en supprimant tous les pixels qui ne sont pas nécessaires dans le calcul de l'orientation. Ensuite, application de toutes les distributions d'énergie de la classe de Cohen sur le profil de projection pour trouver la meilleure distribution qui donne l'orientation.
- Application de quelques règles d'extension pour trouver les zones.
- Extraction des lignes en se basant sur un algorithme de suivi des composantes connexes.
- Séparation de lignes se chevauchant et se connectant en utilisant la morphologie des lettres terminales arabes.

Mots-clés: documents arabes manuscrits, extraction de lignes, estimation de l'orientation, modèle de contour actif (snake), distribution de Wigner-Ville, connexion de lignes.

Abstract

The indexing of handwritten scanned documents poses the problem of lines segmentation, if it fails, disabling the following steps of words extraction and recognition. In addition, the ancient Arabic documents contain annotations in the margins, often composed of lines obliquely oriented. The detection of these lines is important as the rest and is a major challenge for the indexing of these documents. Thus, the segmentation described in this thesis involves the extraction of multi-oriented lines. For this problem, the bibliography has only rudimentary techniques based essentially on the projection of the document image along one direction, which be failed in the case of multi-oriented documents. Given this lack, we have proposed an adaptive approach that first locates the different orientation zones, then based on each local orientation to extract the lines. During my thesis, i particularly invested on the following points :

- Applying an automatic paving using the active contour model (snake).
- Preparation the signal of the projection profile by removing all pixels that are not needed in the orientation estimation. Then, implementation of all energy distributions of Cohen's class on the projection profile to find the best distribution that gives the orientation.
- Applying some extension rules to find the oriented zones.
- Extraction of lines by using an connected components follow-up algorithm.
- Separation of overlapped and touched lines using the morphology of Arabic terminal letters.

Keywords: handwritten Arabic documents, text line extraction, orientation estimation, active contour model (snake), Wigner-Ville distribution, touching lines.

Remerciements

Je tiens tout d'abord à remercier mon directeur de thèse M. Abdel Belaïd, Professeur à l'université de Nancy 2 et responsable de l'équipe READ au LORIA, avec qui j'ai eu la chance et le plaisir de travailler durant ces trois années. Ses conseils aux niveaux scientifique et personnel ont toujours été clairs et m'ont permis d'aboutir à la production de ce travail. Je le remercie aussi pour sa disponibilité et les discussions régulières que nous avons eues.

Je remercie M. Christian Olivier et M. Jean-Yves Ramel pour avoir accepté de rapporter ma thèse et pour les efforts que leur a demandés ce travail. Leurs critiques étaient constructives et leurs indications m'étaient utiles.

Je présente mes vifs remerciements à M. Salvatore-Antoine Tabbone qui m'a fait l'honneur de présider le jury et pour sa lecture attentive du manuscrit.

J'exprime mes sincères remerciements à M. François Auger pour sa collaboration, ses conseils, son enthousiasme et sa disponibilité qui m'ont permis d'approfondir l'étude de la classe de Cohen.

Je tiens à remercier l'association qui a financée ma thèse (Scientific Association for Developement). Je tiens à remercier la bibliothèque nationale de Tunis, la bibliothèque nationale de médecine aux U.S.A, les archives et bibliothèque nationales d'Egypte qui ont bien voulu mettre à disposition du public leurs documents et qui ont constitués notre ressource essentielle de données.

Je tiens aussi à remercier Dr. Ahmad Shahin pour son amitié, ses conseils ainsi que pour l'aide et la motivation qu'il m'a transmis.

Je remercie tous les membres de l'équipe Read qui m'ont permis d'effectuer ce travail dans la bonne humeur. Je dois un grand merci aux membres (actuels et passés) du LORIA que j'ai pu côtoyer durant mon séjour. Dans l'ordre alphabétique : Yolande Belaïd, Rokia Bendaoud, Hubert Cecotti, Julien Clément, Ilham Esslimani, Stéphane Goria, Hatem Hamza, Pierre Humbert, Hanen Maghrebi, Emilie Phillipot, Yves Rangoni et Sizlard Vajda. J'en oublie sûrement, qu'ils veuillent bien m'en excuser. Merci aussi à ceux qui m'ont aiguillé pendant mon enseignement, Azim Roussanaly et Pascal Fontaine et bien sûr toute l'équipe pédagogique du C2I. Un merci bien sûr à Jean-Luc Husson qui m'a encadré lors de mon ATER ainsi que toute l'équipe pédagogique de l'IUT de Saint-Dié des Vosges.

Je remercie ma deuxième famille, c'est-à-dire mes amis et toutes les personnes que j'ai pu rencontrer au cours de mon chemin, pour m'avoir donné le meilleur d'eux-mêmes.

Dans l'ordre alphabétique : Zainab Assaghir, Othman Baker, Ahmad Dib, Ziad El-Ali, Hassan El-Masri, Ahmad Hamad, Kamel Menighed, Bilal Missayké, Mohamad Nassar, Jamal Sabouneh et Zein El Abidin Sabsabi.

Enfin, je veux exprimer toute ma gratitude à mes parents et à ma famille qui m'ont soutenu et encouragé depuis toujours et sans lesquels je ne serais pas parvenu à accomplir ce travail.

Nazih Ouwayed

Je dédie ce travail
à ma mère Fatima,
à mon père Youssef,
Ainsi qu'à mes deux princesses, ma puce Jamal et sa maman Farah.

Table des matières

Table des figures **xi**

Liste des tableaux **xv**

Introduction générale **1**

Partie I Étude bibliographique

Chapitre 1
Description de la base de documents **5**

Chapitre 2
État de l'art **9**

2.1 Approches descendantes 9
2.1.1 Projection 9
2.1.2 Modèle de document 12
2.1.2.1 Découpage en XY (XY-CUT) 12
2.1.2.2 Diagramme de Voronoï 12
2.2 Approches ascendantes 12
2.2.1 Classification par k-ppv 13
2.2.2 Transformée de Hough 15
2.2.3 Technique de lissage 19
2.2.4 Réseau répulsif-attractif 20
2.2.5 Regroupement par l'arbre recouvrant minimal 20
2.2.6 Regroupement par la déformation de contour localisé et propagation 22

2.2.6.1 Le snake paramétré 23
2.2.6.2 Le Level set 23
2.2.6.3 Le Mumford-Shah 26
2.3 Discussion des méthodes 27
2.4 Conclusion 28

Partie II Extraction et séparation de lignes multi-orientées

Architecture du système proposé **33**

Chapitre 1
Maillage automatique **35**

1.1 Introduction 35
1.2 Application du maillage 36
1.3 Modèle de contour actif (*snake*) 38
1.4 Sélection des fenêtres 41
1.5 Expérimentations 45
1.6 Conclusion 46

Chapitre 2
Estimation de l'orientation

2.1 Introduction 47
2.2 Calcul de l'orientation par analyse des histogrammes de projection . . 49
2.3 Les distributions temps-fréquence 51
2.4 Application des distributions temps-fréquence pour l'estimation de l'orientation 59
2.5 Expérimentations 61
2.6 Conclusion 69

Chapitre 3
Extension et correction des zones d'orientation

3.1 Introduction 73
3.2 Détection de l'orientation erronée 74
3.3 Extension de l'orientation 75

3.4 Correction du maillage 76
3.5 Expérimentations 77
3.6 Conclusion 78

Chapitre 4
Suivi de lignes

4.1 Introduction 83
4.2 Algorithme 84
4.3 Expérimentations 85
4.4 Conclusion 88

Chapitre 5
Séparation des lignes connectées

5.1 Introduction 91
5.2 Les types de connexion 93
5.3 L'analyse de la morphologie Arabe 93
5.4 Direction correcte des lettres connectées 94
5.5 Application 95
5.6 Expérimentations 98
5.7 Conclusion 99

Conclusion **101**

Conclusion et perspectives **103**

Annexes **105**

Annexe A
Outils de développement

A.1 Matlab 105
A.2 Notre système 106
A.3 Outil temps-fréquence "TFTB" 107

Annexe B
Expérimentations

Annexe C
Base de documents

C.1 Documents simples 119
C.2 Documents multi-orientés 120
C.3 Documents multi-tailles 120
C.4 Documents avec problèmes au centre 121

Annexe D
Publications

D.1 Revues nationales avec comité de lecture 123
D.2 Conférences internationales avec actes et comité de lecture 123
D.3 Talks 124

Bibliographie **125**

Index des auteurs **133**

Index **137**

Table des figures

1.1 Exemples des quatre catégories de documents manuscrits : (a) Arabes mono-orientés (sources : [BNT, 2009]), (b) Arabes multi-orientés (source : [NLM, 2009]), (c) multi-scripts (Arabe et Latin) (sources : [BNT, 2009]) et (d) hétérogènes (source : [NLM, 2009]). 6
1.2 Exemple de 4 documents arabes manuscrits multi-orientés. Les orientations dans les marges sont de deux types : prolongation des lignes principales (les lignes dans la partie horizontale au milieu du document) (en rouge) ou des nouvelles lignes (en bleu). 7
1.3 Extrait d'un document arabe manuscrit présentant des cas de connexion (unique et multiple) (source : ABNE). Le cas des connexions multiples ne sera pas traité ici. 8

2.1 (a) maillage de document en colonnes et (b) profil de projection de la première colonne à droite [Bennasri *et al.*, 1999]. 10
2.2 Extraction de lignes d'un document manuscrit arabe [Bennasri *et al.*, 1999]. . . . 10
2.3 Étapes d'extraction de lignes : (a) application d'un filtre gaussien, (b) tracé des chemins de suivi, (c) extraction des lignes du centre et (d) résultat final d'extraction de lignes [Nicolaou et Gatos, 2009]. 11
2.4 Étapes d'extraction de lignes : (a) maillage en plusieurs colonnes du document, (b) segmentation des colonnes en 3 types de blocs (petits en violet, moyens en rouge et grands en bleu), (c) résultat finale d'extraction de lignes [Zahour *et al.*, 2004, Zahour *et al.*, 2007]. 13
2.5 Étapes d'extraction de lignes : (a) les masques de regroupement, (b) les points d'ancrages (en gris) et (c) résultat final [Likforman-Sulem et Faure, 1994]. 14
2.6 Étapes d'extraction de lignes : (a) détection des points terminaux (en étoile), de la ligne de base (en trait continu) et de la ligne du centre (en trait interrompu) et (b) résultat final d'extraction de lignes [Feldbach et Tönnies, 2001]. 15
2.7 Illustration de la transformée de Hough pour quelques points d'une ligne droite : (a) ligne originale et (b) sa transformée de Hough (le cercle indique le croisement (un point) des sinusoïdes, source [WIK, 2009]). 16
2.8 Illustration de la transformée de Hough pour une ligne non-droite : (a) ligne originale et (b) sa transformée de Hough (le cercle indique le croisement (une zone), source [TSI, 2009])). 16

2.9 Étapes d'extraction de lignes : (a) document original et (b) résultat de la segmentation en lignes [Likforman-Sulem *et al.*, 1995]. 17
2.10 Exemple d'extraction de lignes : (a) les minima dans la bande verticale à l'extrémité gauche de l'image, (b) les 6 directions et (c) résultat de la segmentation en lignes [Likforman-Sulem *et al.*, 1995]. 18
2.11 Exemple d'extraction de lignes sur un document contenant des notes en Latin [Malleron *et al.*, 2009]. 18
2.12 Exemple d'extraction de lignes sur un document manuscrit Latin (chaque couleur représente une ligne) [Louloudis *et al.*, 2009a]. 19
2.13 Extraction de lignes d'un document arabe (de type Ottman) [Oztop *et al.*, 1999]. 21
2.14 Étapes d'extraction de lignes sur des documents Chinois : (a) résultat de l'arbre recouvrant minimal et (b) résultat de l'extraction des lignes [F. Yin, 2008]. . . . 21
2.15 Étapes d'extraction de lignes sur un document Latin : (a) résultat de graphe et (b) résultat de l'extraction des lignes [Nicolas *et al.*, 2004]. 22
2.16 Étapes d'extraction de lignes : (a) image initial, (b) snake de départ (en rouge) sur l'image initial lissée en appliquant un filtre gaussien, (c) snake déformé et (d) résultat de l'extraction de lignes [Bukhari *et al.*, 2009]. 24
2.17 Étapes d'extraction de lignes pour un document Arabe : (a) image initiale, (b) estimation de la densité des pixels en appliquant un filtre gaussien, (c) estimation initiale des lignes de texte, (d) résultat après 10 itérations par la méthode des lignes de niveaux, (e) résultat après la connexion des lignes de texte voisines et (f) résultat final de l'extraction des lignes [Li *et al.*, 2008]. 25
2.18 Exemple d'extraction de lignes pour trois documents (a) Chinois, (b) Indien et (c) Coréen [Du *et al.*, 2009]. 26

19 Architecture du système *DEMOT*. 34

1.1 Les 3 types de maillage, (a) régulier, (b) irrégulier et (c) semi-régulier. 36
1.2 Maillage par colonne de document : (a) Bangla et (b) Arabe. 37
1.3 Algorithme de maillage automatique (d1, d2 et d3 sont les distances entre les lignes). 37
1.4 Modèle de contour actif (*snake*) : (a) position initiale du snake (cercle rouge), (b) snake après quelques itérations de déformation et (c) snake final [sna, 2009]. . . . 38
1.5 Types de modèle de contour actif (*snake*) : (a) snake ouvert et (b) snake fermé [sna, 2009]. 39
1.6 Énergie externe (indiquée en flèches) pour un objet qui contient des concavités : (a) gradient et (b) flux du vecteur du gradient (GVF) [Xu et Prince, 1997]. . . . 41
1.7 Application du snake pour la détection de lignes, (a) l'image initiale, (b) l'image initiale après une dilatation et une érosion, (c) l'axe majeur tracé pour chaque composante connexe des lignes. L'ellipse englobe la première composante connexe de (b), (d) montre la déformation du snake (c) et (e) donne le résultat final montrant les composantes connexes groupées dans chaque ligne. 42
1.8 Résultats du maillage automatique des documents présentés dans la figure 1.2. . 43
1.9 Exemple d'une fenêtre vide. 44
1.10 Exemple d'une fenêtre presque vide (contient peu de pixels). 45

2.1 Histogrammes de projection d'un document ne comportant qu'une seule ligne, calculés pour des angles de projection de $+30\,^\circ$, $0\,^\circ$ et $-30\,^\circ$. 49
2.2 Histogrammes de projection d'un document comportant plusieurs lignes, calculés pour des angles de projection de $+30\,^\circ$, $0\,^\circ$ et $-30\,^\circ$. 50
2.3 Problèmes d'orientation dans les documents arabes : (a) estimation erronée dûe aux orientations obliques de PAWs, (b) Apparition de maxima locaux à cause du bruit (cercles rouges). 50
2.4 Illustration des propriétés des distributions de la classe de Cohen et plus particulièrement de la distribution de Wigner-Ville (figures a, b et c) et de la distribution pseudo Margenau-Hill (figure d). Le signal $x(t)$ utilisé dans les figures a et b est égal à $\max(\sin(2\pi\, t/64), 0)$. 57
2.5 Exemples de valeurs maximales de la distribution de Wigner-Ville obtenues pour différents histogrammes de projection. 59
2.6 Résultats de l'estimation de l'orientation initiale du document présenté dans la figure 1.2. Les couleurs représentent des zones d'orientation suivant l'échelle des couleurs montrée à droite. 63
2.7 Document Arabe manuscrit incliné de $+14.7\,^\circ$. La position du centre de gravité des pixels noirs de ce document, utilisée pour calculer les histogrammes de projection, est indiquée par une croix. 64
2.8 Courbes d'énergie : maximum des distributions de Wigner-Ville (traits pleins en rouge), de spectrogramme (traits pointillés en bleu) et de la transformée de Fourier discrète (traits mixtes en violet) de la racine carrée des histogrammes de projection du document de la figure 2.7 et maximum de la moyenne de différence entre maxima et minima (traits mixtes en vert) des histogrammes de projection du même document, en fonction de l'angle de projection. 65
2.9 Quelques histogrammes de projection du document de la figure 2.7. 66
2.10 Partie d'un document arabe manuscrit. La largeur et la hauteur de cette image sont de 300 pixels et son inclinaison est de $0\,^\circ$. 69
2.11 Échantillon de 6 documents parmi les 864 documents traités. 70
2.12 Évolution de l'estimation de l'angle d'inclinaison du document de la figure 2.10 en fonction de la largeur du document, pour quatre valeurs différentes de la hauteur. L'inclinaison est estimée en utilisant la distribution de Wigner-Ville. 71

3.1 Les quatres règles d'extension de l'orientation vers les fenêtres voisines : (a) Est-Ouest, (b) Est-Sud-Ouest, (c) Sud-Nord et Nord-Sud et (d) Est-Nord-Ouest. . . 74
3.2 Algorithme de correction de l'orientation : (a) fenêtre avec une orientation erronée et (b) correction de l'orientation. 76
3.3 Résultats d'extension de l'orientation. Chaque couleur représente une zone d'orientation. 79
3.4 Exemple d'un maillage incorrect dans une zone horizontale. 80
3.5 Résultats de la correction du maillage pour les documents représentés dans la figure 1.2.a et b. 80

3.6 Échantillon de 4 documents très complexes où l'approche proposée trouve ses limites. 81

4.1 Étapes de détection de lignes dans une fenêtre. 84
4.2 Attribution des composantes connexes et des symboles diacritiques à leur ligne. . 86
4.3 Résultats de l'algorithme d'extraction de lignes (les cercles bleus indiquent les régions de connexion). 87

5.1 Extrait d'un document représentant quelques lignes qui se connectent. 92
5.2 Cas de connexion entre deux lettres : (a) Arabe, (b) même lettres que (a) avec une autre écriture (c) Latin. 92
5.3 (a) L'alphabet Arabe (la flèche indique la direction de l'écriture), (b) exemple de connexion et (c) connexion causée par le type du script (Diwani). 93
5.4 (a) Exemple de composantes connexes arabes se touchant (la lettre Arabe "ra ر" touche la lettre "alif ا") et (b) algorithme de calcul du vecteur des variations angulaires. 95
5.5 Les étapes de l'application de séparation des composantes connectées. 97
5.6 Cas de séparation incorrecte (le flèche rouge indique le vrai sens de suivi qui n'est pas atteint. 98
5.7 Exemples de résultats de séparation de composantes connectées. 99

A.1 Notre système de traitement d'image et de segmentation des documents "OuwayedImaPro" pour "**Ouwayed Im**age **Pro**cessing". 107

Liste des tableaux

2.1 Les méthodes proposées dans la littérature pour la segmentation des lignes (I. O. : Interval d'Orientation, G : Gris, B : Binaires et C : Couleurs). 29

1.1 Les principes du modèle de contour actif : paramétrique et géométrique. 39
1.2 Variation du taux d'erreur par rapport aux trois paramètres (α, β et μ) pour un nombre d'itérations égal à 25. 45

2.1 Propriétés des distributions temps-fréquence de la classe de Cohen considérées comme nécessaires ou souhaitables dans ce contexte d'application. 56
2.2 Distributions temps-fréquences retenues dans cette étude, avec leurs expressions et leurs propriétés. 58
2.3 Estimations de l'angle d'inclinaison du document de la figure 2.7 obtenues avec les représentations temps-fréquence retenues dans cette étude. 67
2.4 Taux d'estimation correcte de l'inclinaison obtenue avec les représentations temps-fréquence retenues dans cette étude. Le temps moyen correspond à la durée moyenne de traitement d'une image de 500 $\times$ 500 pixels. 67
2.5 Estimation de l'inclinaison des six documents de la figure 2.11 obtenues à l'aide des neuf représentations temps-fréquence retenues dans cette étude. 71

3.1 Résultats de l'extraction des zones d'orientation. 77

4.1 Résultats de l'extraction des zones d'orientation. 86
4.2 Synthèse des résultats obtenus par les méthodes proposées dans la littérature pour la segmentation en lignes de documents. (n.c. : non connu). 89

5.1 Les 4 types de connexion observés dans les documents Arabe manuscrits. 94
5.2 Résultats de l'approche de la séparation des composantes connectées. 98

B.1 Résultats de la détection de zones pour un échantillon de 8 documents appartenant à 4 classes différentes. 110
B.2 Résultats de la détection de lignes pour un échantillon de 8 documents appartenant à 4 classes différentes. 114

Introduction générale

Le problème étudié concerne la segmentation en lignes de documents arabes anciens. Des exemples de ces documents se trouvent dans la figure 1.2.

L'écriture ancienne, quelque soit la langue, est par nature complexe du fait de son irrégularité due au côté manuel du script. Rarement, les scripteurs utilisaient des lignes support (ou calques) pour écrire, ce qui occasionne des sinuosités dans les lignes d'écriture. De plus, à cause du style calligraphique de l'écriture arabe, des ligatures sont facilement introduites entre les parties de mots et des attachements sont produits entre les mots des lignes successives. Ajoutons à cela que comme le document n'existait que sur papier, sa mise à jour se faisait directement sur le texte lui-même, ce qui conduisait soit à prolonger les lignes dans les marges, soit à ajouter des blocs entiers de lignes dans ces marges.

Tous ces artefacts complexifient le problème de la segmentation en lignes qui est essentiellement contextuel, dans l'Arabe, alors que la plupart des techniques de segmentation existantes dans la littérature sont plutôt "physiques", cherchant essentiellement des alignements parallèles de composantes connexes. Le problème de cette segmentation "contextuelle" ou "dynamique" en lignes est donc nouveau et complexe, ce qui constitue le défi lancé dans cette thèse.

La littérature évoque des techniques de segmentation globale et locale. Les techniques globales procèdent en général par projection suivant une direction, ce qui ne peut fonctionner ici à cause de la sinuosité des lignes et de leur multi-orientation. Les techniques locales essaient de réaliser la projection dans des zones plus réduites et donc contenant moins de sinuosités. Ces techniques sont efficaces quand les lignes sont mono-orientées mais perdent vite de leur efficacité quand les lignes sont multi-orientées, comme dans notre cas.

Devant ce constat d'échec des méthodes classiques, nous avons proposé une méthode qui commence par un repérage des zones d'orientation, laquelle orientation est ensuite exploitée avantageusement pour une extraction appropriée des lignes. L'extraction de zones d'orientation se fait à l'aide d'un maillage minimal au début puis élargi par association de mailles voisines de même orientation. Le calcul d'orientation utilise la distribution d'énergie de la classe de Cohen, plus précise que la méthode de différenciation des minima-maxima dans les histogrammes de projection. Cette étape de calcul d'orientation est fondamentale. Elle garantit les étapes suivantes d'extraction et de suivi de lignes. L'approche se termine par une phase finale de séparation des lignes connectées fondée

sur l'exploitation de la morphologie des lettres terminales. La méthode proposée nécessite le déploiement de plusieurs techniques. L'approche proposée exige seulement que les documents soient binaires et nettoyés du bruit de numérisation.

Le mémoire est organisé de la manière suivante :

- La première partie propose une étude préalable du problème. Elle se divise en deux chapitres. Dans le premier chapitre, nous décrivons la base de documents étudiés dans cette thèse. Dans le deuxième chapitre, les méthodes pour l'extraction des lignes sont étudiées et détaillées.
- La deuxième partie présente notre approche pour l'extraction et la séparation des lignes. Cette partie se divise en cinq chapitres. Le premier chapitre décrit le maillage automatique. Dans le deuxième chapitre, la méthode proposée pour l'estimation de l'orientation est décrite en présentant la classe de Cohen et l'application des ses distributions. Le troisième chapitre est consacré à l'algorithme de fission-fusion pour grouper les pavés selon leur orientation afin de trouver les zones multi-orientées. L'algorithme de suivi pour l'extraction des lignes est détaillé dans le quatrième chapitre. On termine cette partie par l'exposé de la méthode de séparation des lignes connectées.
- Le mémoire se termine par une conclusion et des perspectives qui montreront un prolongement possible du système proposé.

Première partie

Étude bibliographique

Chapitre 1

Description de la base de documents

De manière générale, on peut classer les documents manuscrits dans quatre catégories en fonction de leur aspect physique, et ceci quelque soit la langue. Ces classes sont :

1. **Documents mono-orientés :** les lignes dans cette classe sont orientées suivant une seule direction. La figure 1.1.a montre un document Arabe manuscrit avec une orientation horizontale.
2. **Documents multi-orientés** : les lignes dans ces documents occupent ou non toute la largeur du document ou sont rangées par blocs. Elles ont des orientations différentes. La figure 1.1.b donne un exemple de cette classe de documents.
3. **Documents multi-scripts :** il s'agit ici de documents écrits par plusieurs personnes différentes, conduisant à des écritures ou scripts différents. Cela arrivait souvent par le passé, des personnes se succédaient pour compléter un même document, ou collaboraient à l'écriture d'un même ouvrage. De nos jours, certains documents modernes peuvent aussi être multi-scripts, comme par exemple les cartes de voeux, les écrits destinés à différentes nationalités, etc. La figure 1.1.c montre un document manuscrit multi-scripts (Arabe et Latin). La zone d'écriture arabe contient deux orientations.
4. **Documents hétérogènes** : ces types de documents contiennent à la fois du texte et des images ou des illustrations. Beaucoup de documents cartographiques, mécaniques ou architecturaux sont de ce type. Les documents manuscrits de ce type peuvent contenir plusieurs orientations correspondant à des lignes de cotation, ou d'illustration des dessins. La figure 1.1.d montre un document Arabe manuscrit hétérogène.

Les documents étudiés dans cette thèse sont des anciens documents manuscrits arabes mono et multi-orientés. Ils sont issus de trois grandes bibliothèques : la Bibliothèque Nationale de Tunis (BNT [1]), la Bibliothèque Nationale de médecine des U.S.A (BNMU [2])

1. http ://www.bibliotheque.nat.tn/
2. http ://www.nlm.nih.gov/hmd/arabic/welcome.html

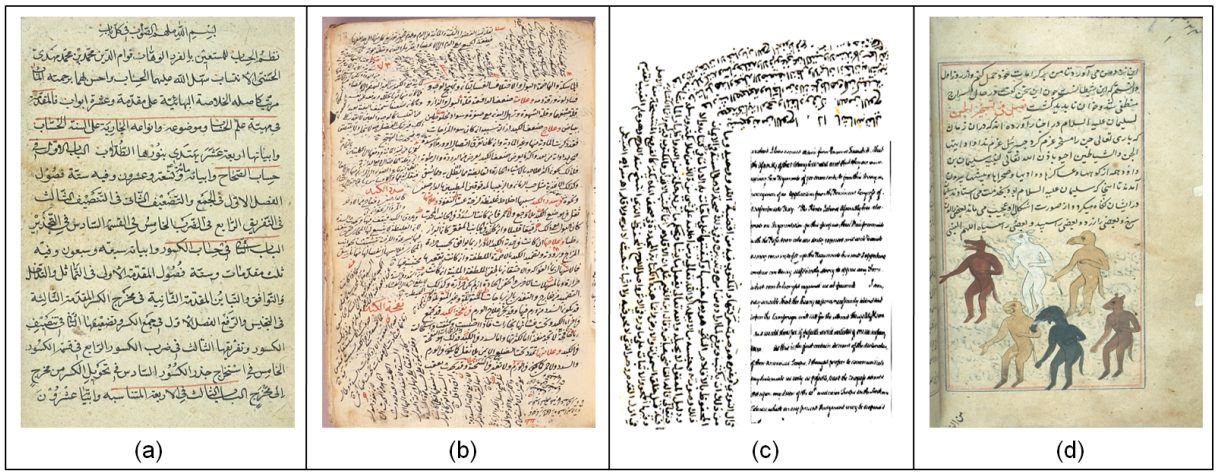


FIGURE 1.1: Exemples des quatre catégories de documents manuscrits : (a) Arabes mono-orientés (sources : [BNT, 2009]), (b) Arabes multi-orientés (source : [NLM, 2009]), (c) multi-scripts (Arabe et Latin) (sources : [BNT, 2009]) et (d) hétérogènes (source : [NLM, 2009]).

et les Archives et Bibliothèque Nationales d'Egypte (ABNE [3]).

Ces documents contiennent en général une zone centrale horizontale et des annotations multi-orientées dans les marges. Lors de la collecte de nos données, nous avons recensé 180 documents (100 multi-orientés, 72 mono-orientés, 3 multi-scripts et 5 hétérogènes). Le pourcentage des documents multi-orientés est de 55.5%, ce qui explique la motivation de la recherche entreprise dans cette thèse. Les documents de la BNT sont de faible résolution (72-96 *dpi*) tandis que ceux de la BNMU et ABNE sont de moyenne et haute résolution (150-300 *dpi*). La figure 1.2 montre un exemple de 4 documents choisis arbitrairement parmi nos 100 documents multi-orientés. On remarque bien dans ces 4 documents la présence d'annotations orientées dans les marges. Par exemple, le document de la figure 1.2.b contient 4 annotations.

Dans cette thèse, nous travaillerons sur les documents multi-orientés en vue d'extraire les lignes, et sur les documents mono-orientés pour estimer l'orientation globale. La difficulté de la segmentation en lignes de documents multi-orientés provient de l'irrégularité de l'orientation des lignes et surtout de la présence aléatoire de prolongement des lignes dans les marges. De plus, les lignes successives se connectent facilement par les ascendants et les descendants qui sont souvent nombreux et très accentués dans les écritures arabes, en vue de donner à l'écriture un attrait calligraphique (voir figure 1.3). Des lignes de plusieurs orientations différentes peuvent coexister dans des zones d'une même page (voir la figure 1.2). La présence massive de points diacritiques complique la tâche. La variation des formes des lettres, inhérente à l'écriture arabe, multiplie d'autant les problèmes de séparation de lignes et de mots dans les lignes [Likforman-Sulem et Faure, 1994].

3. http ://portal.unesco.org/ci/photos/showgallery.php/cat/559

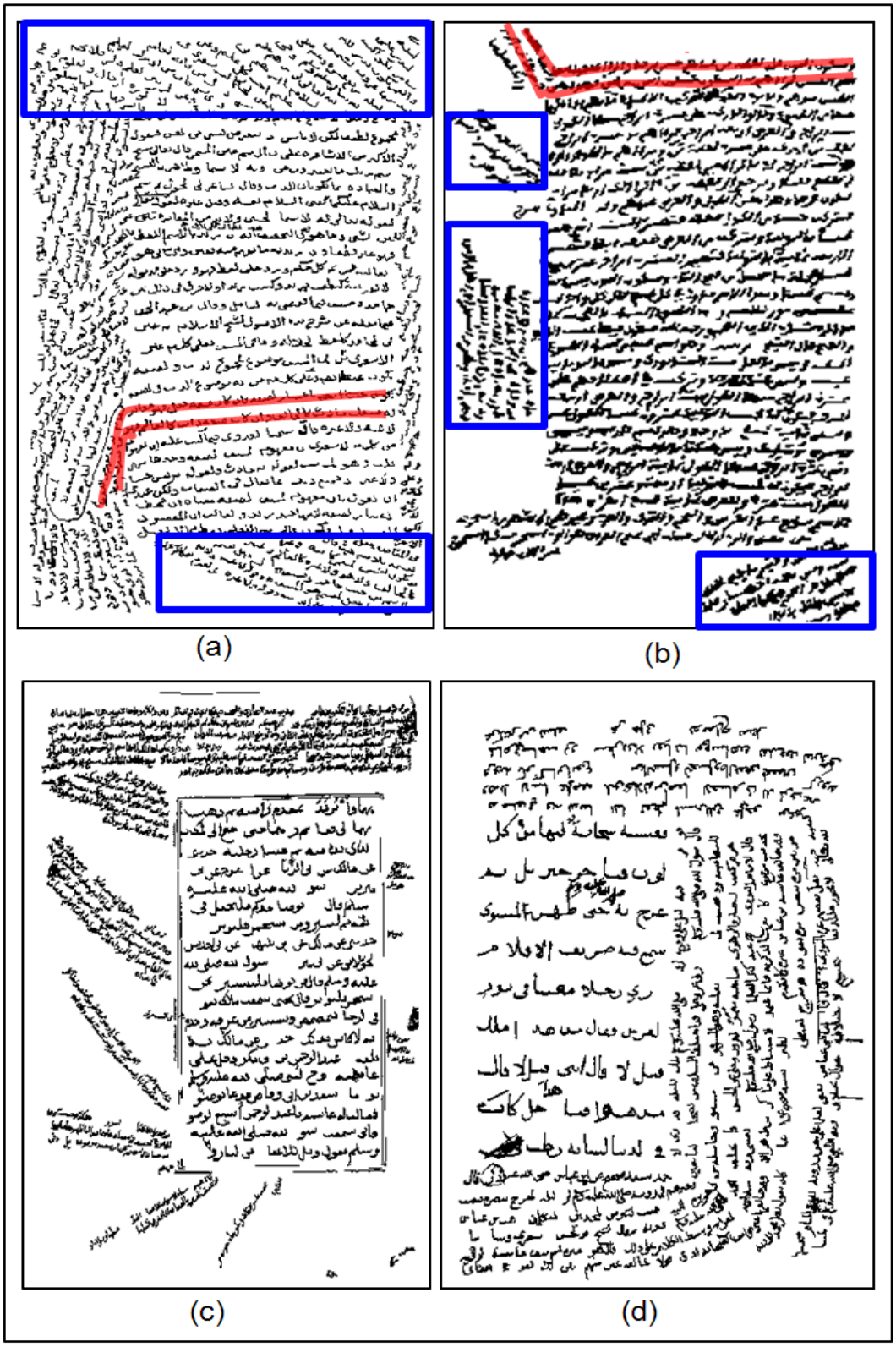


FIGURE 1.2: Exemple de 4 documents arabes manuscrits multi-orientés. Les orientations dans les marges sont de deux types : prolongation des lignes principales (les lignes dans la partie horizontale au milieu du document) (en rouge) ou des nouvelles lignes (en bleu).

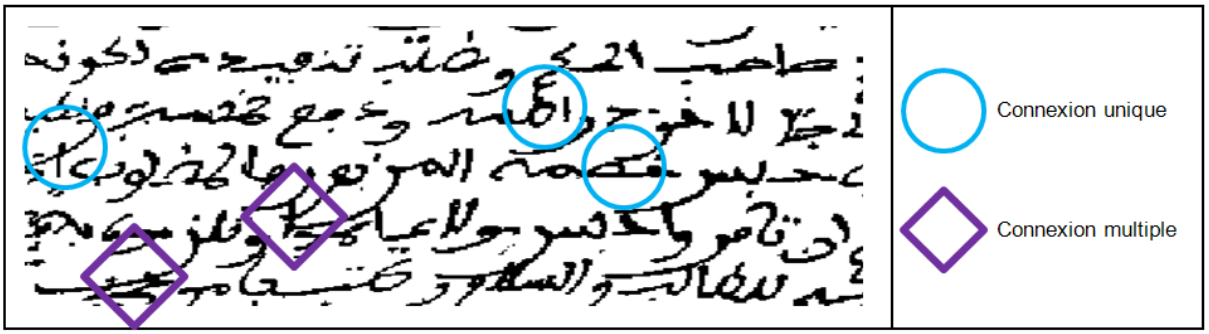


FIGURE 1.3: Extrait d'un document arabe manuscrit présentant des cas de connexion (unique et multiple) (source : ABNE). Le cas des connexions multiples ne sera pas traité ici.

Chapitre 2

État de l'art

Un document numérisé est, à la base, une image de pixels. Nous nous focaliserons dans ce chapitre sur les techniques de segmentation d'images en lignes. Les techniques existantes dans la littérature sont réparties en techniques descendantes et ascendantes.

2.1 Approches descendantes

Les approches descendantes considèrent le document dans sa globalité et une seule orientation est en principe recherchée. Dans cette classe, le document est décomposé en allant du plus large (blocs) au plus fin (CCX [4] ou pixel). Dans cette catégorie, nous trouvons la technique de la projection et la méthode guidée par le modèle de document.

2.1.1 Projection

Pour détecter les lignes, on suit en général les maxima du profil de projection du document entier ou d'une bande de ce document en respectant l'orientation trouvée. Le profil de projection est obtenu en additionnant les valeurs des pixels sur l'axe de l'orientation du document binaire.

Ensuite, les maxima et les minima du profil sont déterminés et les lignes entre deux minima consécutifs sont recherchées. Ces composantes connexes forment une ligne si la taille du pic engendré correspond à la largeur de la ligne. Bennasri et al. ont proposé dans [Bennasri *et al.*, 1999], une méthode pour extraire les lignes d'un texte manuscrit arabe, en utilisant la projection. D'abord le document est divisé en plusieurs colonnes afin de corriger le problème de sinuosité. Puis, les points de départ de toutes les lignes sont détectées en utilisant les minima du profil de projection partielle (de chaque colonne, voir figure 2.1). Ensuite, un suivi de contour partiel de chaque ligne est effectué : d'abord dans le sens de l'écriture, puis dans le sens opposé. À l'issue de cette opération, les lignes

4. Composante ConneXe.

adjacentes qui se chevauchent sont alors séparées (voir figure 2.2). Les points diacritiques, sur lesquels demeure un doute sont marqués et leur affectation à l'une ou l'autre des lignes voisines définitive est validée lors de la reconnaissance. La méthode a été testée sur une centaine de textes arabes manuscrits mono-orientés et multi-scripteurs contenant 2000 lignes. Le taux d'extraction atteint par cette méthode est de 98%.

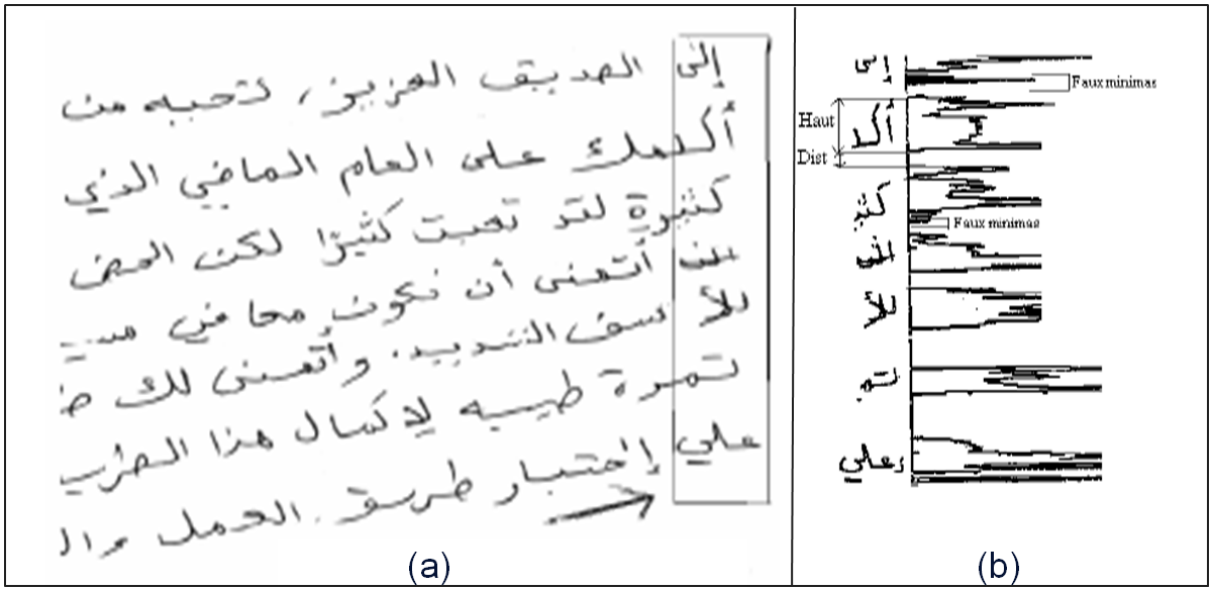


FIGURE 2.1: (a) maillage de document en colonnes et (b) profil de projection de la première colonne à droite [Bennasri *et al.*, 1999].

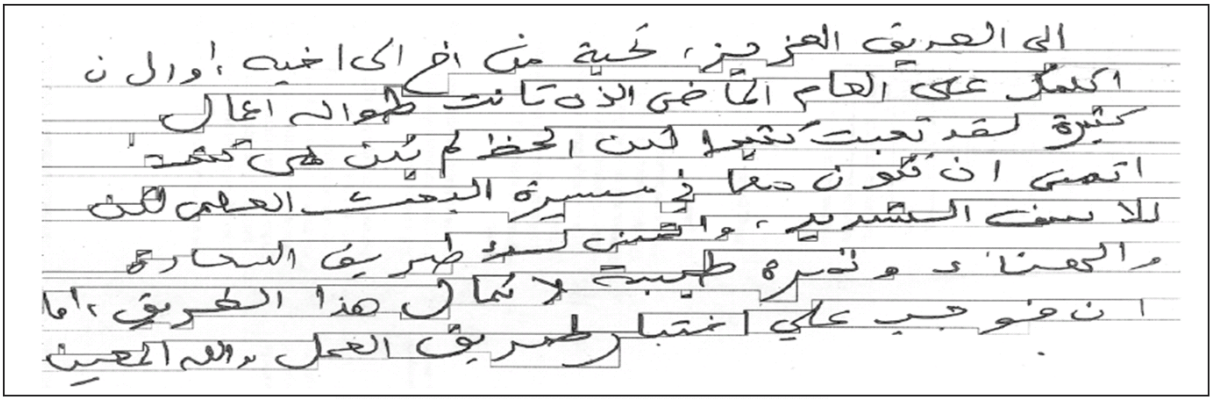


FIGURE 2.2: Extraction de lignes d'un document manuscrit arabe [Bennasri *et al.*, 1999].

Dans [Nicolaou et Gatos, 2009], Nicolaou et al. ont proposé une technique pour segmenter les lignes des documents manuscrits latins en utilisant des traceurs (axes) des minima. Ces minima sont estimés en utilisant l'histogramme de projection verticale. L'approche est basée sur l'hypothèse topologique que, pour chaque ligne du texte, il existe un

chemin allant d'un côté de l'image à l'autre qui ne traverse qu'une seule ligne (voir figure 2.3.b). Ces chemins correspondent aux alignements des minima locaux des profils de projections des lignes. Pour extraire les lignes, on applique avant la projection un filtre gaussien afin d'élargir les espaces entre les lignes et de diminuer les effets de bruit (voir figure 2.3.a). Une étape de post-traitement est appliquée pour attribuer les pixels manquants aux lignes les plus proches (voir figures 2.3.c et 2.3.d). Cette méthode a été testée sur une sous-base de documents d'ICDAR 2007 composée de 20 documents contenant 476 lignes et 80 documents contenant 1771 lignes. Cette technique atteint un taux d'extraction qui est égal à 98.6%.

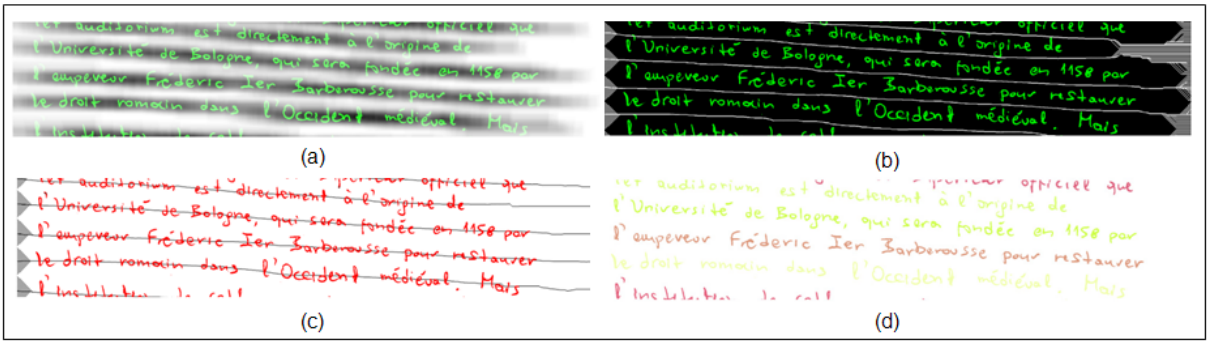


FIGURE 2.3: Étapes d'extraction de lignes : (a) application d'un filtre gaussien, (b) tracé des chemins de suivi, (c) extraction des lignes du centre et (d) résultat final d'extraction de lignes [Nicolaou et Gatos, 2009].

Shapiro et al. ont proposé dans [Shapiro *et al.*, 1993] une méthode d'estimation de l'orientation globale du document en appliquant une transformée de Hough à l'image entière. De cette transformation est déduit un angle d'orientation qui est ensuite utilisé pour faire des projections. Le nombre de maxima du profil de projection donne le nombre de lignes. Les maxima locaux sont rejetés en comparant leurs valeurs à celles des maxima les plus élevés. La recherche de lignes est effectuée entre deux maxima dans le profil de projection. Cette technique a été testée sur un ensemble de 200 pages avec une tâche de séparation des mots. Les auteurs donnent quelques résultats pour montrer la performance et l'efficacité de l'approche. Le taux d'extraction n'est pas mentionné.

Dans le travail de Antonacopoulos et Karatzas [Antonacopoulos et Karatzas, 2004], chaque minimum dans le profil de l'histogramme de projection est considéré comme un séparateur entre les lignes du texte. Ces séparateurs sont des points potentiels pour la segmentation en lignes. Un vecteur de distance entre les points potentiels est ensuite construit afin de trouver la distance la plus fréquente. Cette distance est considérée comme le seuil de segmentation. Lorsque deux points potentiels voisins ont une distance très proche du seuil, les composantes connexes entre ces deux points forment la ligne du texte. La méthode est appliquée sur des registres imprimés de la seconde guerre mondiale, qui comportent des espaces réguliers entre les lignes. La structure logique est utilisée pour

dériver les régions du texte où les noms d'intérêt peuvent être trouvés. Les auteurs donnent quelques résultats d'extraction de lignes sans toutefois mentionner de taux d'extraction.

2.1.2 Modèle de document

La méthode guidée par le modèle de document consiste à découper récursivement le document jusqu'au la détection des objets recherchés (paragraphes ou lignes). Dans ce principe, nous trouvons le découpage en XY et le diagramme de Voronoï.

2.1.2.1 Découpage en XY (XY-CUT)

L'algorithme de découpage en XY a été introduit par Nagy et al. dans [Nagy *et al.*, 1988]. Le principe consiste à découper un document binaire horizontalement et verticalement en plusieurs rectangles. Le découpage continue dans chaque rectangle d'une manière récursive jusqu'au un condition soit satisfait. Le condition d'arrêt est défini selon l'application souhaitée. Dans ce travail, les profils de projection horizontaux et verticaux sont étudiés pour définir les conditions d'arrêt afin d'extraire les lignes. La méthode a été appliquée sur des articles latins imprimés. Les auteurs donnent quelques résultats d'extraction de lignes sans toutefois mentionner de taux d'extraction. l'algorithme de découpage en XY a été utilisé pour des documents qui ne contiennent pas beaucoup des variations.

2.1.2.2 Diagramme de Voronoï

L'algorithme de Voronoï consiste à appliquer un pavage pour un document afin de trouver les zones souhaitées. Le pavage peut avoir différents types selon l'application souhaitée ce qui rend l'algorithme de Voronoï applicable pour des documents contenant beaucoup de variations. Shi et al. ont proposé dans [Lu *et al.*, 2004] une méthode pour l'extraction des mots en utilisant le diagramme de Voronoï. Tout d'abord, les contours des composantes connexes dans le document sont extraites afin de définir les points de pavage. Ensuite, les pavés sont regroupés entre eux afin de trouver les zones souhaitées (mots). Le regroupement des pavés se base sur la distance entre les composantes connexes. La méthode a été appliquée sur 904 documents latins imprimés. Cette technique atteint un taux de regroupement qui est égal à 99%.

2.2 Approches ascendantes

Les approches ascendantes sont basées sur les éléments de bas niveau de l'image comme les pixels ou les composantes connexes. Dans cette catégorie, nous trouvons la classification par k-ppv [5], à nouveau la transformée de Hough, la technique de lissage (smearing), la

5. plus proches voisins.

technique de réseau répulsif-attractif, le regroupement par l'arbre recouvrant minimal (MST [6]) et le regroupement par la déformation de contour localisé et propagation.

2.2.1 Classification par k-ppv

L'idée de la classification par k-ppv est de regrouper les éléments de bas niveau entre eux selon plusieurs critères comme la proximité afin de former les lignes. Plusieurs travaux ont été proposés dans la littérature qui utilisent ce principe.

Une approche pour l'extraction des lignes des anciens manuscrits arabes a été proposée par Zahour et al. dans [Zahour *et al.*, 2004, Zahour *et al.*, 2007]. Au départ, le document est divisé en plusieurs colonnes de même taille (voir figure 2.4.a). Puis, chaque colonne du document est segmentée en trois types de blocs de texte : les petits blocs qui représentent généralement les symboles diacritiques, les blocs moyens qui correspondent au corps du texte et les grands blocs qui reflètent le chevauchement entre les mots des lignes voisines (voir figure 2.4.b). Enfin, les blocs sont regroupés entre eux en utilisant les distances euclidiennes entre les ordonnées des bases des blocs. On compare les blocs de la colonne i avec ceux des colonnes $i-1$ et $i+1$, sauf pour la première et la dernière colonne. Les blocs dont la distance entre les ordonnées de leur base inférieure est minimale sont appariés pour former une ligne de texte (voir figure 2.4.c). L'approche a été testée sur 100 anciens documents manuscrits arabes. Le taux d'extraction atteint 96%.

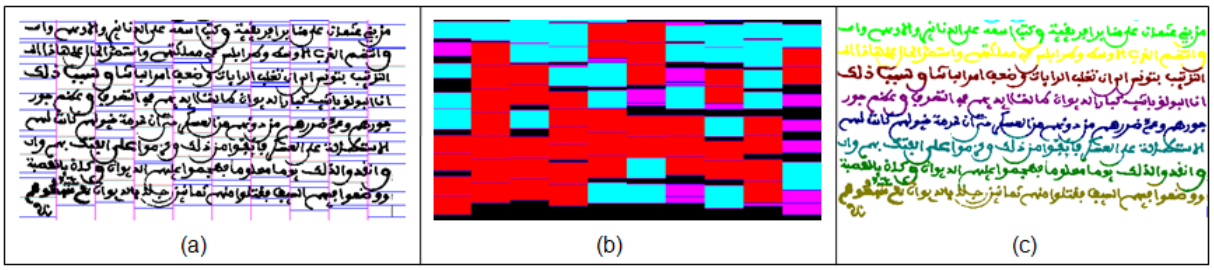


FIGURE 2.4: Étapes d'extraction de lignes : (a) maillage en plusieurs colonnes du document, (b) segmentation des colonnes en 3 types de blocs (petits en violet, moyens en rouge et grands en bleu), (c) résultat finale d'extraction de lignes [Zahour *et al.*, 2004, Zahour *et al.*, 2007].

Dans [Likforman-Sulem et Faure, 1994], les auteurs ont proposé une méthode itérative basée sur la perception du regroupement pour former les alignements. Au début, les composantes connexes dans l'image sont extraites. Puis, quatre masques correspondant à une discrétisation de l'espace en quatre directions (0°, 45°, 90°, 135°) sont définis (voir figure 2.5.a). Chaque masque m est de forme carrée et contient une zone d'excitation sensible à la densité d'écriture dans la zone. Ces masques sont appliqués sur toutes les

6. Minimum Spanning Tree.

composantes connexes afin d'extraire les points d'ancrage. Ces points sont des composantes connexes dont l'orientation est fiable (la réponse de cette composante connexe au masque m est $\geq$80%). Ils ont une forme compacte et « allongée » le long d'une direction : segment d'écriture sans hampe ni jambage, lettre longiligne (voir figure 2.5.b). Ensuite, les composantes connexes sont regroupées en se basant sur la théorie de la "Gestalt theory" dont trois des principes qui permettent à l'oeil humain de percevoir des groupements d'éléments sont retenus. Ces principes sont notamment la proximité, la similarité et la continuité de la direction. Une composante connexe c_1 sera groupée avec la composante c_2 si c_1 est suffisamment proche de c_2, si c_1 est dans la direction de l'alignement (donnée par le point d'ancrage) et si c_1 est de taille similaire à c_2. Ce regroupement construit les lignes de texte (voir figure 2.5.c). Cette technique a été expérimentée sur 56 documents (35 brouillons de texte et 21 enveloppes postales). Elle atteint un taux d'extraction de 86% pour les brouillons et de 96.3% pour les enveloppes postales.

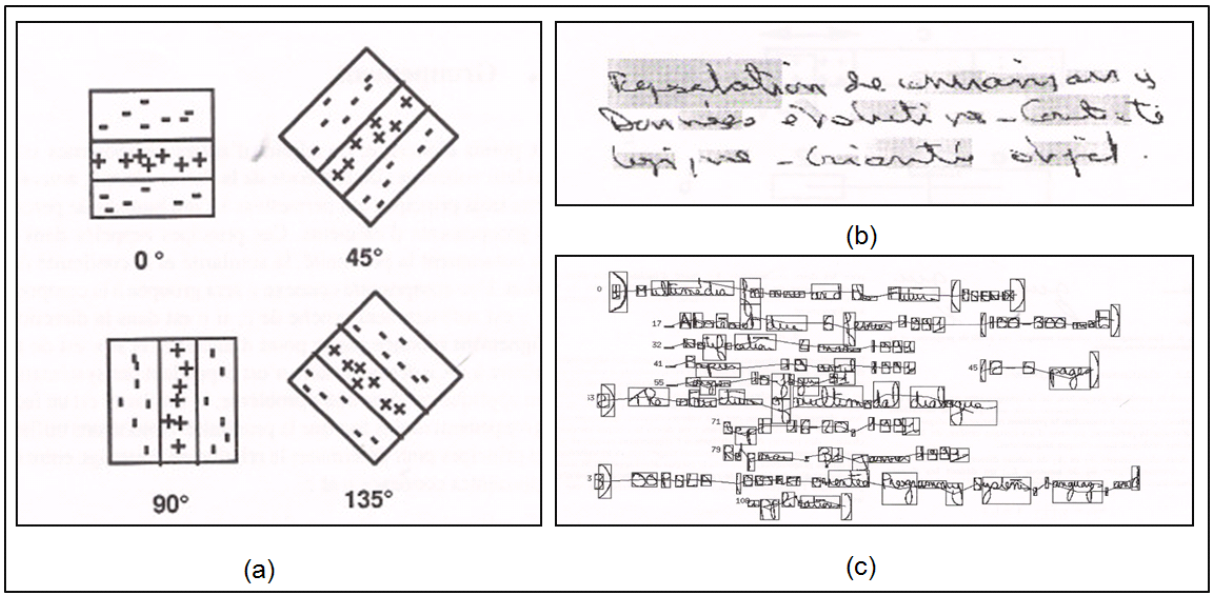


FIGURE 2.5: Étapes d'extraction de lignes : (a) les masques de regroupement, (b) les points d'ancrages (en gris) et (c) résultat final [Likforman-Sulem et Faure, 1994].

Feldbach et Tönnies ont proposé dans [Feldbach et Tönnies, 2001] une méthode pour extraire les lignes des documents de registres des anciennes églises. Ces documents comprennent des fluctuations, des connexions entre les lignes. La méthode employée utilise les lignes de base et les lignes de centre pour extraire les lignes de texte. Les lignes de base sont extraites en regroupant les points minima voisins du squelette des composantes connexes. Pour trouver ces points minima, un algorithme d'analyse des codes des segments successifs (dans le codage de Freeman ou chain code) est utilisé sur le squelette de chaque composante connexe. Ensuite, la ligne de centre de l'écriture est estimée en utilisant l'orientation de la ligne de base et les points terminaux dans le skeleton (voir figure 2.6.a). Ces aligne-

ments sont ensuite regroupées selon divers critères comme la taille, la distance entre les lignes et selon leurs orientations. Une étape de correction des conflits entre les alignements est appliquée pour améliorer le taux de segmentation (voir figure 2.6.b). Cette méthode a été appliquée sur 7 documents contenant 61 paragraphes appartenant à deux documents de l'année 1649 (2 documents), et à 5 documents des années 1727, 1741, 1812, 1817 et 1838. Le taux de bonne segmentation varie entre 90% et 97% selon les valeurs des critères utilisés.

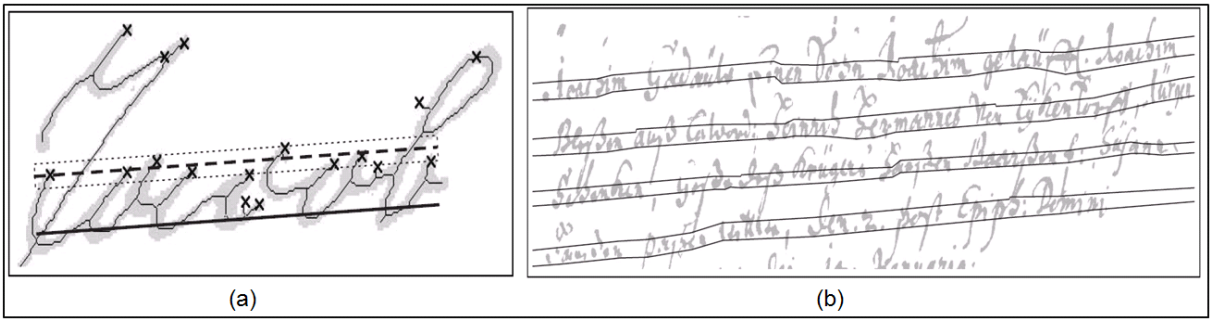


FIGURE 2.6: Étapes d'extraction de lignes : (a) détection des points terminaux (en étoile), de la ligne de base (en trait continu) et de la ligne du centre (en trait interrompu) et (b) résultat final d'extraction de lignes [Feldbach et Tönnies, 2001].

2.2.2 Transformée de Hough

La transformée de Hough est très utilisée pour l'extraction des lignes dans les images [Duda et Hart, 1972]. Un ensemble de points (voir figures 2.7.a et 2.8.a) qui forment une ligne dans l'espace réel (i.e. image) correspond à une zone d'accumulation dans l'espace de Hough (voir figures 2.7.b et 2.8.b). Dans cet espace, chaque point (x_i,y_i) de la ligne est exprimée par l'équation paramétrique suivante :

$$\rho = x_i cos(\theta) + y_i sin(\theta) \tag{2.1}$$

où θ est l'angle d'inclinaison par rapport à l'axe des abscisses (x) et ρ l'éloignement de la droite par rapport à l'origine du système de coordonnées (x,y).

Chaque point de l'image est projeté dans l'espace des paramètres (ou espace de Hough). Le point (x_i,y_i) est représenté par la sinusoïde d'équation 2.1 qui correspond à toutes les droites auxquelles ce point pourrait appartenir. La propriété de colinéarité des points dans l'espace réel se traduit dans l'espace des paramètres par l'intersection de leur sinusoïde respective.

Likforman-Sulem et al. ont proposé dans [Likforman-Sulem *et al.*, 1995] une approche utilisant la transformée de Hough pour l'extraction des lignes des documents manuscrits latins. Cette méthode utilise comme points votants les centres de gravité des composantes

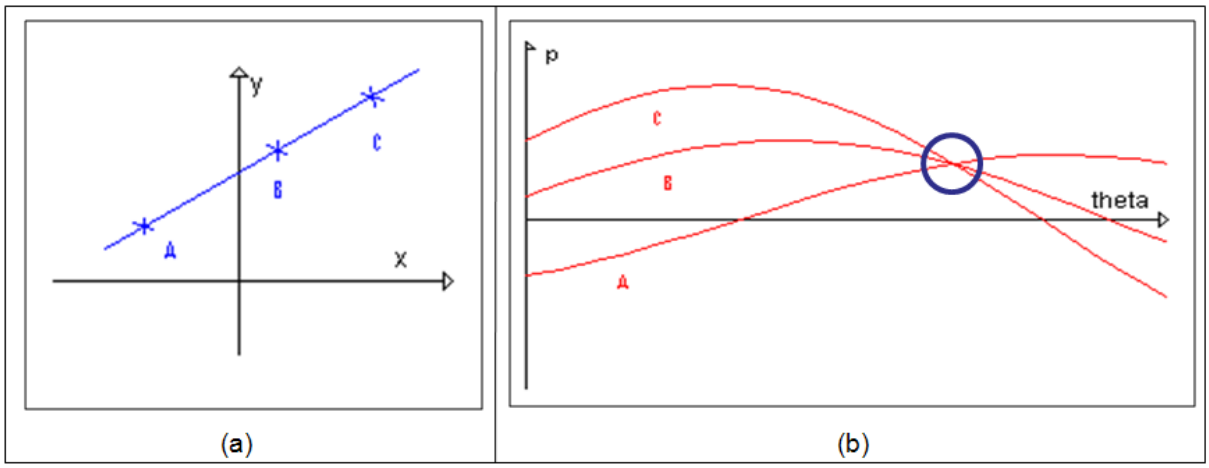


FIGURE 2.7: Illustration de la transformée de Hough pour quelques points d'une ligne droite : (a) ligne originale et (b) sa transformée de Hough (le cercle indique le croisement (un point) des sinusoïdes, source [WIK, 2009]).

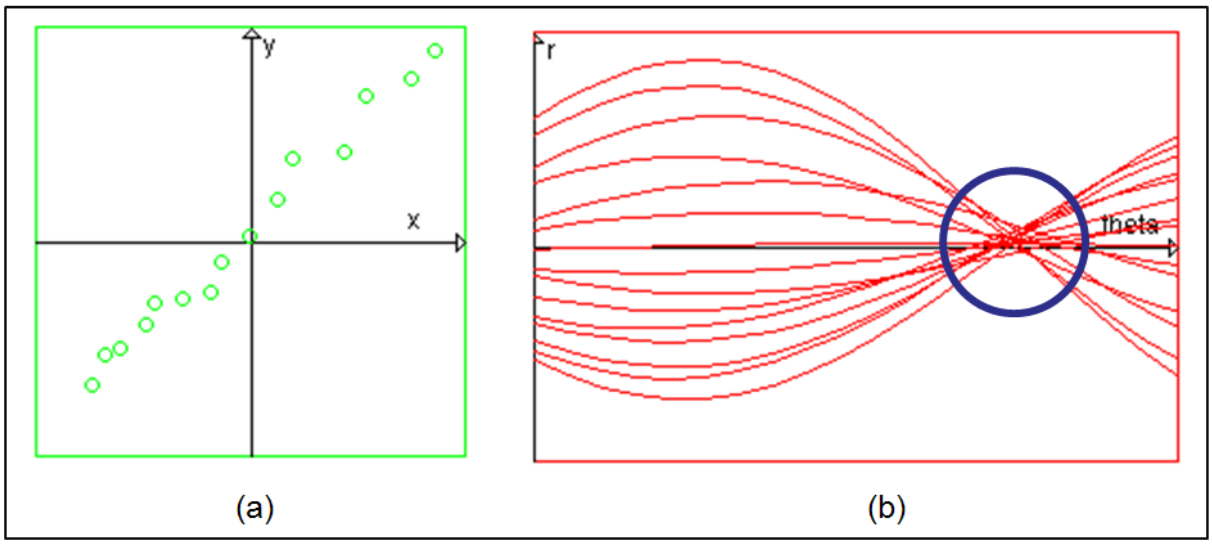


FIGURE 2.8: Illustration de la transformée de Hough pour une ligne non-droite : (a) ligne originale et (b) sa transformée de Hough (le cercle indique le croisement (une zone), source [TSI, 2009])).

connexes. Les alignements (qui sont moins colinéaires que des lignes parfaites) qui ont beaucoup de points votants correspondent à des zones de fortes accumulations dans le domaine de Hough. Ces alignements sont validés en utilisant les informations contextuelles comme les voisines internes et externes. Cette approche a été testée sur plusieurs documents, y compris des brouillons, des blocs d'adresses et des lettres. Les auteurs ne donnent pas de mesure sur la qualité de la segmentation employée. Ils montrent l'efficacité de l'approche par quelques illustrations sur quelques documents.

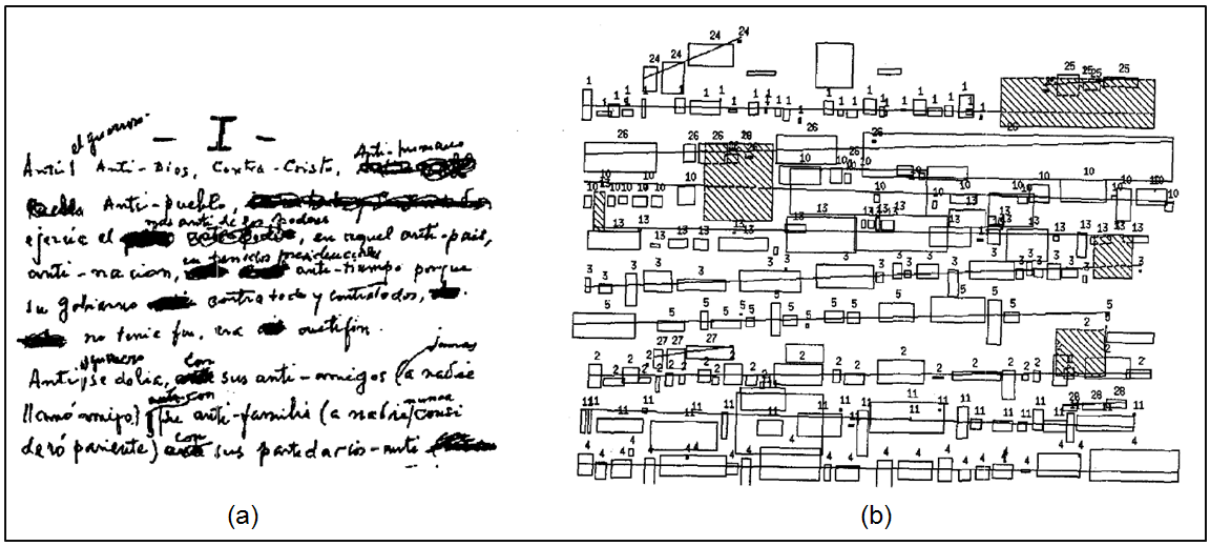


FIGURE 2.9: Étapes d'extraction de lignes : (a) document original et (b) résultat de la segmentation en lignes [Likforman-Sulem *et al.*, 1995].

La transformée de Hough a également été appliquée sur des documents manuscrits par Pu et Shi dans [Pu et Shi, 1998]. On utilise comme points votants les points minima des composantes connexes dans une bande verticale sur la gauche de l'image (voir figure 2.10.a). Les alignements sont recherchés en regroupant les cellules dans l'espace de Hough dans 6 directions (voir figure 2.10.b). Ces alignements correspondent aux successions de composantes connexes qui débutent chaque ligne. Ensuite, une fenêtre glissante est associée à chaque alignement permettant à un algorithme de suivre les composantes connexes restantes en se servant de la proximité des composantes connexes initiales et de l'orientation trouvée par Hough. Cette technique a été testée sur plus de 100 documents manuscrits latins. Les auteurs ne mesurent pas la qualité de l'extraction des lignes. Ils montrent l'efficacité de l'approche par quelques illustrations sur quelques documents (voir figure 2.10.c).

Dans [Malleron *et al.*, 2009], la transformée de Hough est utilisée avec une méthode de regroupement des composantes connexes. Pour cela, les composantes connexes sont extraites puis les contours et les bords de ces composantes sont détectés. L'orientation du document est ensuite estimée en projetant ces bords, 18 orientations ont été étudiées

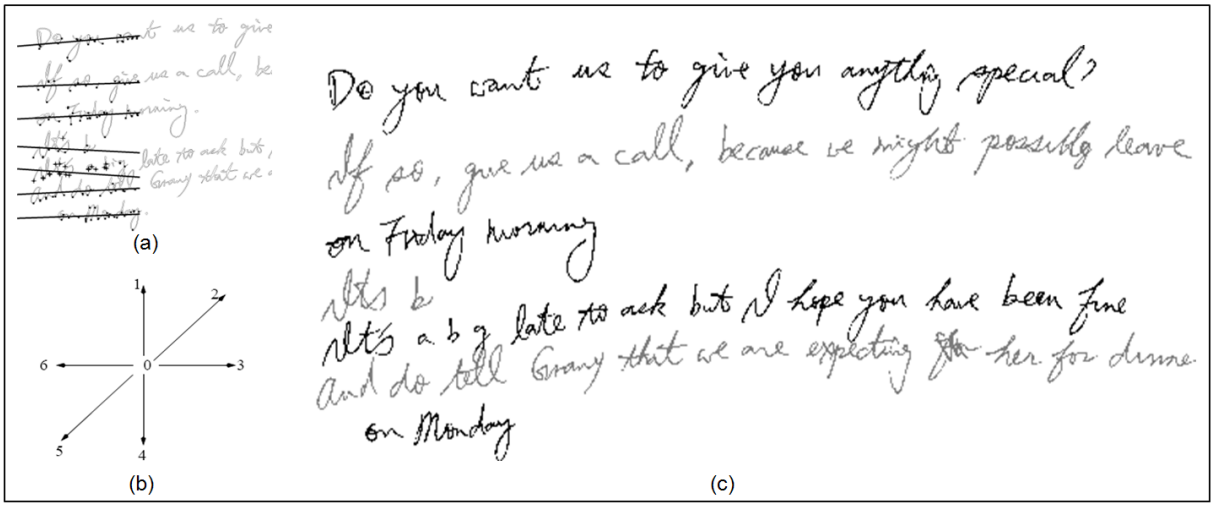


FIGURE 2.10: Exemple d'extraction de lignes : (a) les minima dans la bande verticale à l'extrémité gauche de l'image, (b) les 6 directions et (c) résultat de la segmentation en lignes [Likforman-Sulem *et al.*, 1995].

afin de regrouper les composantes connexes qui ont les mêmes orientations. L'approche a été testée sur 200 documents manuscrits (lettres, textes imprimés, notes, dictionnaires, recueillies). La figure 2.11 montre l'efficacité d'extraction de lignes sur un document contenant des notes. Le taux d'alignement correct est de 88%.

FIGURE 2.11: Exemple d'extraction de lignes sur un document contenant des notes en Latin [Malleron *et al.*, 2009].

Louloudis et al. ont proposé dans [Louloudis *et al.*, 2009a], une technique pour l'extraction des lignes et des mots des anciens manuscrits grecs. La transformée de Hough est appliquée aux composantes connexes en utilisant les centres de gravité de leurs rectangles englobants comme points votants. Ces rectangles sont estimés en calculant la taille moyenne des caractères dans le document. Le système proposé a été testé sur la base de documents d'ICDAR2007 qui se divise en 80 documents contenant 1773 lignes et en 40 anciens documents manuscrits contenant 1095 lignes (voir figure 2.12). Le taux d'extraction de lignes est de 97%.

FIGURE 2.12: Exemple d'extraction de lignes sur un document manuscrit Latin (chaque couleur représente une ligne) [Louloudis *et al.*, 2009a].

2.2.3 Technique de lissage

La technique de lissage (RLS [7]), consiste à noircir les petits espaces entre les pixels noirs consécutifs sur la direction horizontale, ce qui aura pour résultat de les connecter. Les boîtes qui englobent les composantes connexes successives dans l'image lissée forment les lignes.

Dans [Shi et Govindaraju, 2004], les auteurs ont proposé une méthode en utilisant l'algorithme de "fuzzy run length". Ce lissage crée des alignements qui vont être analysés selon plusieurs heuristiques pour former les lignes de texte. L'alignement dont la taille n'excède pas une valeur prédéfinie est supprimé. Cette méthode a été testée sur 1864

7. Run-Length Smoothing.

documents représentant des anciens manuscrits latins et des reçus de colis USPS [8]. Le taux d'extraction de lignes atteint 93%.

Dans [Bourgeois, 1997, Bourgeois *et al.*, 2001], les auteurs ont proposé une méthode pour l'extraction des lignes. Cette méthode est appliquée à 20 livres imprimés en niveaux de gris du XVIe siècle. Les lignes sont extraites en appliquant un lissage de type RLSA [9] adapté à une image en niveaux de gris. Au lieu de connecter une série de pixels noirs et blancs, le gradient de l'image est dilaté selon la direction horizontale avec un angle d'inclinaison qui varie entre $\pm 30^{\circ}$. Le processus peut être appliqué dans plusieurs orientations si nécessaire. Le taux de segmentation atteint est égal à 95%.

2.2.4 Réseau répulsif-attractif

Oztop et al. ont proposé dans [Oztop *et al.*, 1999], une approche fondée sur le réseau répulsif-attractif. Cette technique travaille directement sur les images en niveaux de gris. Au début, les lignes de base sont construites une à une du haut de l'image vers le bas en regroupant les ordonnées des pixels voisins. Ces lignes sont les forces répulsives et les pixels sont les forces attractives pour ce réseau (réseau des forces répulsives-attractives). Pour extraire les lignes, les relations entre les pixels et les lignes de base sont étudiées en passant les forces dans ce réseau. Les pixels qui forment la vraie ligne de texte sont gardés, les autres pixels sont supprimés. La technique a été testée sur plusieurs documents anciens arabes (de type Ottman) et latins. Les auteurs ne donnent pas le taux d'extraction. Ils se contentent de monter les résultats d'extraction de lignes sur quelques documents (voir figure 2.13).

2.2.5 Regroupement par l'arbre recouvrant minimal

Yin et Liu dans [F. Yin, 2008] ont adapté l'algorithme d'agglomération de Kruskal à la recherche de lignes de textes. A partir d'une distance définie pour mieux caractériser les CCs proches (2 CCs successives sur la même ligne seront considérées plus proches que 2 CCs sur des lignes différentes), l'algorithme de Kruskal réunit deux à deux les paires de CCs les plus proches. Ensuite, cette distance est appliquée sur les paires de noeuds proches et ainsi de suite jusqu'à rassembler tous les noeuds (voir figure 2.14). En coupant cet arbre à une hauteur estimée représenter un éloignement entre des lignes successives, les regroupements obtenus correspondent aux lignes de texte (voir figure 2.14). Cette approche a été appliquée sur plusieurs documents manuscrits chinois comportant 8664 lignes de texte de la base des données HIT-MW [Su *et al.*, 2007]. Le taux de segmentation en lignes est égal à 95.02%.

Dans [Nicolas *et al.*, 2004], une autre approche a été proposée pour les documents

8. United States Postal Service.
9. Run Length Smoothing Algorithm.

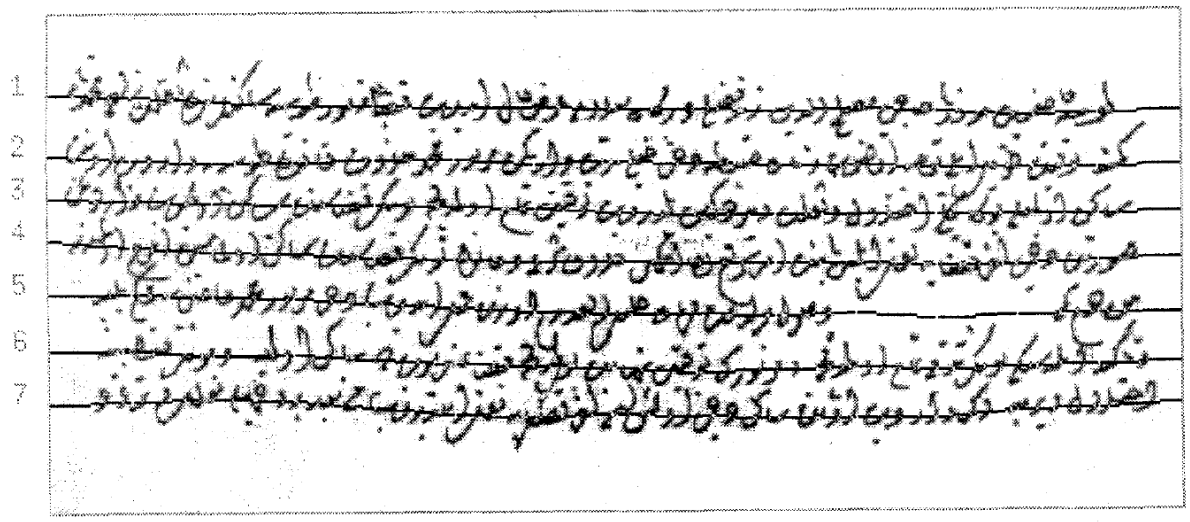


FIGURE 2.13: Extraction de lignes d'un document arabe (de type Ottman) [Oztop *et al.*, 1999].

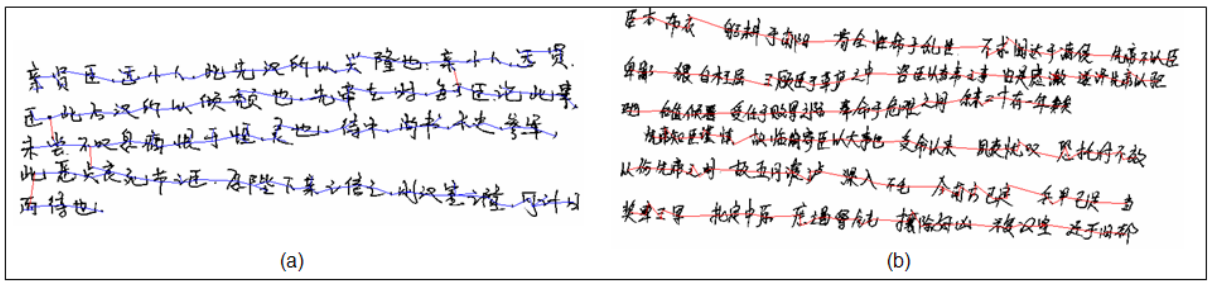


FIGURE 2.14: Étapes d'extraction de lignes sur des documents Chinois : (a) résultat de l'arbre recouvrant minimal et (b) résultat de l'extraction des lignes [F. Yin, 2008].

manuscrits latins de Gustave FLAUBERT [10]. Cette approche est basée sur l'intelligence artificielle en utilisant un système de production. Ce système consiste à parcourir un graphe (arbre) pour trouver les chemins qui représentent les lignes de texte. Les noeuds dans ce graphe sont les composantes connexes et la relation entre les noeuds est la distance euclidienne (voir figure 2.15). Chaque noeud a une probabilité qui est mesurée par rapport à la distance entre les composantes connexes. En passant ce graphe dans le système de production, les auteurs peuvent trouver à la fin, les chemins (lignes de texte) les plus convenables qui contiennent les composantes connexes voisines (voir figure 2.15). Le taux d'extraction n'est pas mentionné.

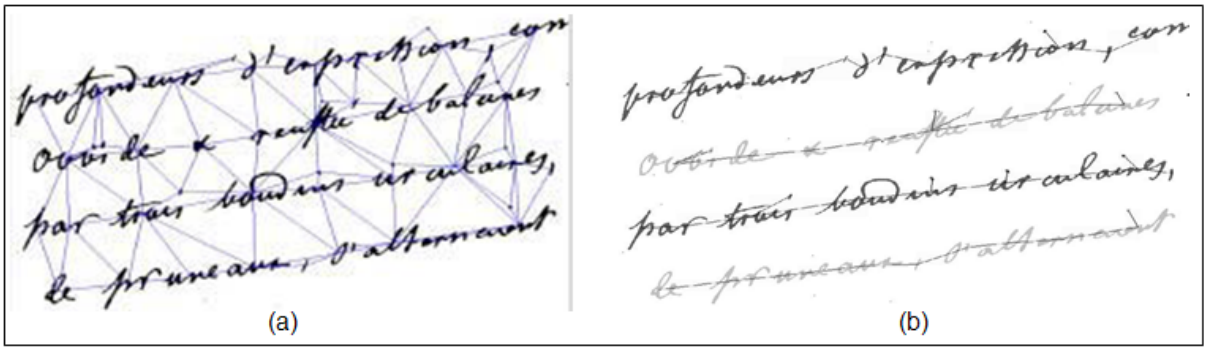


FIGURE 2.15: Étapes d'extraction de lignes sur un document Latin : (a) résultat de graphe et (b) résultat de l'extraction des lignes [Nicolas *et al.*, 2004].

2.2.6 Regroupement par la déformation de contour localisé et propagation

Les modèles des contours déformables sont utilisés dans plusieurs domaines de traitement d'image comme l'analyse des images médicales [Mcinerney et Terzopoulos, 1996]. Ces modèles sont basés sur l'évolution des contours des images. Pour extraire une zone (paragraphe, ligne de texte, région, etc.) d'une image, on pose une forme (contour de départ) sur l'image, puis on applique le mécanisme des contours déformables pour faire évoluer cette forme afin qu'elle trouve la zone cible. Les mécanismes d'évolution sont des fonctions d'énergie. La zone cible sera trouvée en minimisant ces fonctions d'énergie.

Plusieurs modèles de contours déformables existent dans la littérature :

- Le modèle de contour actif (snake) paramétrique [Kass *et al.*, 1987].
- Le snake géométrique [Caselles *et al.*, 1995].
- La méthode de Level set [Pluempitiwiriyawej *et al.*, 2005, Osher et Paragios, 2003, Sethian, 1985].

10. Gustave Flaubert, né à Rouen le 12 décembre 1821 et mort à Canteleu, au hameau de Croisset, le 8 mai 1880, est un écrivain français [WIK, 2009].

– Le B-spline ou le B-snake [Leitner et Cinquin, 1993].
– Le modèle de Mumford-Shah [Ramlau et Ring, 2007].

L'utilisation de chaque modèle dépend de l'application étudiée et du type d'images traitées. Par exemple, le snake géométrique a été beaucoup utilisé pour la segmentation d'images médicales complexes.

Dans le domaine de l'extraction des lignes des documents manuscrits, on a utilisé les trois modèles suivants : le snake paramétré, la méthode de Level set et le modèle de Mumford-Shah.

2.2.6.1 Le snake paramétré

Le snake est un ensemble de points d'un contour $v(s) = (x(s), y(s))$, où $x(s), y(s)$ sont les coordonnées de ces points et $s \in [0, 1]$ [Kass *et al.*, 1988]. s est une approximation linéaire entre 0 et 1 des coordonnées des points. L'approximation s représente l'abscisse curviligne, permettant de paramétrer la longueur de l'arc dans un intervalle très petit. Le contour final peut être obtenu en minimisant la fonction d'énergie suivante :

$$\begin{aligned} E^*_{\text{snake}} &= \int_0^1 E_{\text{snake}}(v(s))\, ds \\ &= \int_0^1 \left[E_{\text{int}}(v(s)) + E_{\text{image}}(v(s)) + E_{\text{con}}(v(s))\right]\, ds \end{aligned} \tag{2.2}$$

L'énergie du snake se décompose en trois types :
– E_{int} (énergie interne) qui permet au snake de se rétrécir ou de s'allonger.
– E_{image} (énergie de l'image ou énergie extérieure) qui pousse le snake vers les lignes et les contours des objets qui se trouvent dans l'image (i. e. gradient)
– E_{con} (énergie du contexte) qui exprime certaines contraintes supplémentaires qui peuvent être imposées par l'utilisateur vu le snake qu'il veut obtenir.

Bukhari et al. ont proposé dans [Bukhari *et al.*, 2009] une méthode de snake paramétré pour l'extraction des lignes de documents manuscrits. Le contour ou snake de départ est la ligne centrale des lignes du texte (voir figure 2.16.b). Cette ligne est estimée en se basant sur les valeurs de l'intensité dans une image de document en niveaux gris. Ensuite, le mécanisme de minimisation d'énergie de snake est appliqué pour trouver les lignes du texte (voir figure 2.16.c et d). Le système proposé a été testé sur la base de documents d'ICDAR2007 qui se divise en 80 documents contenant 1770 lignes. Le taux d'extraction de lignes est de 96.3%.

2.2.6.2 Le Level set

[Sumengen, 2004] Dans la méthode de lignes des niveaux (Level set), le contour se déforme selon sa dérivée partielle et un champ de vecteur externe (gradient). La formule de ce mécanisme s'écrit de la manière suivante :

$$\frac{\partial f}{\partial t} + S_N \nabla f = b_k \left|\nabla f\right| \tag{2.3}$$

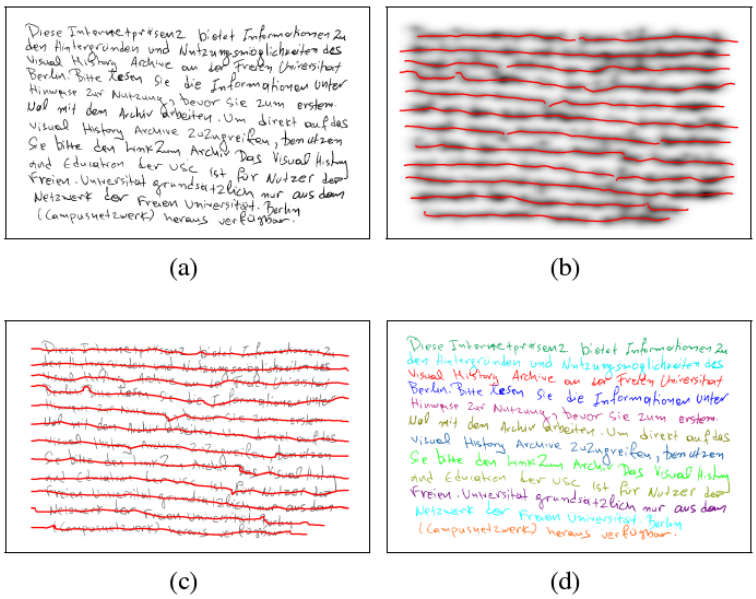


FIGURE 2.16: Étapes d'extraction de lignes : (a) image initial, (b) snake de départ (en rouge) sur l'image initial lissée en appliquant un filtre gaussien, (c) snake déformé et (d) résultat de l'extraction de lignes [Bukhari *et al.*, 2009].

où f est une fonction implicite qui représente les contours de l'image au niveau zéro, t désigne le temps, $\frac{\partial f}{\partial t}$ représente le mouvement de la fonction f dans l'image. Ce mouvement est contrôlé par la courbure de contour b_k, la vitesse de mouvement S_N et ∇f le gradient de contour.

Dans la méthode de lignes des niveaux, le contour fermé de la composante connexe (ou d'alignement trouvé par un filtre gaussien) est la fonction implicite f qui a des valeurs négatives à l'intérieur, des valeurs positives à l'extérieur et des valeurs nulles au contour. La fonction f est déformée selon l'équation 2.3. Grâce à la dérivée partielle et aux paramètres b_k et S_N, les composantes connexes (ou alignements) qui sont dans la même direction vont être regroupées pour former les vrais lignes de texte.

Li et al. ont proposé dans [Li *et al.*, 2008] une telle technique pour l'extraction des lignes de documents manuscrits complexes. La méthode démarre par l'application d'un filtre gaussien avec un noyau anisotropique pour estimer la densité des pixels (voir figure 2.17.b). Ensuite, une carte de probabilité est construite donnant la probabilité qu'un pixel appartienne à une ligne du texte (voir figure 2.17.c). Ensuite, le modèle de Level set est appliqué pour déformer le contour afin de trouver les lignes (voir figure 2.17.d). Enfin, deux étapes de post-traitement sont appliquées pour regrouper les lignes de texte voisines qui ont une distance faible entre eux (voir figure 2.17.e) et pour couper les lignes qui se connectent (voir figure 2.17.f). Cette méthode a été testée sur 2691 lignes de texte arabe, 1672 lignes de texte chinois, 1365 lignes de texte indien et 1932 lignes de texte coréen. Les taux d'extraction de lignes sont 84.6% (textes arabes), 92% (textes chinois), 95% (textes

indiens) et 96% (textes coréens).

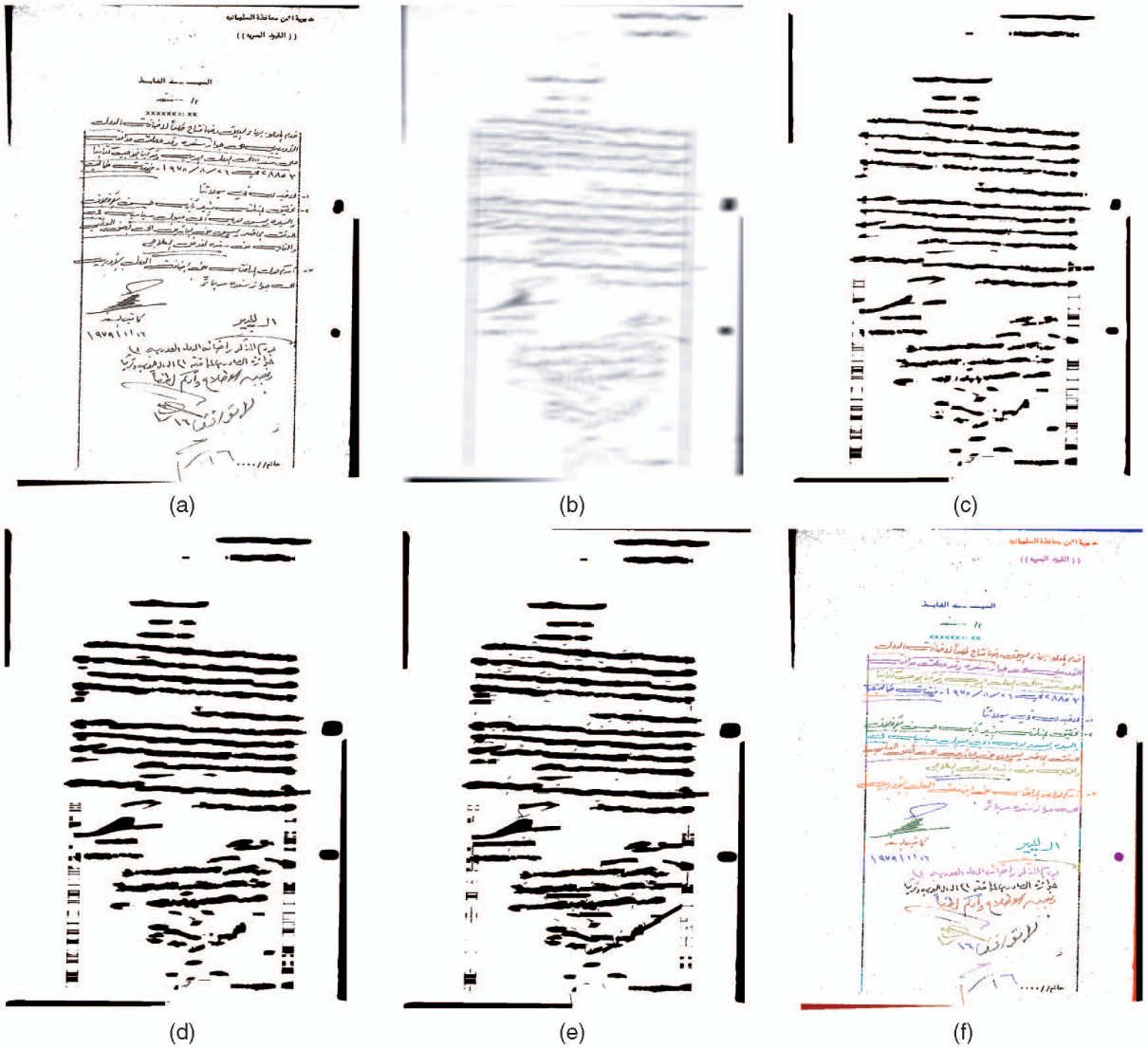


FIGURE 2.17: Étapes d'extraction de lignes pour un document Arabe : (a) image initiale, (b) estimation de la densité des pixels en appliquant un filtre gaussien, (c) estimation initiale des lignes de texte, (d) résultat après 10 itérations par la méthode des lignes de niveaux, (e) résultat après la connexion des lignes de texte voisines et (f) résultat final de l'extraction des lignes [Li *et al.*, 2008].

2.2.6.3 Le Mumford-Shah

Le modèle de Mumford-Shah [Mumford et Shah, 1989] déforme les contours en minimisant la fonction d'énergie suivante :

$$E(u, C) = \int_{\Omega} |u - u_0|^2 \, dxdy + \mu \int_{\Omega \backslash C} |\nabla u|^2 \, dxdy + v. |C| \tag{2.4}$$

où u_0 est l'image initiale, u l'image lissée après l'application d'un filtre gaussien, C l'ensemble des courbes (ou snakes) de départ, $|C|$ l'ensemble des longueurs des courbes, Ω le domaine de l'image, $\Omega \backslash C$ le domaine de l'image sans l'ensemble des courbes, ∇u le gradient de l'image u et μ et v les paramètres qui équilibrent l'effet des autres termes dans cette équation.

Du et al. ont proposé une telle méthode dans [Du *et al.*, 2009] pour l'extraction des lignes de documents manuscrits latins. L'extraction des contours de départ est similaire dans [Li *et al.*, 2008]. L'extraction des lignes se fait par l'application de modèle de Mumford-Shah. Le système proposé a été testé sur 100 documents chinois, 96 documents indiens et 100 documents coréens. Les taux d'extraction de lignes sont de 98% pour les documents chinois, 98% pour les documents indiens et 96% pour les documents coréens. La figure 2.18 montre les résultas de trois documents (Chinois, Indien et Coréen).

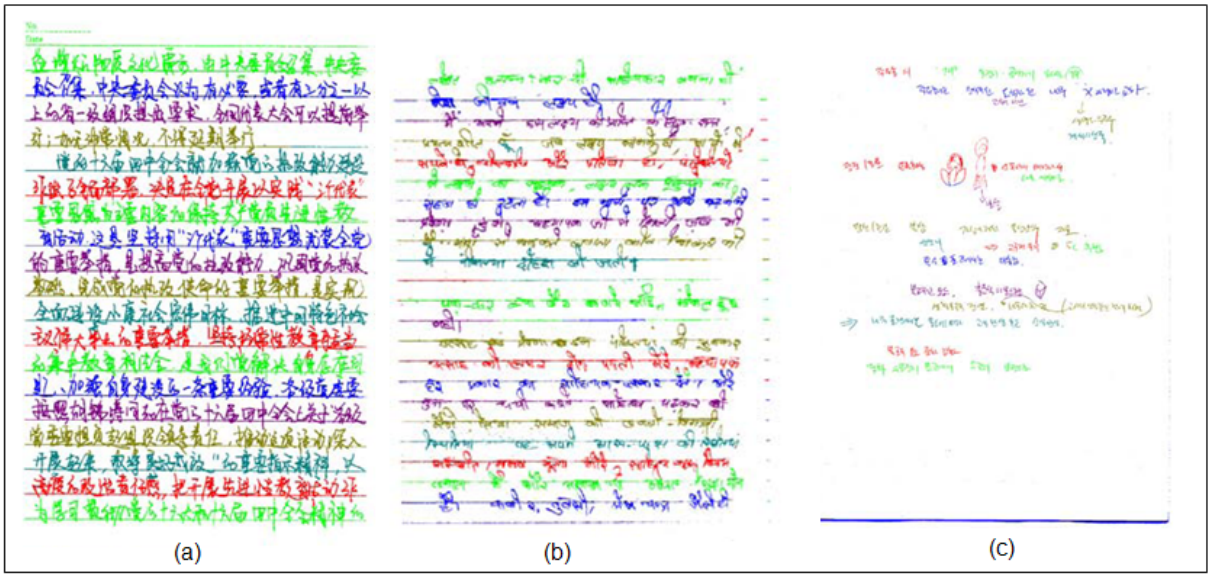


FIGURE 2.18: Exemple d'extraction de lignes pour trois documents (a) Chinois, (b) Indien et (c) Coréen [Du *et al.*, 2009].

2.3 Discussion des méthodes

Le tableau 2.1, résume toutes les méthodes évoquées, réparties selon 15 critères : types des lignes (droites, orientées et cursives), types des documents (imprimés, manuscrits, multi-orientés, intervalle d'orientation (I. O.), latins, chinois, indiens, arabes, persans, urdus, nature du contenu (C : Couleur, G : Gris et B : Binaire) et maillage.

Toutes ces approches sont soit trop globales, procédant par projection ou par recherche d'alignements, soit trop locales, opérant par suivi de composantes connexes. Elles trouvent ici leurs limites face à la mauvaise qualité et à la multi-orientation des documents. La plupart de ces méthodes ont été appliquées sur des documents ayant une seule orientation. L'adaptation de ces approches est impossible si on veut extraire toutes les orientations.

En effet, la technique de projection, par exemple, ne peut pas être appliquée parce qu'elle demande des paramètres comme l'orientation de l'écriture et elle est sensible lorsque les documents présentent du bruit, ce qui est le cas des anciens documents arabes. Ce bruit peut créer des faux maxima dans le profil de projection ce qui implique un suivi des lignes erroné.

Les méthodes guidées par le modèle de document sont applicables pour des documents imprimés. Elles trouvent ici leurs limites à cause la multi-orientation et l'écriture manuscrite.

De même, la technique de classification par k-ppv nécessite des paramètres comme la proximité et l'orientation des composantes connexes qui ne sont pas faciles à extraire dans les documents anciens, du fait de la présence d'orientations multiples et du bruit.

L'adaptation de la transformée de Hough pour des documents multi-orientés est très difficile à cause de la multiplicité des orientations, de la sinuosité des lignes et donc de la non régularité des orientations. L'extraction des points votants est très difficile à cause des problèmes de représentativité et de stabilité.

L'application de la méthode de lissage est impossible à utiliser parce qu'elle est conçue pour des documents mono-orientés.

La technique de réseau répulsif-attractif demande une bonne détection de la ligne de base. Cette ligne est l'élément fondamental pour trouver les forces répulsives et attractives parce qu'elle contient le maximum des pixels. Une fausse détection conduit à une fausse extraction des lignes. Dans nos documents, l'extraction des lignes de base est une tâche difficile parce que l'écriture arabe présente beaucoup de fluctuations par rapport à la ligne horizontale.

La méthode de regroupement par l'arbre recouvrant minimal demande des documents avec des espacements réguliers entre les lignes car elle est basée sur l'apprentissage des distances qui doivent être caractéristiques des espacements entre composantes connexes dans les lignes et entre les lignes.

Les méthodes basées sur les modèles déformables sont très avantageuses et très importantes par rapport aux méthodes descendantes et ascendantes. Ces avantages viennent du fait que ces modèles sont indépendants du script traité. Ces modèles ont été appliqués

sur des documents arabes, latins, Chinois et urdus. Ils ne demandent pas beaucoup de connaissances et d'information sur le document (comme la taille des caractères, la géographie des composantes connexes, etc.). De même, ces méthodes peuvent être utilisées pour des écritures cursives grâce au pouvoir de déformation du contour. Malgré ces avantages, il est très difficile d'appliquer ces modèles sur les documents manuscrits arabes multi-orientés, parce qu'ils demandent beaucoup de paramètres comme le nombre de dilatation et d'érosion, le noyau du filtre gaussien et le nombre d'itérations. De même, il est très difficile de définir le contour et de l'initialiser près des lignes de texte recherchées. Ces modèles peuvent convenir pour des angles d'orientation faibles mais pas pour toutes les orientations comme dans le cas de nos documents. De plus, ces méthodes nécessitent une étape de post-traitement pour séparer les lignes connectées.

2.4 Conclusion

Nous avons présenté dans ce chapitre un grand nombre de méthodes permettant de segmenter les documents manuscrits. Parmi ces méthodes, certaines peuvent être utilisées pour la segmentation de documents arabes manuscrits comme la technique de projection et la technique de la classification par k-ppv. D'autres techniques comme la méthode basée sur la transformé du Hough, la technique de lissage, la méthode du réseau répulsif-attractif et la méthode basée sur le regroupement par l'arbre recouvrant minimal nécessitent une adaptation importante pour être pouvoir être appliquées à nos documents.

Classe	Catégorie	Auteur	Type des lignes			Nature du contenu											Maillage
			Droit	Orienté	Cursive	Imprimé	Manuscrit	Multi-orienté	I. O.	Latin	Chinois	Indien	Arabe	Persan	Urdu	Niveau	
Descendant	Projection	[Nicolaou et Gatos, 2009]	x	x			x			x						G,B	
		[Bennasri *et al.*, 1999]	x	x			x						x			B	x
		[Shapiro *et al.*, 1993]	x	x			x			x						B	
		[Antonacopoulos et Karatzas, 2004]	x			x				x						B	
	Modèle de document	[Nagy *et al.*, 1988]	x			x				x						B	x
		[Lu *et al.*, 2004]	x	x		x				x						B	x
Ascendant	Classification par k-ppv	[Likforman-Sulem et Faure, 1994]	x	x	x		x			x						G,B	
		[Zahour *et al.*, 2004]	x	x	x		x						x			B	
		[Zahour *et al.*, 2007]	x	x	x		x						x			B	
		[Zahour *et al.*, 2008]	x	x	x		x						x			B	
		[Feldbach et Tönnies, 2001]	x	x			x			x						G	
	Transformée de Hough	[Likforman-Sulem *et al.*, 1995]	x	x			x			x						B	
		[Pu et Shi, 1998]	x	x	x		x			x						B	
		[Malleron *et al.*, 2009]	x	x		x	x		$\pm 45^{\circ}$	x						B	
		[Louloudis *et al.*, 2009a]	x	x	x	x	x			x						G,B	
	Lissage	[Shi et Govindaraju, 2004]	x	x			x									B	
		[Bourgeois *et al.*, 2001]	x	x		x	x									G	
	Répulsif-attractif	[Oztop *et al.*, 1999]	x	x	x		x			x			x	x		G	
	Arbre couvrant minimal	[F. Yin, 2008]	x	x			x				x					B	
		[Nicolas *et al.*, 2004]	x	x			x			x						B	
	Snake paramétré	[Bukhari *et al.*, 2009]	x	x	x	x	x	x		x			x		x	G,B	
	Level set	[Li *et al.*, 2008]	x	x	x	x	x	x		x	x	x	x			B	
	Mumford-Shah	[Du *et al.*, 2009]	x	x	x		x	x		x						G	

TABLEAU 2.1: Les méthodes proposées dans la littérature pour la segmentation des lignes (I. O. : Interval d'Orientation, G : Gris, B : Binaires et C : Couleurs).

Deuxième partie

Extraction et séparation de lignes multi-orientées

Architecture du système proposé

Le système proposé se divise en trois niveaux (voir figure 19) :
- Niveau 1 : détection des zones multi-orientées.
- Niveau 2 : extraction de lignes.
- Niveau 3 : séparation des lignes connectées.

Dans le premier niveau, les zones multi-orientées sont extraites du document en appliquant un maillage automatique, puis une méthode de regroupement des mailles voisines est opérée pour réunir les mailles de même orientation. L'orientation dans chaque maille est estimée en appliquant une méthode de distribution d'énergie issue de la classe de Cohen. Plusieurs étapes de correction de l'orientation et du maillage sont appliquées afin de corriger les erreurs produites par ces méthodes.

Dans le second niveau, les lignes dans chaque zone sont détectées en se basant sur l'orientation de la zone. Une étape de correction du suivi est opérée en vue d'affecter les symboles diacritiques à leurs lignes.

Dans le troisième niveau, les lignes connectées sont séparées par analyse de la morphologie des lettres terminales arabes qui sont à l'origine de ces connexions.

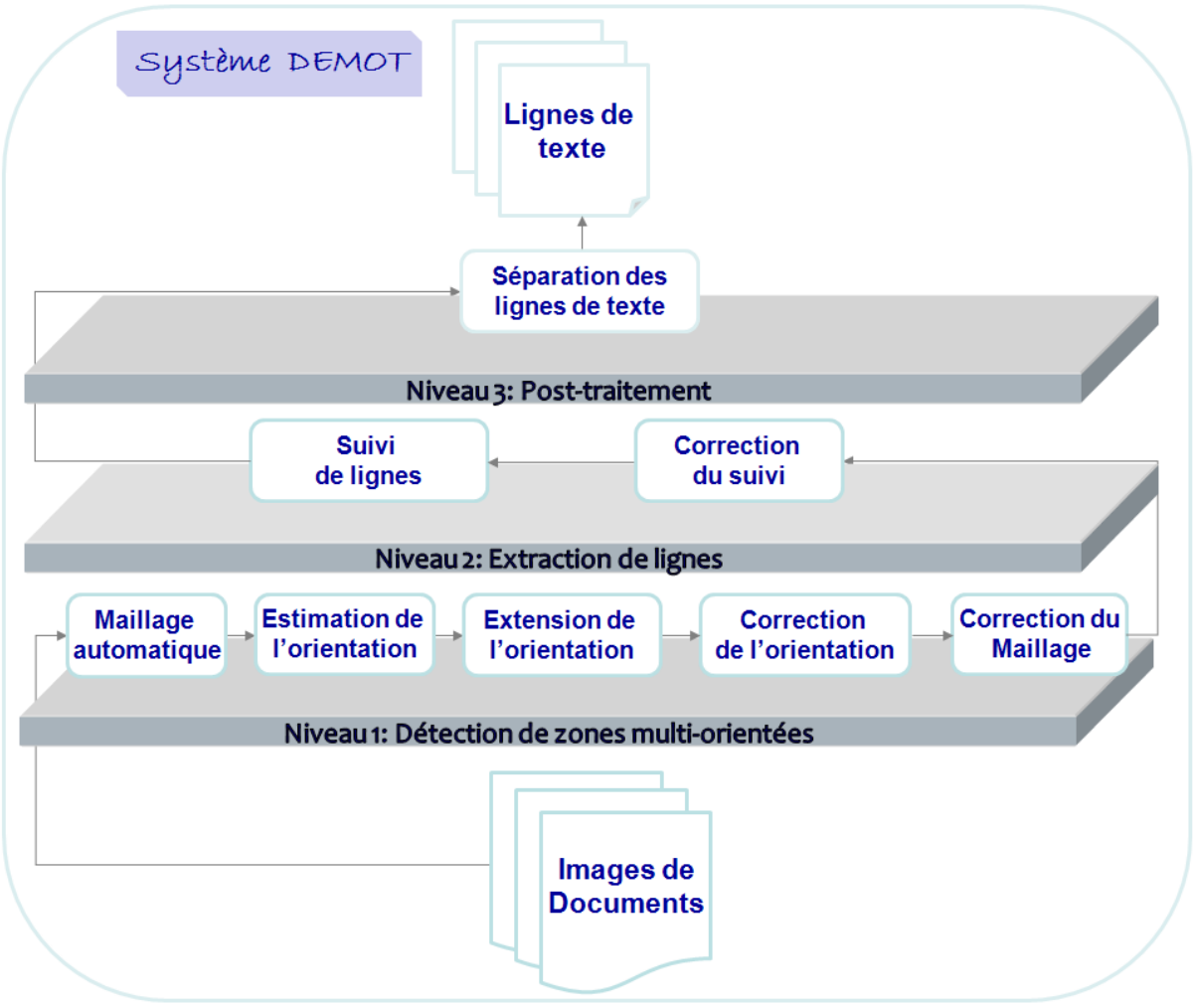


FIGURE 19: Architecture du système $DEMOT$.

Chapitre 1

Maillage automatique

Dans ce chapitre, nous présentons l'algorithme de maillage automatique dans le but de diviser l'image en plusieurs fenêtres rectangulaires. Tout d'abord, nous commençons par l'application du maillage automatique, puis, nous expliquons le modèle de contour actif (snake) utilisé pour définir ce maillage. Enfin, nous montrerons nos résultats et nous terminerons le chapitre par une conclusion qui établira un bilan de cette technique.

Sommaire

1.1 Introduction 35
1.2 Application du maillage 36
1.3 Modèle de contour actif (*snake*) 38
1.4 Sélection des fenêtres 41
1.5 Expérimentations 45
1.6 Conclusion 46

1.1 Introduction

Un maillage est une partition d'une zone géographique en plusieurs unités contiguës dont la forme et la taille peuvent être régulières (voir figure 1.1.a), irrégulières (voir figure 1.1.b) ou semi-régulières (voir figure 1.1.c) [Brunet *et al.*, 1992]. Les mailles peuvent avoir la forme d'un triangle, un rectangle, un cercle, un carré, etc. et peuvent être appliquées en 2D ou en 3D. Le choix dépend de l'application visée. Le maillage est utilisé dans plusieurs domaines afin de construire des réseaux. On trouve le maillage dans les SIG[11] par exemple, en vue de construire des cartes et de créer des limites géographiques.

Dans les domaines du traitement d'image et de l'analyse de documents, le maillage est souvent utilisé pour des besoins de segmentation en régions ou de localisation d'objets

11. Système d'Information Géographique.

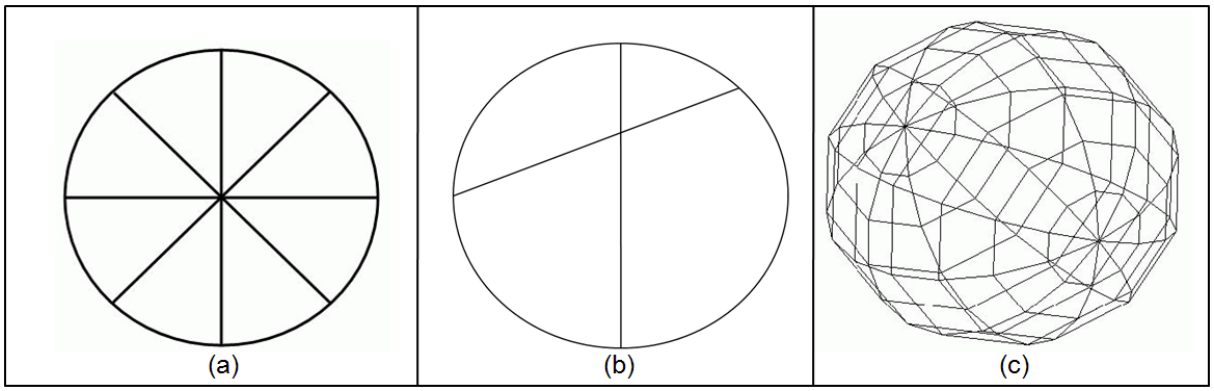


FIGURE 1.1: Les 3 types de maillage, (a) régulier, (b) irrégulier et (c) semi-régulier.

particuliers. Par exemple dans les manuscrits traités, le maillage est utilisé afin de segmenter le document en plusieurs zones orientées. Dans la littérature, plusieurs approches ont été proposées afin de détecter les lignes dans les documents orientés en utilisant le maillage. Zahour et al. ont proposé dans [Zahour *et al.*, 2008] une approche pour l'extraction des lignes connectées dans les documents manuscrits arabes. L'approche proposée est basée sur un maillage qui divise le document en plusieurs colonnes (strips). Ces colonnes de faible largeur permettent aux auteurs d'opérer des projections horizontales précises en vue de détecter, à l'aide des maxima, les lignes de base locales. Dans [Pal et Datta, 2003], Pal et Datta ont proposé une approche pour l'extraction des lignes de documents banglas. Un maillage régulier divise le document en plusieurs colonnes. Ensuite, les profils de projection de ces colonnes sont utilisés pour extraire les lignes de base. Ce type de maillage en colonnes serait inutile dans le cas de nos documents multi-orientés car les colonnes ne révéleront pas d'orientation homogène. Pour cela, nous nous concentrons sur un maillage plus petit et rectangulaire qui divise l'image du document en plusieurs pavés de taille identique mais adaptative afin de donner à toutes les orientations la chance d'être extraites et étendues à des zones plus larges. Ci-après, nous allons expliquer en détail le maillage automatique opéré dans $DEMOT$.

1.2 Application du maillage

Le document est partitionné en petites fenêtres de taille d'abord régulière ($w \times h$). Cette taille est automatiquement générée en se basant sur l'idée qu'une fenêtre doit contenir environ 3 lignes pour être capable de produire un histogramme de projection représentatif de l'orientation. Cela donne lieu à plusieurs étapes.

Tout d'abord, une première fenêtre ou maille de taille arbitraire (15×15 pixels) est désignée au milieu de l'image (voir figure 1.3.a). Ensuite, l'approche du snake est appliquée pour extraire des lignes de texte (voir les explications dans la section 1.3). La largeur de la

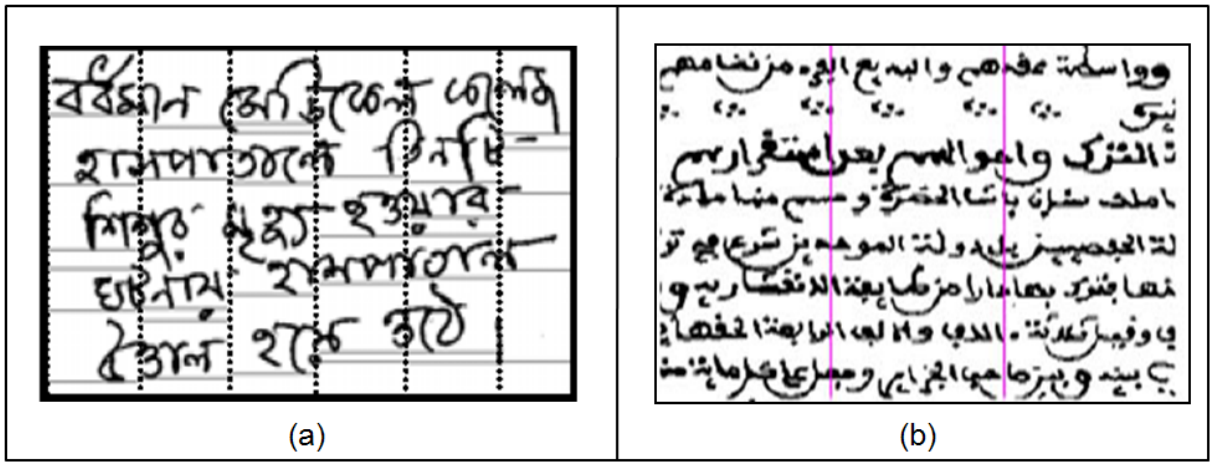


FIGURE 1.2: Maillage par colonne de document : (a) Bangla et (b) Arabe.

fenêtre est agrandie jusqu'à ce que le snake fournisse au moins 3 lignes. Une fois les lignes trouvées, la hauteur moyenne $\bar{h}$ des lignes est estimée ainsi que la distance moyenne $\bar{g}$ entre les lignes. La distance entre les lignes est estimée en utilisant l'écart minimum entre les enveloppes convexes des lignes [Mahadevan et Nagabushnam, 1995] (voir figure 1.3.b). La fenêtre finale a une taille égale à $w \times h$, où $w = h = 3 \times \bar{h} + 2 \times \bar{g}$ (voir figure 1.3.a).

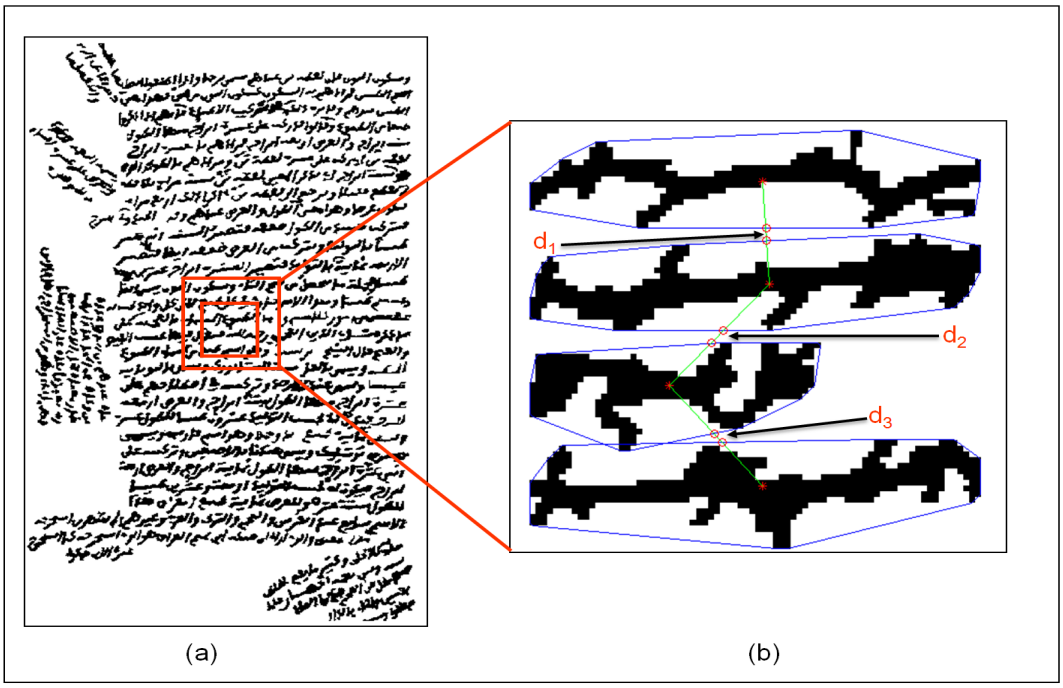


FIGURE 1.3: Algorithme de maillage automatique (d1, d2 et d3 sont les distances entre les lignes).

1.3 Modèle de contour actif (*snake*)

Le modèle de contour actif a été introduit par Kass, Witkin et Terzopolous en 1987 et 1988 [Kass *et al.*, 1987, Kass *et al.*, 1988]. [Rousselle, 2003] L'état de l'art contient de nombreux articles qui parlent du modèle de contour actif comme la revue des méthodes de segmentation d'images médicales [Pham *et al.*, 2000], le livre sur la vision active [Blake et Yuille, 1992], les contributions et applications des modèles déformables en vision par ordinateur [Bascle, 1994], la revue des surfaces déformables [Montagnat *et al.*, 2001]. Le modèle de contour actif a été utilisé dans plusieurs domaines : reconnaissance des formes [Choi *et al.*, 2001], suivi des scènes [Allili et Ziou, 2007], segmentation d'images [Zheng *et al.*, 2007], etc..

Le snake est défini comme un contour virtuel qui va être emmené vers un contour réel dans l'image en utilisant un mécanisme de minimisation d'énergie et une méthode itérative pour la déformation progressive. Par exemple, la figure 1.4.a nous montre le snake de départ qui a été initialisé près de la forme cible. Dans la figure 1.4.b le snake est déformé en appliquant le phénomène du modèle de contour actif et la figure 1.4.c représente le snake final (contour cherché).

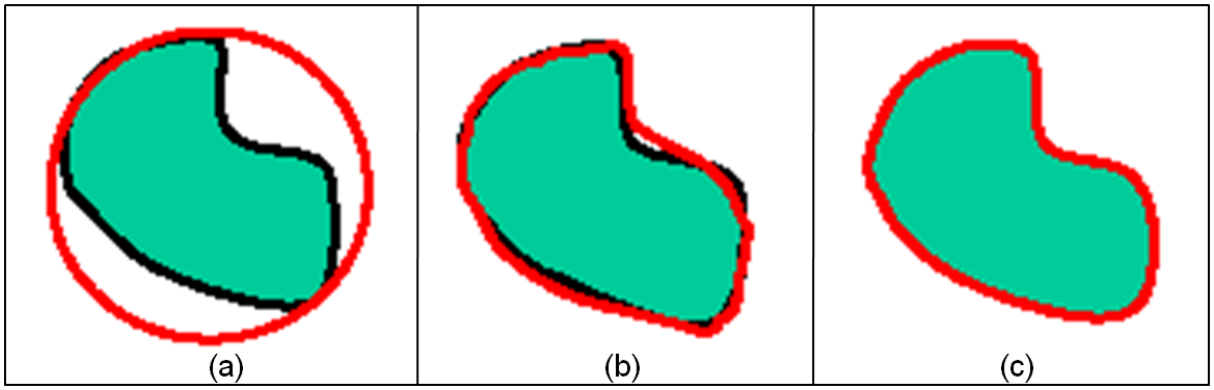


FIGURE 1.4: Modèle de contour actif (*snake*) : (a) position initiale du snake (cercle rouge), (b) snake après quelques itérations de déformation et (c) snake final [sna, 2009].

Le snake est un ensemble de points $v(s) = (x(s), y(s))$, où $x(s), y(s)$ qui sont les coordonnées du contour et $s \in [0, 1]$. s est une approximation linéaire entre 0 et 1 des coordonnées des points. Cette approximation représente l'abscisse curviligne, permettant de paramétrer la longueur de l'arc dans un intervalle unitaire. Le snake peut être ouvert ou fermé selon l'application (voir figure 1.5).

Dans la littérature, deux types de contour actif ont été proposés : contour actif paramétrique [Kass *et al.*, 1987] et contour actif géométrique [Caselles *et al.*, 1995]. Le tableau 1.1 explique la différence entre les deux types selon 5 éléments : les paramètres, la détection (un ou plusieurs objets), la topologie, l'initialisation et l'utilisation.

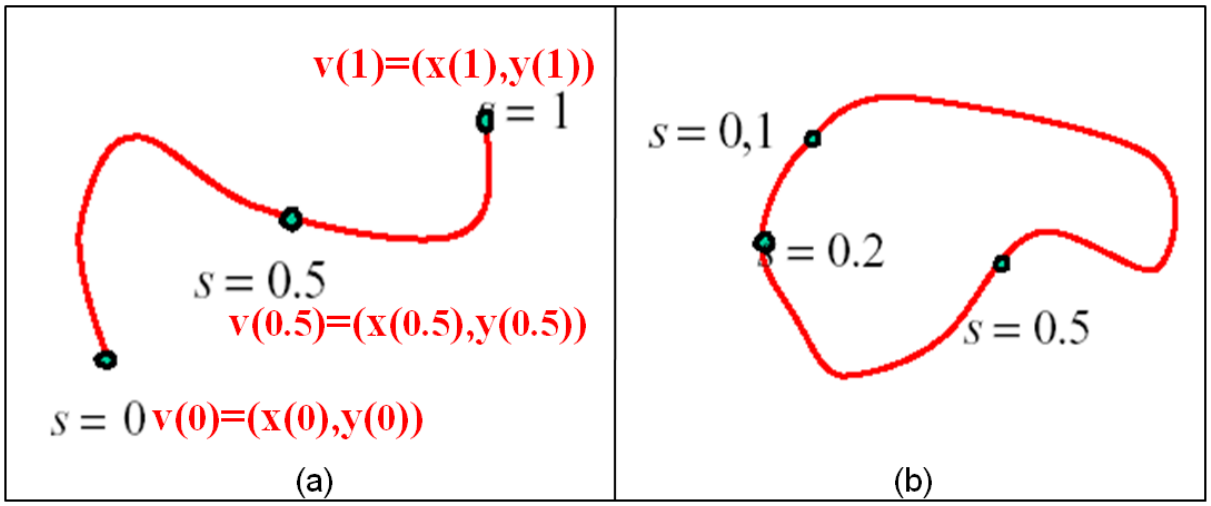


FIGURE 1.5: Types de modèle de contour actif (*snake*) : (a) snake ouvert et (b) snake fermé [sna, 2009].

Contour actif paramétrique	**Contour actif géométrique**
Résultat dépendant de la paramétrisation (e.g. paramètres d'élasticité et de rigidité)	Résultat dépendant des informations géométriques de la courbe (e.g. vitesse de la courbure)
Détection d'un seul objet	Détection simultanée de plusieurs objets
Changement impossible de la topologie	Changement possible de la topologie
Initialisation proche du contour désiré	Initialisation non nécessairement proche du contour désiré
Utilisation pour des contours simples	Utilisation pour des contours complexes (e.g. images médicales)

TABLEAU 1.1: Les principes du modèle de contour actif : paramétrique et géométrique.

Le contour final peut être obtenu en minimisant la fonction d'énergie suivante :

$$\begin{aligned} E^*_{\text{snake}} &= \int_0^1 E_{\text{snake}}(v(s))\, ds \\ &= \int_0^1 \left[E_{\text{int}}(v(s)) + E_{\text{image}}(v(s)) + E_{\text{con}}(v(s))\right]\, ds \end{aligned} \tag{1.1}$$

L'énergie du snake se décompose en trois types d'énergie E_{int} (énergie interne), E_{image} (énergie de l'image ou énergie extérieure) et E_{con} (énergie du contexte) :

- $E_{\text{int}}(v(s)) = \int_0^1 \alpha(s) \left|\frac{dv}{ds}\right|^2 ds + \int_0^1 \beta(s) \left|\frac{d^2v}{ds^2}\right|^2 ds$, où α est le coefficient d'élasticité et β est le coefficient de rigidité. [gvf, 2009] Cette énergie se décompose en deux parties. La première correspond à l'énergie d'élasticité qui permet à la courbe de se rétrécir ou de s'allonger. Remarquons que cette énergie est minimale pour : $\left|\frac{dv}{ds}\right| = 0$; $v(s) = \textit{constante}$, autrement-dit, lorsque tous les points du snake sont situés à la même position. La deuxième correspond à l'énergie de lissage qui est minimale pour $\left|\frac{d^2v}{ds^2}\right| = 0$; i.e. $v(s) = c.s$, où c est une constante, autrement-dit, lorsque les points du snake sont tous alignés. Notons que $v''(s)$ correspond à la courbure du snake au point s.
- E_{image} est l'énergie externe de l'image qui pousse le snake vers les lignes et les contours des objets qui se trouvent dans l'image. L'énergie externe traditionnelle est située sur les contours externes. Cela oblige à initialiser le snake à proximité du contour cible. Par ailleurs, les valeurs du gradient ont des sens inverses sur les deux côtés du même contour, ce qui empêche le snake d'entrer dans les concavités (voir figure 1.6.a). Pour cette raison, Xu et al. [Xu et Prince, 1997] ont proposé un nouveau type d'énergie externe qui permet d'initialiser le snake loin du contour cible et d'aller vers les concavités (voir figure 1.6.b). Cette énergie est nommée flux du vecteur du gradient (GVF : Gradient Vector Flow). Le flux du vecteur du gradient est défini comme un vecteur $V(x, y) = (u(x, y), v(x, y))$ qui minimise la fonction d'énergie suivante en utilisant une méthode itérative :

  $$\varepsilon = \iint \mu(u_x^2 + u_y^2 + v_x^2 + v_y^2) + |\nabla f|^2 |V - \nabla f|^2 dx\, dy \tag{1.2}$$

  où V peut être trouvé par les deux équations d'Euler :

  $$\begin{aligned} \mu\nabla^2 u - (u - f_x)({f_x}^2 + {f_y}^2) &= 0 \\ \mu\nabla^2 v - (v - f_y)({f_x}^2 + {f_y}^2) &= 0 \end{aligned} \tag{1.3}$$

  où μ est un paramètre de réglage de bruit. Pour trouver u et v., Xu et al. proposent dans [Xu et Prince, 1997] plusieurs implémentations numériques.
- E_{con} est l'énergie de contexte qui exprime certaines contraintes supplémentaires qui peuvent être imposées par l'utilisateur vu le snake qu'il veut obtenir.

Dans notre application, nous avons utilisé le modèle de contour actif paramétrique en prenant l'axe majeure de la composante connexe comme le snake de départ (snake ouvert), la GVF comme énergie externe et une énergie de contexte nulle. Le modèle de

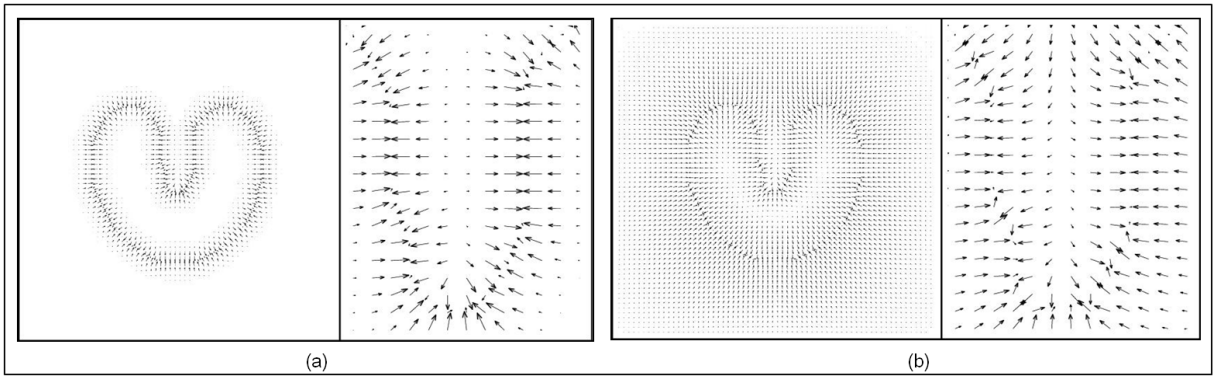


FIGURE 1.6: Énergie externe (indiquée en flèches) pour un objet qui contient des concavités : (a) gradient et (b) flux du vecteur du gradient (GVF) [Xu et Prince, 1997].

contour paramétrique a été choisi parce que son application est simple et parce que nous cherchons un seul contour (contour de la composante connexe) et dans la fenêtre.

Pour détecter les alignements, l'algorithme procède comme suit : tout d'abord, on applique une dilatation et une érosion à l'image initiale en vue d'agrandir le contour (voir figure 1.7.b). Ensuite, on cherche l'axe majeur de chaque composante connexe par une méthode basée sur les descripteurs de Fourier [Maître, 2000, Zahn et Roskies, 1972] (voir figure 1.7.c). Pour assurer que les lignes soient détectées, on incrémente la taille de l'axe majeur par un seuil égal à 1/4 de la largeur moyenne des composantes connexes. Ce seuil est obtenu expérimentalement. Puis, on applique le mécanisme de minimisation des énergies du snake afin de le déformer et de le pousser vers le contour de la composante connexe (voir figure 1.7.d). Enfin, les composantes connexes qui appartiennent à la même ligne sont regroupées (voir figure 1.7.e). L'algorithme 1 décrit ces étapes. La figure 1.8 montre les résultats du maillage automatique des documents présentés dans la figure 1.2. Nous pouvons remarquer dans la figure 1.8 qu'une majorité de fenêtres contiennent au moins trois lignes.

1.4 Sélection des fenêtres

Afin de réduire le temps d'exécution, nous rejetons dans l'étape suivante (estimation de l'orientation) les fenêtres contenant peu de pixels (e.g. vides ou presque vides), parce que leur calcul d'inclinaison sera erroné ou insignifiant. Lorsque ce type de fenêtre contient du texte, on l'attribue à la fenêtre voisine qui partage la plupart de ses composantes. Par exemple, dans la figure 1.9.a, la composante connexe c_1 (resp. c_2) traverse les fenêtres B et D (resp. A, B et C), mais son centre de gravité c_{g1} (resp. c_{g2}) appartient à une fenêtre B (resp. A). Après déplacement des composantes connexes dans les fenêtres A et B, nous créons deux fenêtres vides C et D (voir la figure 1.9.b). Dans la figure 1.10.a, la composante connexe c_1 est déplacée dans la fenêtre B. Ainsi, la fenêtre A devient presque vide (contient du texte avec peu de pixels). Elle est fusionnée avec la fenêtre B (voir

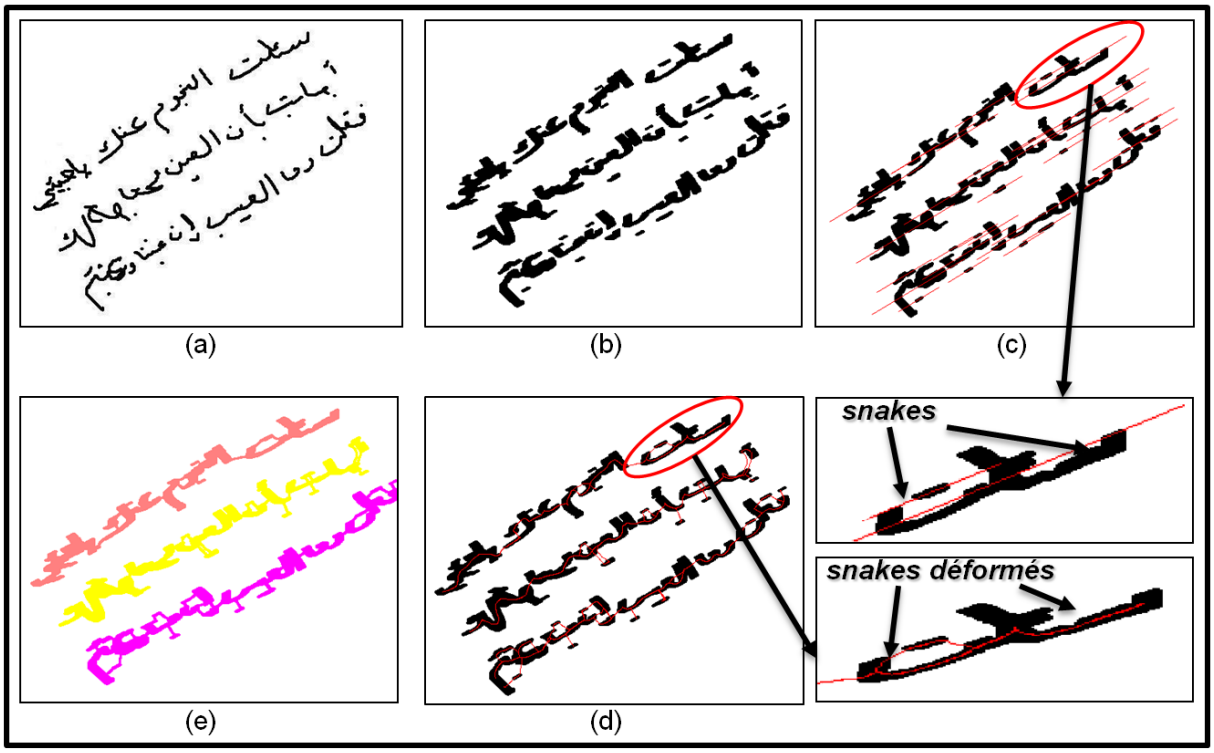


FIGURE 1.7: Application du snake pour la détection de lignes, (a) l'image initiale, (b) l'image initiale après une dilatation et une érosion, (c) l'axe majeur tracé pour chaque composante connexe des lignes. L'ellipse englobe la première composante connexe de (b), (d) montre la déformation du snake (c) et (e) donne le résultat final montrant les composantes connexes groupées dans chaque ligne.

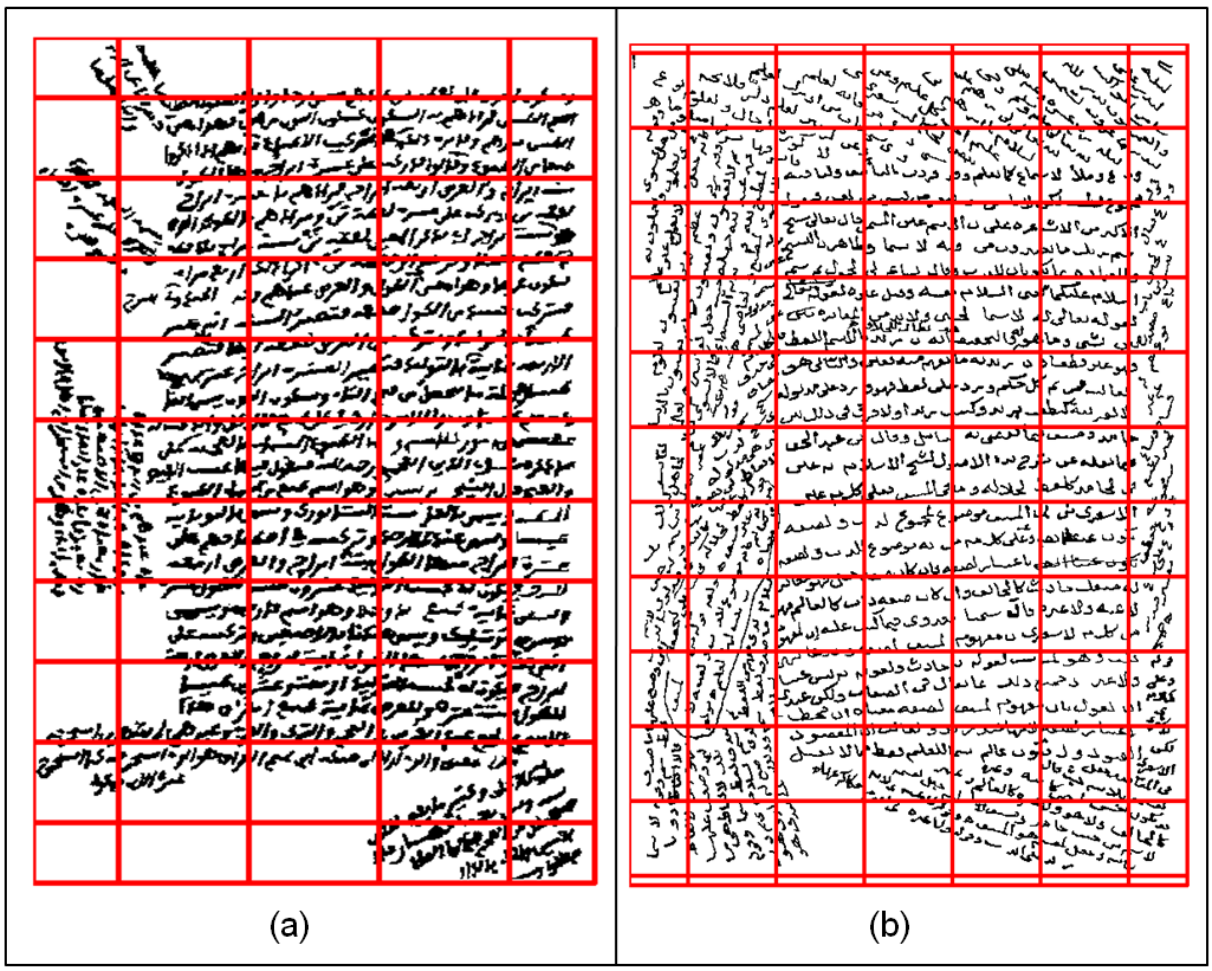


FIGURE 1.8: Résultats du maillage automatique des documents présentés dans la figure 1.2.

Algorithme 1: Détection des alignements en appliquant le modèle de contour actif (snake).

Données : Fenêtre w_i
Résultat : Composantes connexes de chaque alignement

Calculer les composantes connexes (CCXs) de w_i ;
j←1 ;
pour *chaque CCX* **faire**
  Calculer l'axe majeur $MajorAxe(j)$;
  j←j+1 ;
fin
j←1 ;
pour *chaque* $MajorAxe$ **faire**
  Définir le *snake* avec les pixels de $MajorAxe(j)$;
  Incrémenter la longueur de $MajorAxe(j)$;
  Déformer le snake ;
  j←j+1 ;
fin
Recalculer les composantes connexes $CCXnew$;
Chercher les CCX qui ∈ à chaque alignement ;

figure 1.10.c).

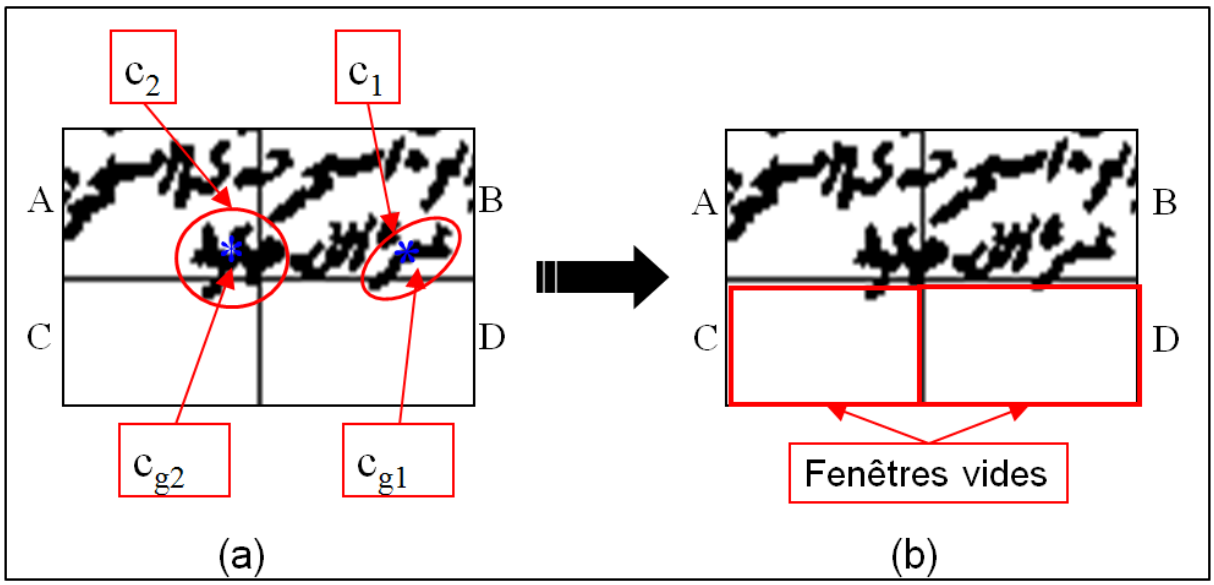


FIGURE 1.9: Exemple d'une fenêtre vide.

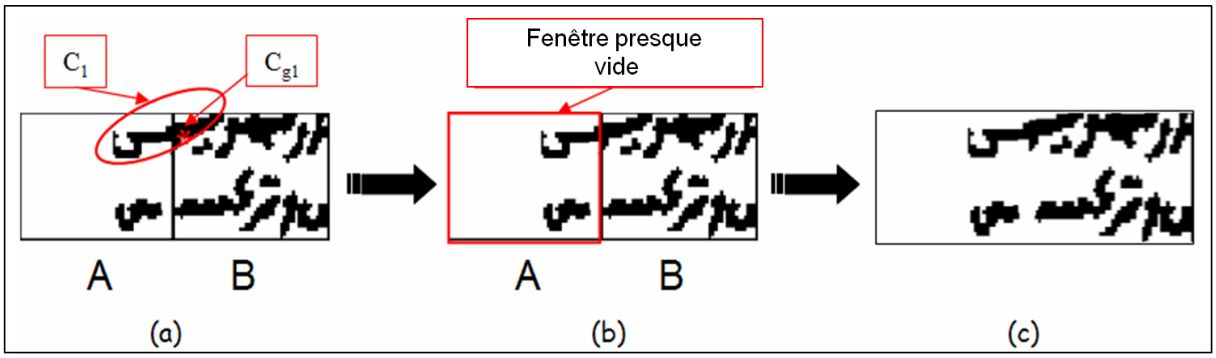


FIGURE 1.10: Exemple d'une fenêtre presque vide (contient peu de pixels).

1.5 Expérimentations

Nous avons testé la technique de maillage sur 36 anciens documents manuscrits arabes multi-orientés. Ces documents sont bien choisis pour représenter 4 classes différentes : documents simples (15), documents multi-orientés (15), documents multi-tailles (3) et documents avec problèmes au centre (3). Ces nombres suffisent pour avoir une idée sur la faisabilité ou non de la méthode. Ils seront augmentés ultérieurement. Ces classes sont décrits dans l'Annexe C.

La qualité de division dépend des paramètres de l'algorithme du modèle de contour actif (snake) qui sont : le coefficient d'élasticité α, le coefficient de rigidité β, le facteur de bruit μ de l'énergie externe (GVF : Gradient Vector Flow) et le nombre d'itérations de l'algorithme de minimisation du mécanisme d'énergie. L'erreur arrive lorsque le nombre des lignes est moins de 3 dans chaque fenêtre. Le tableau 1.2 donne la variation du taux d'erreur moyen par rapport aux trois paramètres α, β et μ. Le nombre d'itérations a été fixé à 25. On remarque que le taux d'erreur augmente lorsque la valeur de α augmente et β et μ diminuent. Le taux d'erreurs est minimal (0.30%) pour $\alpha = 1$, $\beta = 0$ et $\mu = 0.2$.

α	β	μ	**Taux d'erreur (%)**
0	1	1.2	0.81
0.2	0.8	1	0.67
0.4	0.6	0.8	0.56
0.6	0.4	0.6	0.52
0.8	0.2	0.4	0.34
1	**0**	**0.2**	**0.30**

TABLEAU 1.2: Variation du taux d'erreur par rapport aux trois paramètres (α, β et μ) pour un nombre d'itérations égal à 25.

1.6 Conclusion

Nous avons présenté dans ce chapitre, le maillage automatique des documents. Ce maillage a pour but de détecter toutes les zones d'orientations dans les images de documents. L'idée est basée sur le modèle de contour actif (*snake*). Le snake est appliqué afin de trouver une fenêtre initiale qui contient au moins trois lignes, à partir de laquelle on étend le maillage sur toute l'image. Le nombre des lignes est fixé à 3 pour avoir un histogramme de projection représentatif de l'écriture et capable d'estimer l'orientation exacte de la fenêtre.

Chapitre 2

Estimation de l'orientation

Nous présentons dans ce chapitre, la méthode d'estimation de l'orientation en nous basant sur la classe de Cohen. Le chapitre commence par une explication de la méthode d'analyse du profil de projection pour l'estimation de l'orientation. Puis, les théories des distributions temps-fréquence énergétiques seront détaillées afin d'expliquer leurs propriétés et leurs caractéristiques. Ensuite, leur emploi pour l'estimation de l'orientation sera exposé. Enfin, nous montrerons des expérimentations et nous terminerons le chapitre par une conclusion.

Sommaire

2.1 Introduction 47
2.2 Calcul de l'orientation par analyse des histogrammes de projection 49
2.3 Les distributions temps-fréquence 51
2.4 Application des distributions temps-fréquence pour l'estimation de l'orientation 59
2.5 Expérimentations 61
2.6 Conclusion 69

2.1 Introduction

Plusieurs techniques ont été proposées dans la littérature pour l'estimation de l'orientation d'une partie ou de tout un document.

Dans [Akiyama et Hagita, 1990, Pavlidis et Zhou, 1992, Sharif et Movahhedinia, 2008, Zhang et Hu, 2008], les histogrammes de projection de l'ensemble du document ou d'une bande de celui-ci sont calculés pour différents angles. Puis, les différences entre les pics et les vallées de ces histogrammes sont calculées pour estimer l'angle d'inclinaison du document. D'autres travaux [Inglis, 1999, Pal et Chaudhuri, 1996, Sarfraz *et al.*, 2007] utilisent la classification par k-ppv, où les composantes connexes sont regroupées en k-

groupes. L'orientation au sein de chaque groupe est d'abord déterminée, puis celle de l'image est obtenue en faisant la moyenne des orientations des groupes. On trouve également des travaux qui calculent des valeurs cumulées des corrélations entre les paires des lignes consécutives [Chaudhuri et Chaudhuri, 1997, Pal et Chaudhuri, 1997].

Dans [Postl, 1986, Srihari et Govindraju, 1989, Aradhya *et al.*, 2006], la transformée de Hough a été utilisée en choisissant comme points votants des points probablement colinéaires. Ces points sont les pixels les plus bas [Postl, 1986] des composantes connexes. L'inconvénient de cette transformation est qu'elle nécessite un grand espace mémoire et sa durée d'exécution est rédhibitoire. La valeur maximale du module carré de la transformée de Fourier a aussi été utilisée pour estimer l'angle d'inclinaison [Peake et Tan, 1997, Yin, 2001]. Les moindres carrés ont été utilisés pour évaluer l'inclinaison du texte en recherchant l'axe le plus représentatif des données [Belaid et Belaid, 1992, Inglis, 1999, Babu *et al.*, 2006]. Deux points au hasard sur cet axe permettent alors d'estimer l'orientation.

Dans [Chen et Haralick, 1994, Das et Chanda, 2001, Dhandra *et al.*, 2006], les opérations morphologiques comme la dilatation, l'érosion, l'ouverture et la fermeture sont utilisées pour transformer les lignes du texte en bandes noires de pixels et pour supprimer les connexions entre ces bandes. Deux points dans chaque bande (ligne) sont identifiés pour mesurer l'inclinaison de la ligne et l'angle moyen de toutes ces inclinaisons est considéré comme l'orientation de l'ensemble du document. Sauvola et al. [Sauvola *et al.*, 1997] ont également estimé l'orientation en cherchant des alignements (appelés "runs") qui correspondent à des lignes du texte espacées par des zones qui ne contiennent pas de pixels noirs. Cette méthode a été appliquée pour les documents latins imprimés. Elle ne pourrait probablement pas fournir de bons résultats pour des documents arabes manuscrits car ces derniers ne possèdent généralement pas d'interligne constante, et parce que les lignes sont très proches et se touchent dans la plupart des cas.

D'autres approches utilisent les caractéristiques particulières de l'alphabet de chaque langue. Par exemple, dans [Pal et Chaudhuri, 1997], la ligne supérieure qui se trouve dans tous les caractères des alphabets indiens (Devnagari et Bangla) est utilisée pour estimer l'orientation.

Dans ce chapitre, nous présentons une nouvelle méthode d'estimation de l'angle d'inclinaison de documents arabes manuscrits, qui s'appuie sur une analyse temps-fréquence des histogrammes de projection du document.

Dans [Kavallieratou *et al.*, 1999, Kavallieratou *et al.*, 2002], Kavallieratou et al. ont déjà utilisé la distribution de Wigner-Ville (DWV) pour estimer l'orientation globale des documents latins imprimés ou manuscrits. Dans le présent travail, nous avons souhaité déterminer d'abord les propriétés des représentations temps-fréquence qui semblent souhaitables pour une telle application, afin d'établir ensuite une liste de représentations, dont les performances ont enfin été évaluées.

2.2 Calcul de l'orientation par analyse des histogrammes de projection

La technique classique d'analyse des histogrammes de projection [Zhang et Hu, 2008, Akiyama et Hagita, 1990, Pavlidis et Zhou, 1992, Sharif et Movahhedinia, 2008] permet de percevoir plusieurs caractéristiques du contenu d'un document écrit, en particulier la densité du document, les interlignages, la distribution des blocs et bien sûr l'inclinaison. Pour estimer cette dernière, deux cas peuvent se présenter :

Cas 1 : la fenêtre ne contient qu'une seule ligne. L'orientation est estimée en calculant les histogrammes pour des angles de projection espacés de quelques degrés, retenant celui qui est le plus effilé (voir figure 2.1, où l'inclinaison correcte est de 0 °).

Cas 2 : la fenêtre contient plusieurs lignes. L'orientation est déterminée par l'histogramme de projection qui a le plus grand écart moyen entre les pics et les vallées, possédant donc les pics les plus élevés et les vallées les plus basses ou creuses (voir figure 2.2, où l'inclinaison correcte est de 0 °).

En général, on souhaite se trouver dans le deuxième cas où le grand nombre de pics et vallées assurent un examen plus précis de l'orientation.

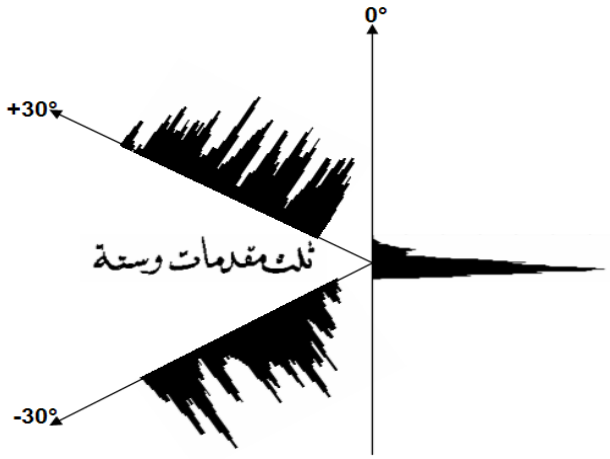


FIGURE 2.1: Histogrammes de projection d'un document ne comportant qu'une seule ligne, calculés pour des angles de projection de +30 °, 0 ° et −30 °.

Cette approche classique par l'examen de l'histogramme de projection ne conduit pas toujours à une estimation correcte de l'orientation. Cela peut arriver dans deux cas. Dans le premier cas, les mots arabes sont composés de PAWs (Part of Arabic Words) qui sont presque verticaux ou obliques, et qui faussent le calcul de l'orientation de la ligne. En outre, les écarts entre les distances entre les PAWs sont très étroits, essentiellement dans le cas de textes manuscrits. Par exemple, l'angle estimé pour la ligne horizontale dans la figure 2.3.a par l'analyse du profil de l'histogramme de projection est égal à +45 ° tandis que l'angle exact est 0 °. Le deuxième cas se produit des documents manuscrits complexes

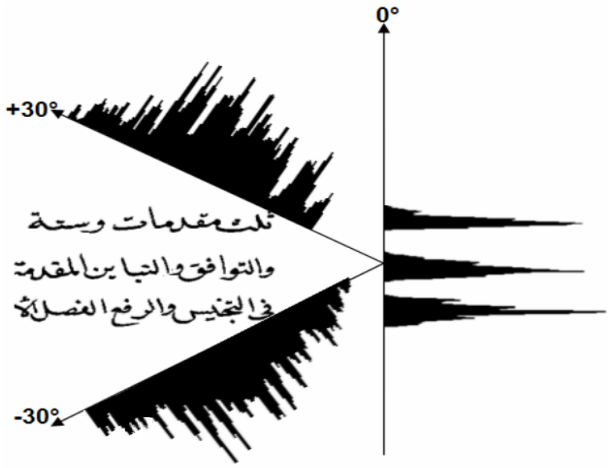


FIGURE 2.2: Histogrammes de projection d'un document comportant plusieurs lignes, calculés pour des angles de projection de $+30\,°$, $0\,°$ et $-30\,°$.

qui présentent beaucoup de bruit, de symboles diacritiques, de connexions entre les lignes (voir figure 2.3.b). Ces artefacts, ajoutés à l'irrégularité de l'écriture arabe et la présence de hampes et de jambages très allongés créent des maxima locaux dans les histogrammes de projection, et diminuent la valeur de la différence moyenne entre les pics et les vallées. Pour pouvoir disposer d'un estimateur plus robuste, nous avons donc envisagé d'utiliser des représentations temps-fréquence de ces histogrammes.

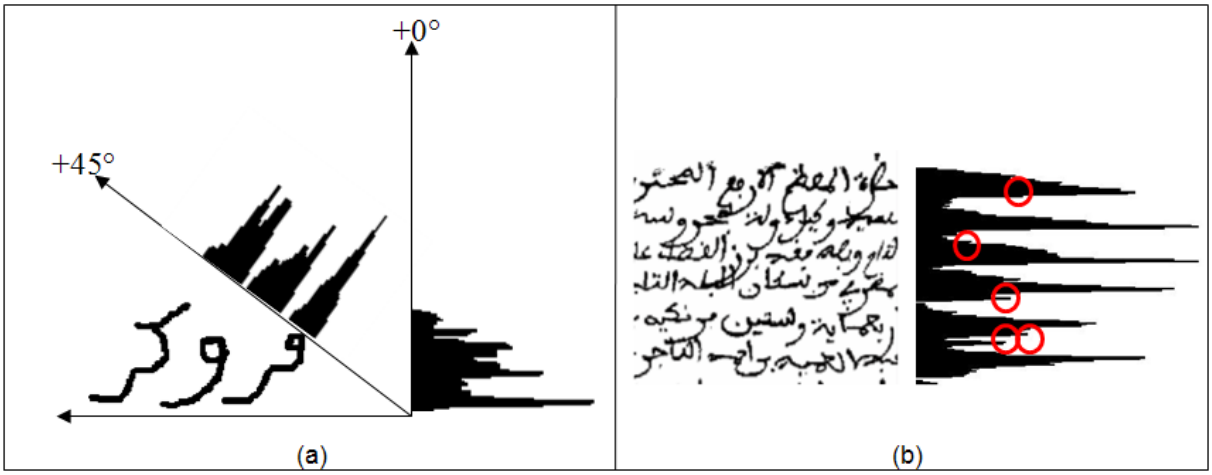


FIGURE 2.3: Problèmes d'orientation dans les documents arabes : (a) estimation erronée dûe aux orientations obliques de PAWs, (b) Apparition de maxima locaux à cause du bruit (cercles rouges).

2.3 Les distributions temps-fréquence

Les approches classiques de traitement du signal comme la transformée de Fourier ne parviennent pas à décrire simultanément la variation d'un signal par rapport au temps et à la fréquence. Les distributions temps-fréquence énergétiques vont au delà de ce qu'autorisent ces approches classiques en analysant la « non-stationnarité » d'un signal et en distribuant son énergie dans un espace à la fois temporel et fréquentiel. D'après [Hlawatsch et Auger, 2005], l'énergie E_x d'un signal $x(t)$ a pour expression

$$E_x = \int_{-\infty}^{+\infty} |x(t)|^2 \, dt = \int_{-\infty}^{+\infty} |\widehat{x}(f)|^2 \, df \tag{2.1}$$

où $\widehat{x}(f)$ est la transformée de Fourier du signal $x(t)$. Cette énergie étant une quantité quadratique, les distributions temps-fréquence énergétiques doivent conserver cette propriété.

En 1966, Cohen [Cohen, 1966, Escudié et Gréa, 1976], ont montré qu'un nombre important de distributions temps-fréquence énergétiques peuvent être vues comme des cas particuliers d'une expression générale

$$C_x(t,f) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \phi_{r-d}(\tau,\xi) A_x(\tau,\xi) e^{j2\pi(t\xi - f\tau)} d\xi d\tau, \tag{2.2}$$

où $A_x(\tau,\xi)$ est la fonction d'ambiguïté à bande étroite du signal, définie par

$$A_x(\tau,\xi) = \int_{-\infty}^{+\infty} x(t+\tau/2)\, x^*(t-\tau/2) e^{-j2\pi\xi t}\, dt$$

La classe de Cohen contient toutes les distributions temps-fréquence quadratiques qui vérifient la propriété d'invariance par translation temporelle ou fréquentielle [Flandrin, 1993, Hlawatsch et Auger, 2005, Hlawatsch et Boudreaux-Bartels, 1992]. Chacun des membres de cette classe se distingue par un noyau $\phi_{r-d}(\tau,\xi)$, qui joue un rôle déterminant dans la qualité des images fournies et dans les propriétés [Hlawatsch et Boudreaux-Bartels, 1992, Auger et Doncarli, 1992, Classen et Mecklenbrauker, 1980, Cohen, 1989, Flandrin, 1999, Flandrin *et al.*, 1992] qu'il vérifie.

L'utilisation de ces distributions est très répandue, notamment dans des applications de la reconnaissance des formes [Boashash *et al.*, 1987, Cristobal *et al.*, 1986] ou pour le traitement des signaux sismiques [Boles et Boashash, 1988].

Dans notre contexte d'application, un ensemble de propriétés nous ont semblé nécessaires ou souhaitables. Elles sont regroupées dans le tableau 2.1 et commentées ci-dessous :

Propriété P1 : valeur réelle

$$\forall t, f, C_x(t,f) = C_x^*(t,f) \tag{2.3}$$

où $C_x^*(t,f)$ est le conjugué de $C_x(t,f)$.

Cette propriété permet d'associer à chaque point du plan temps-fréquence un nombre réel et non un nombre complexe, ce qui réduit la quantité de mémoire nécessaire au stockage de la distribution. Cela permet aussi de construire facilement une représentation graphique de la distribution temps-fréquence, par exemple par une image indiquant par des niveaux de gris entre blanc (énergie forte) et noire (énergie faible), la valeur de la distribution en chaque point. Enfin, cela permet d'utiliser une relation d'ordre pour comparer les valeurs entre elles et en extraire le maximum. Par exemple, la valeur maximale de la distribution de Wigner-Ville du signal analytique [Boashash, 1992] du signal de la figure 2.4.a est 47.117.

Propriété P2 : covariance vis-à-vis des translations temporelles et fréquentielles

$$\forall t_1, f_1, y(t) = x(t - t_1)e^{j2\pi f_1 t}$$

$$\Rightarrow C_y(t, f) = C_x(t - t_1, f - f_1) \tag{2.4}$$

Cela veut dire qu'un simple décalage du texte dans l'image se traduira par un décalage de la représentation temps-fréquence de l'histogramme sans changer la valeur de son maximum. Ceci est intéressant car il montre que la distribution n'est pas sensible à la localisation des lignes dans la fenêtre. Par exemple, sur la figure 2.4.c, les deux images sont obtenues par translation verticale du document. Elles conduisent à des versions translatées des histogrammes de projection et donc à des versions translatées des représentations temps-fréquence, qui ont donc le même maximum absolu.

Propriété P3 : covariance vis-à-vis des changements d'échelle

$$\forall a > 0, y(t) = a^{-1/2}x(t/a)$$

$$\Rightarrow C_y(t, f) = C_x(t/a, af) \tag{2.5}$$

Grâce à cette propriété, un agrandissement ou une réduction du document se traduira par une dilatation ou une compression de son histogramme, qui se traduira par une modification en conséquence de sa représentation temps fréquence. Par exemple, le maximum de la représentation de Wigner-Ville du signal analytique du signal $y(t) = 2\,x(4t)$ (voir figure 2.4.b) est le même que celle de $x(t)$.

Propriété P4 : conservation de l'énergie

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} C_x(t, f)dtdf = \int_{-\infty}^{+\infty} |x(t)|^2\, dt \tag{2.6}$$

Cette propriété permet de garantir que l'énergie de l'histogramme de projection est exactement distribuée dans le plan temps-fréquence.

Propriété P5 : positivité

$$\forall x(t), \forall t, f, C_x(t, f) \geq 0 \tag{2.7}$$

Cette propriété permet d'interpréter plus facilement les valeurs de la représentation comme une mesure d'énergie.

Propriété P6 : conservation des valeurs nulles du signal

$$x(t_0) = 0 \Rightarrow \forall f, C_x(t_0, f) = 0 \tag{2.8}$$

Cette propriété permet d'avoir une représentation nulle à toutes les « fréquences » aux « instants » où l'histogramme de projection est nul. Ceci permet de retrouver dans la représentation temps-fréquence les vallées de cet histogramme de projection (voir figure 2.4.d).

Propriété P7 : conservation du support du signal

Si

$$\forall t, |t| > T \Rightarrow x(t) = 0$$

alors

$$\forall t, |t| > T \Rightarrow C_x(t, f) = 0 \tag{2.9}$$

Ceci permet d'avoir une représentation temps-fréquence nulle à l'extérieur de l'intervalle dans lequel l'histogramme de projection prend des valeurs non nulles (la représentation temps-fréquence ne déborde pas du support de l'histogramme).

Pour ce problème d'estimation de l'inclinaison d'un document numérisé, les propriétés P1, P2, P3 et P4 nous paraissent fortement souhaitables, alors que les propriétés P5, P6 et P7 nous semblent moins indispensables, mais susceptibles de fournir des estimations plus précises.

Les propriétés retenues nous ont permis d'établir ensuite une liste de représentations susceptibles de fournir de bonnes estimations de l'inclinaison. Ces représentations sont énumérées dans le tableau 2.2, avec leurs expressions et leurs propriétés. Nous allons détailler dans la suite ces distributions qui peuvent servir à l'estimation de l'orientation.

Wigner-Ville : l'élement le plus étudié de la classe de Cohen est certainement la distribution de Wigner-Ville [Wigner, 1932]. Elle peut être vue comme l'unique distribution à valeurs réelles qui vérifie les propriétés nécessaires pour les applications

classiques de traitement du signal. Son noyau $\phi_{r-d}(\tau, \xi)$ est égal à 1, ce qui donne la formule suivante :

$$W_x(t,f) = \int_{-\infty}^{+\infty} x(t+\tau/2)x^*(t-\tau/2)e^{-j2\pi f\tau}d\tau \tag{2.10}$$

L'utilisation pratique de cette distribution est rendue difficile d'une part par son expression, qui fait intervenir une intégrale sur toutes les valeurs du signal, et d'autre part par ses termes d'interférence, à la fois nombreux et élevés.
L'interférence est produite parce que la représentation d'un signal égal à une somme de deux signaux n'est pas égale à la somme des représentations de chacun des deux signaux : si $y(t) = x_1(t) + x_2(t)$, alors

$$C_y(t,f) = C_{x_1}(t,f) + C_{x_2}(t,f) + C_{x_1x_2}(t,f) + C_{x_2x_1}(t,f) \tag{2.11}$$

avec

$$C_{x_1x_2}(t,f) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \phi_{t-r}(t-\upsilon,\tau)x_1(\upsilon+\tau/2)x_2^*(\upsilon-\tau/2)e^{-j2\pi f\tau}d\upsilon d\tau \tag{2.12}$$

Pseudo Wigner-Ville : pour éviter le problème d'interférence, la distribution de *pseudo Wigner-Ville* [Classen et Mecklenbrauker, 1980] fait intervenir une fenêtre $h(\tau)$ qui rend la distribution calculable, réduit le nombre de termes d'interférence (en supprimant les interférences entre composantes éloignées dans le temps) et réduit leur amplitude (par un lissage dans la direction fréquentielle : $\phi_{t-f}(t,f) = \widehat{h}(f)$). Son noyau $\phi_{r-d}(\tau,\xi)$ est égal à $h(\tau)$, ce qui donne l'expression suivante :

$$PW_x(t,f) = \int_{-\infty}^{+\infty} h(\tau)x(t+\tau/2)x^*(t-\tau/2)e^{-j2\pi f\tau}d\tau \tag{2.13}$$

Pseudo Wigner-Ville lissée : pour obtenir des résultats plus lisibles encore, la représentation *pseudo Wigner-Ville lissée* [Flandrin, 1993, Flandrin et Martin, 1984] utilise une fenêtre supplémentaire $g(t)$, afin d'opérer un lissage temporel indépendant du lissage fréquentiel. Son noyau $\phi_{r-d}(\tau,\xi)$ est égal à $h(\tau)\widehat{g}(\xi)$, ce qui donne l'expression suivante :

$$PWL_x(t,f) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(t-\upsilon)h(\tau)x(\upsilon+\tau/2)x^*(\upsilon-\tau/2)e^{-j2\pi f\tau}d\tau d\upsilon \tag{2.14}$$

Margenau-Hill : [Margenau et Hill, 1961] c'est la partie réelle de la distribution de Rihaczek. La distribution de *Rihaczek* [Rihaczek, 1992] vérifie un nombre important de propriétés, ce qui paraît miraculeux au regard de la simplicité de sa définition.

Son noyau $\phi_{r-d}(\tau,\xi)$ est égal à $e^{-j\pi\xi\tau}$, ce qui donne la formule suivante :

$$x(t)\widehat{x}^*(f)e^{-j2\pi ft} \tag{2.15}$$

Son noyau $\phi_{r-d}(\tau,\xi)$ est égal à $cos(\pi\xi\tau)$, ce qui donne l'expression suivante :

$$\frac{1}{2}\int_{-\infty}^{+\infty}\left[x(t+\tau)x^*(t)+x(t)x^*(t-\tau)\right]e^{-j2\pi f\tau}d\tau \tag{2.16}$$

Pseudo Margenau-Hill : [Hippenstiel et Oliviera, 1990] c'est une version à la fois calculable et plus lisible. Son noyau $\phi_{r-d}(\tau,\xi)$ est égal à $h(\tau)cos(\pi\xi\tau)$, ce qui conduit à :

$$\frac{1}{2}\int_{-\infty}^{+\infty}h(\tau)\left[x(t+\tau)x^*(t)+x(t)x^*(t-\tau)\right]e^{-j2\pi f\tau}d\tau \tag{2.17}$$

Spectrogramme : le spectrogramme est utilisé pour analyser des signaux lentement non-stationnaires ou comportant un grand nombre des composantes [Stankovic, 1994]. Son noyau $\phi_{r-d}(\tau,\xi)$ est égal à $A_h^*(\tau,\xi)$, ce qui donne :

$$\left|\int_{-\infty}^{+\infty}x(u)h^*(u-t)e^{-j2\pi fu}du\right|^2 \tag{2.18}$$

Choi-Williams : cette distribution [Choi et Williams, 1989] réduit les termes d'interférences en utilisant des noyaux de type produit, de la forme $\phi_{r-d}(\tau,\xi)=P(\tau\xi)$. P correspond à la réponse fréquentielle d'un filtre passe-bas, dont la réponse impulsionnelle est infinie. Son noyau $\phi_{r-d}(\tau,\xi)$ est égal à $exp(-4\pi^2\xi^2\tau^2/\sigma)$, ce qui conduit à :

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\frac{exp(-\frac{\upsilon^2}{4\tau^2/\sigma})}{\sqrt{4\pi\tau^2/\sigma}}x(t-\upsilon+\tau/2)x^*(t-\upsilon-\tau/2)e^{-j2\pi f\tau}d\upsilon d\tau \tag{2.19}$$

Born-Jordan : c'est la plus ancienne représentation de type produit dont la réponse impulsionnelle est finie [Born et Jordan, 1925]. Son noyau $\phi_{r-d}(\tau,\xi)$ est égal à $sinc(\pi\xi\tau)$, ce qui donne :

$$\int_{-\infty}^{+\infty}\int_{-|\tau|/2}^{|\tau|/2}\frac{1}{|\tau|}x(t-\upsilon+\tau/2)x^*(t-\upsilon-\tau/2)e^{-j2\pi f\tau}d\upsilon d\tau \tag{2.20}$$

Zhao-Atlas-Marks : cette distribution [Zhao *et al.*, 1990] réduit les termes d'interférence mais elle n'est pas de type produit. Son noyau $\phi_{r-d}(\tau,\xi)$ est égal à $h(\tau)\,|\tau|\,sinc(\pi\xi\tau)$, ce qui conduit à :

$$\int_{-\infty}^{+\infty}\int_{-|\tau|/2}^{|\tau|/2}h(\tau)x(t-\upsilon+\tau/2)x^*(t-\upsilon-\tau/2)e^{-j2\pi f\tau}d\upsilon d\tau \tag{2.21}$$

No	**Propriété**	**Expression**
P1	Valeur réelle	$\forall t, f, C_x(t,f) = C_x^*(t,f)$ où $C_x^*(t,f)$ est le conjugué de $C_x(t,f)$
P2	Covariance vis-à-vis des translations temporelles et fréquentielles	$\forall\, t_1,\ f_1,$ $y(t) = x(t-t_1)\, e^{j2\pi f_1 t} \Rightarrow$ $C_y(t,f) = C_x(t-t_1, f-f_1)$
P3	Covariance vis-à-vis des changements d'échelle	$\forall\, a > 0,$ $y(t) = a^{-1/2}\, x(t/a) \Rightarrow$ $C_y(t,f) = C_x(t/a, af)$
P4	Conservation de l'énergie	$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} C_x(t,f) dt df$ $= \int_{-\infty}^{+\infty} \lvert x(t)\rvert^2 dt$
P5	Positivité	$\forall\, x(t), \forall\, t, f,\ C_x(t,f) \geq 0$
P6	Conservation des valeurs nulles du signal	$x(t_0) = 0 \Rightarrow$ $\forall\, f,\ C_x(t_0, f) = 0$
P7	Conservation du support du signal	si $\forall\, t, \lvert t\rvert > T \Rightarrow x(t) = 0,$ alors $\forall\, t, \lvert t\rvert > T \ \Rightarrow\ C_x(t,f) = 0$

TABLEAU 2.1: Propriétés des distributions temps-fréquence de la classe de Cohen considérées comme nécessaires ou souhaitables dans ce contexte d'application.

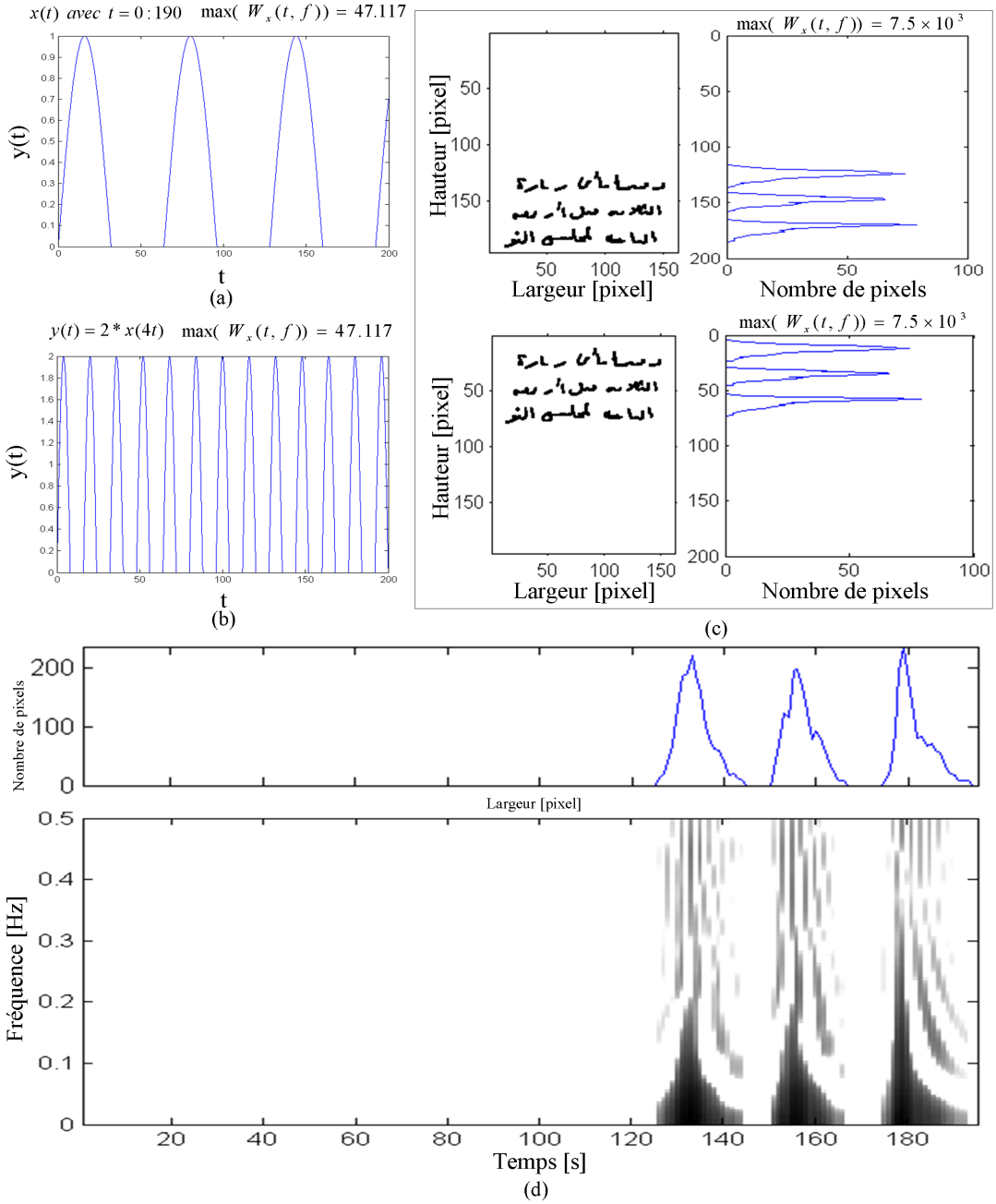


FIGURE 2.4: Illustration des propriétés des distributions de la classe de Cohen et plus particulièrement de la distribution de Wigner-Ville (figures a, b et c) et de la distribution pseudo Margenau-Hill (figure d). Le signal $x(t)$ utilisé dans les figures a et b est égal à $\max(\sin(2\pi\, t/64), 0)$.

Distribution	**Expression**	**Propriétés**						
		P1	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**
Wigner-Ville	$\int_{-\infty}^{+\infty} x(t+\tau/2)\, x^*(t-\tau/2)\, e^{-j2\pi f\tau} d\tau$	x	x	x	x			x
Pseudo Wigner-Ville	$\int_{-\infty}^{+\infty} h(\tau)\, x(t+\tau/2)\, x^*(t-\tau/2)\, e^{-j2\pi f\tau} d\tau$	x	x		x			x
Pseudo Wigner-Ville lissée	$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(t-\upsilon)\, h(\tau) x(\upsilon+\tau/2) x^*(\upsilon-\tau/2) e^{-j2\pi f\tau} d\tau d\upsilon$	x	x		x			
Margenau-Hill	$\frac{1}{2}\int_{-\infty}^{+\infty} [x(t+\tau)x^*(t) + x(t)x^*(t-\tau)]\, e^{-j2\pi f\tau} d\tau$	x	x	x	x		x	x
Pseudo Margenau-Hill	$\frac{1}{2}\int_{-\infty}^{+\infty} h(\tau)\, [x(t+\tau)x^*(t) + x(t)x^*(t-\tau)]\, e^{-j2\pi f\tau} d\tau$	x	x	x	x		x	x
Spectrogramme	$\left\lvert \int_{-\infty}^{+\infty} x(u) h^*(u-t) e^{-j2\pi f u} du \right\rvert^2$	x	x	x	x	x		
Choi-Williams	$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \frac{\exp(-\frac{\upsilon^2}{4\tau^2/\sigma})}{\sqrt{4\pi\tau^2/\sigma}} x(t-\upsilon+\tau/2)\, x^*(t-\upsilon-\tau/2) e^{-j2\pi f\tau} d\upsilon\, d\tau$	x	x	x	x			
Born-Jordan	$\int_{-\infty}^{+\infty}\int_{-|\tau|/2}^{|\tau|/2} \frac{1}{\lvert\tau\rvert} x(t-\upsilon+\tau/2) x^*(t-\upsilon-\tau/2)\, e^{-j2\pi f\tau} d\upsilon\, d\tau$	x	x	x	x			x
Zhao-Atlas-Marks	$\int_{-\infty}^{+\infty}\int_{-|\tau|/2}^{|\tau|/2} h(\tau) x(t-\upsilon+\tau/2) x^*(t-\upsilon-\tau/2) e^{-j2\pi f\tau} d\upsilon\, d\tau$	x	x					x

TABLEAU 2.2: Distributions temps-fréquences retenues dans cette étude, avec leurs expressions et leurs propriétés.

2.4 Application des distributions temps-fréquence pour l'estimation de l'orientation

Lorsqu'on calcule l'histogramme de projection d'un document suivant une direction de projection, on obtient, lorsque cette direction correspond à l'orientation du document, un histogramme dans lequel chaque ligne d'écriture conduit à un maximum local nettement marqué (voir figure 2.2). Chaque paragraphe du document se révèle dans l'histogramme de projection par une succession périodique de pics et de vallées dont la période est sensiblement constante. Cette succession périodique constitue alors une composante de l'histogramme délimitée dans le « temps » par la taille du paragraphe et qui oscille à une « fréquence » déterminée par la largeur de l'interligne.

Comme tous les pixels s'accumulent aux mêmes endroits, les maxima locaux ont des niveaux d'énergie plus élevés que pour les autres directions de projection, ce qui explique que l'on peut estimer l'orientation d'un document en cherchant l'angle de projection pour laquelle la distribution temps-fréquence obtenue à partir de cet histogramme localise la plus grande quantité d'énergie en un point quelconque du plan temps-fréquence. À titre d'illustration, la figure 2.5 montre l'évolution de la valeur maximale de la distribution de Wigner-Ville lorsque le nombre de pics et de vallées augmente et lorsque les vallées se creusent dans le signal.

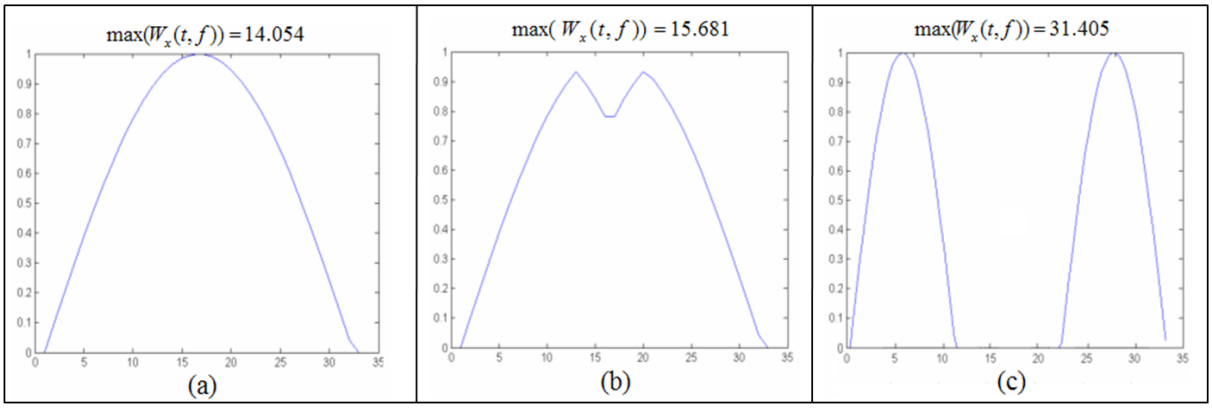


FIGURE 2.5: Exemples de valeurs maximales de la distribution de Wigner-Ville obtenues pour différents histogrammes de projection.

L'utilisation des représentations temps-fréquence de la classe de Cohen nous semble particulièrement justifiée dans ce contexte. En effet, l'histogramme de projection d'un document numérisé, obtenu pour une direction de projection donnée, est un signal non-stationnaire, constitué de composantes qui correspondent aux paragraphes du document et qui oscillent à une fréquence déduite de la largeur de l'interligne. Lorsque la direction de projection est égale à la direction principale du document, l'énergie est alors très localisée en « temps » (la direction perpendiculaire à la direction de projection) et en

« fréquence » (la fréquence spatiale associée), ce qui justifie d'estimer cet angle en recherchant la direction de projection pour laquelle la représentation temps-fréquence possède un maximum global. Ceci nécessite l'emploi d'une représentation à valeurs réelles, donc quadratique, car les représentations linéaires (telles que la transformée de Fourier court-terme [Gribonval, 2005] ou la transformée en ondelettes continue [Blu et Lebrun, 2005]) sont à valeurs complexes. Enfin, même si l'acquisition de l'image du document fait intervenir un agrandissement ou une réduction, qui s'apparentent à une compression ou à une dilatation, le document en lui-même est caractérisé par un interligne choisi par son auteur, dont la variation conduit, sur l'histogramme de projection, davantage à une modulation de fréquence qu'à un changement d'échelle : une augmentation de l'interligne ne conduit pas nécessairement à une augmentation de la taille des caractères. C'est pour cette raison que les représentations temps-fréquence, si possible également invariantes par changement d'échelle, nous semblent plus pertinentes pour cette application que des représentations affines [Ovarlez *et al.*, 2005].

Les représentations temps-fréquence utilisées étant quadratiques, c'est à la racine carrée des histogrammes que nous les appliquons. Les valeurs des représentations possèdent alors la même dimension physique que les histogrammes et correspondent donc au nombre de pixels noirs situés sur une droite parallèle à la direction de projection et espacés d'un certain interligne. La propriété de conservation de l'énergie permet alors de conserver l'intégrale de l'histogramme, c'est à dire le nombre total de pixels noirs contenus dans l'image. La propriété de positivité correspond au souhait de ne pas avoir de valeurs négatives lorsque le nombre total de pixels noirs est distribué dans le plan temps-fréquence. Les propriétés de conservation des valeurs nulles ou des supports correspondent au souhait d'avoir des effectifs nuls dans le plan temps-fréquence aux endroits où l'histogramme ne dénombre aucun pixel. Le calcul de la représentation temps-fréquence de la racine carrée des histogrammes et non pas des histogrammes, comme il est fait dans [Kavallieratou *et al.*, 1999, Kavallieratou *et al.*, 2002], renforce donc la signification des valeurs de la représentation et de ses propriétés. Les estimations de l'inclinaison obtenues dans les deux cas ne diffèrent cependant pas de manière significative.

Pour estimer l'angle d'inclinaison, la méthode que nous proposons consiste donc à calculer le signal analytique $x_a(t)$ de la racine carrée de l'histogramme de projection $x(t)$ du document, auquel a été retranchée sa valeur moyenne. Le signal analytique est obtenu à partir du signal $x(t)$ en lui retirant ses fréquences négatives [Boashash, 1992]. L'histogramme $x(t)$ est obtenu en projetant le document suivant un axe qui passe par le centre de gravité de l'image, qui constitue un point déduit du contenu de l'image, et non pas de ses dimensions et de son cadrage, et incliné d'un certain angle par rapport à l'axe horizontal de l'image, utilisé comme référence angulaire (voir figure 2.9). Nous calculons ensuite la représentation temps-fréquence de ce signal analytique. L'angle pour lequel la valeur maximale de cette représentation temps-fréquence est la plus élevée constitue l'estimation de l'angle d'inclinaison du document. À titre d'exemple, la figure 2.8 montre l'évolution,

en fonction de l'angle de projection, du maximum des distributions de Wigner-Ville, de spectrogramme et de la transformée de Fourier discrète calculée à partir de la racine carrée des histogrammes de projection du document de la figure 2.7 et du maximum de la moyenne de différence entre maxima et minima des histogrammes de projection du même document. Les coordonnées du point (t, f) du plan temps-fréquence où la valeur maximale de la représentation est obtenue pour l'inclinaison estimée pourraient être utilisées pour estimer l'emplacement du centre de gravité du paragraphe principal du document et la largeur de son interligne.

Pour réduire le coût de calcul et la quantité de mémoire nécessaire, une technique de recherche du maximum en trois phases est utilisée. Dans la première phase, les histogrammes de projection sont calculés pour des angles allant de $-85\,^\circ$ à $+90\,^\circ$ avec un pas de $5\,^\circ$. Dans la seconde phase, les angles de projection utilisés varient de $-4\,^\circ$ à $+4\,^\circ$ autour de l'inclinaison $\hat{\theta}_1$ estimée à la première phase, avec un pas de $1\,^\circ$. Dans la troisième phase, l'estimation est affinée en utilisant des angles de projection entre $-0.9\,^\circ$ et $+0.9\,^\circ$ autour de l'inclinaison $\hat{\theta}_2$ obtenue à la phase précédente, avec un pas de $0.1\,^\circ$. Dans chacune des trois phases, l'inclinaison est estimée en cherchant l'angle de projection pour lequel la plus grande valeur maximale de la représentation temps-fréquence est obtenue. Ces trois phases pour estimer l'orientation sont résumées dans l'algorithme 2.

Le tableau 2.3 montre les estimations de l'inclinaison obtenues de cette manière avec les représentations temps-fréquence retenues dans cette étude, pour le document de la figure 2.7. L'angle exact est de $+14.7\,^\circ$. L'algorithme de l'estimation de l'orientation est appliqué pour chaque fenêtre afin d'estimer les orientations initiales dans le document. La figure 2.6 présente les résultats de l'estimation de l'orientation dans chaque fenêtre (pavé) pour les documents présentés dans la figure 1.2. Chaque orientation est représentée par une couleur.

2.5 Expérimentations

Pour étudier l'efficacité de notre approche, nous l'avons testée sur un corpus de 864 documents arabes manuscrits. Ces documents ont été obtenus à partir des 72 documents mono-orientés, par orientation suivant 12 angles variant de $-75\,^\circ$ à $+90\,^\circ$. Les essais ont été effectués sur un PC sous Windows XP, équipé d'un processeur Pentium M de 1.4 GHz et d'une mémoire cache de 1 Go, avec Matlab R2007b complété par la boîte à outils temps-fréquence `tftb` [Auger *et al.*, 1996].

Le tableau 2.3 montre les estimations de l'inclinaison du document de la figure 2.7 obtenues avec les représentations temps-fréquence retenues dans cette étude. L'ensemble de ces estimations se situent entre $-2.1\,^\circ$ et $+0.5\,^\circ$ de la valeur exacte, égale à $14.7\,^\circ$. Pour ce document, la distribution de Wigner-Ville semble être la seule à donner un résultat plus intéressant que l'analyse des histogrammes de projection. L'estimation de l'inclinaison obtenue en cherchant l'angle de projection qui maximise le module carré de la transformée de

Algorithme 2: Estimation de l'orientation en appliquant la distribution de Wigner-Ville.

Données : Document D_i
Résultat : Orientation θ_i de D_i

$j \leftarrow 1$;
pour $\theta \leftarrow -75\,^\circ$ *jusqu'à* $\theta \leftarrow 90\,^\circ$ **faire**
  Calculer le signal analytique $x_a(j)$ de la racine carrée de l'histogramme de projection $x(j)$ du D_i selon $\theta(j)$;
  Calculer la valeur d'énergie $Energy1(j)$ par DWV ;
  $\theta \leftarrow \theta + 15\,^\circ$;
  j←j+1 ;
fin
Calculer $maxenergy1$ en appliquant $Max(Energy1)$;
Calculer θ_i correspondant au maximum de $Energy1$;
$j \leftarrow 1$;
pour $\theta \leftarrow \theta_i\,^\circ - 4\,^\circ$ *jusqu'à* $\theta_i\,^\circ + 4\,^\circ$ **faire**
  Calculer le signal analytique $x_a(j)$ de la racine carrée de l'histogramme de projection $x(j)$ du D_i selon $\theta(j)$;
  Calculer la valeur d'énergie $Energy2(j)$ par DWV ;
  $\theta \leftarrow \theta + 1\,^\circ$;
  j←j+1 ;
fin
Calculer $maxenergy2$ en appliquant $Max(Energy2)$;
si $maxenergy2 > maxenergy1$ **alors**
  θ_i correspondant au maximum de $Energy2$;
  $j \leftarrow 1$;
  pour $\theta \leftarrow \theta_i\,^\circ - 0.9\,^\circ$ *jusqu'à* $\theta_i\,^\circ + 0.9\,^\circ$ **faire**
    calculer le signal analytique $x_a(j)$ de la racine carrée de l'histogramme de projection x(j) du D_i selon $\theta(j)$;
    Calculer la valeur d'énergie $Energy3(j)$ par DWV ;
    $\theta \leftarrow \theta + 0.1\,^\circ$;
    j←j+1 ;
  fin
  Calculer $maxenergy3$ en appliquant $Max(Energy3)$;
  si $maxenergy3 > maxenergy2$ **alors**
    θ_i correspondant au maximum de $Energy3$;
    sinon
      θ_i correspondant au maximum de $Energy2$;
    fin
  fin
  sinon
    θ_i correspondant au maximum de $Energy1$;
  fin
fin

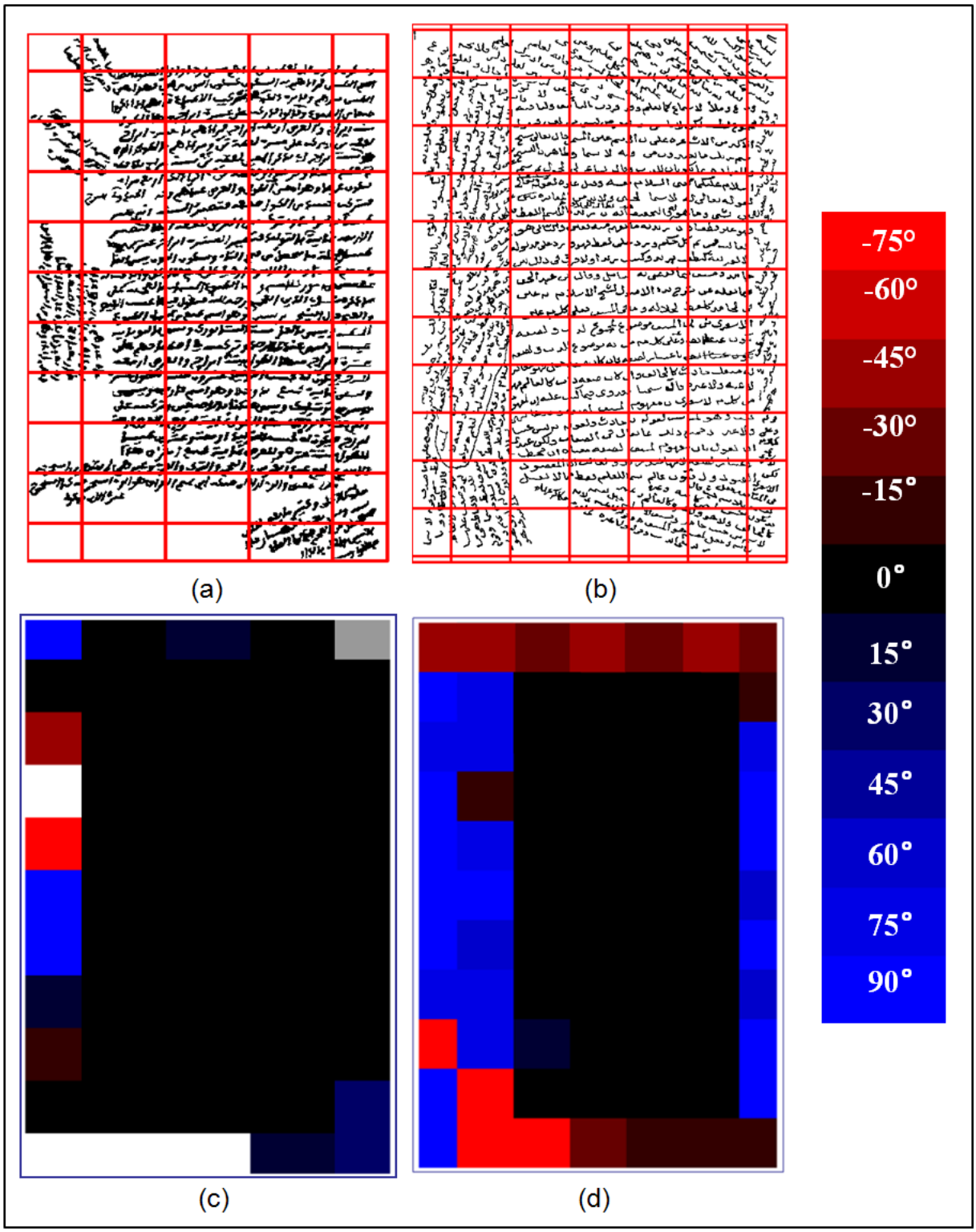


FIGURE 2.6: Résultats de l'estimation de l'orientation initiale du document présenté dans la figure 1.2. Les couleurs représentent des zones d'orientation suivant l'échelle des couleurs montrée à droite.

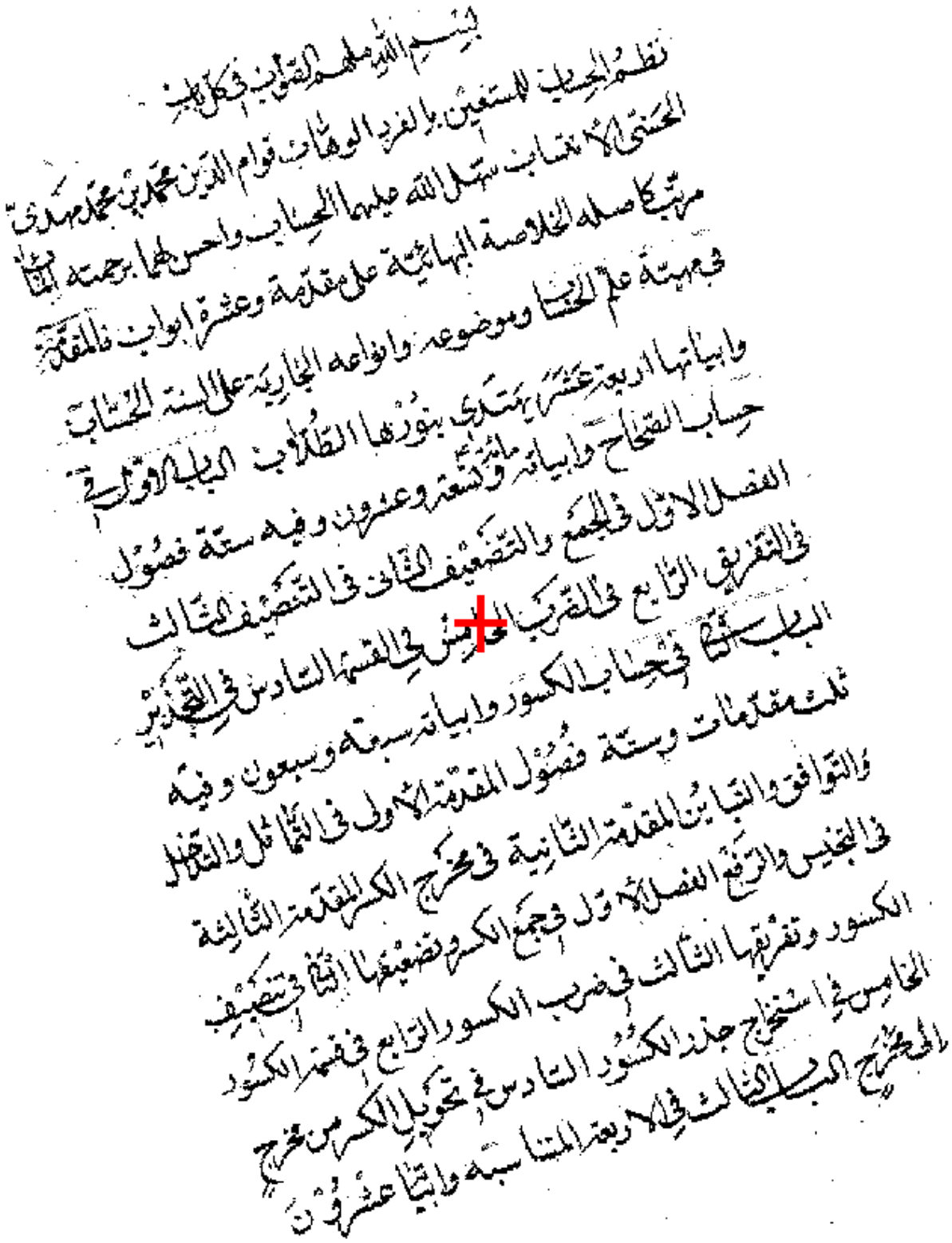


FIGURE 2.7: Document Arabe manuscrit incliné de $+14.7°$. La position du centre de gravité des pixels noirs de ce document, utilisée pour calculer les histogrammes de projection, est indiquée par une croix.

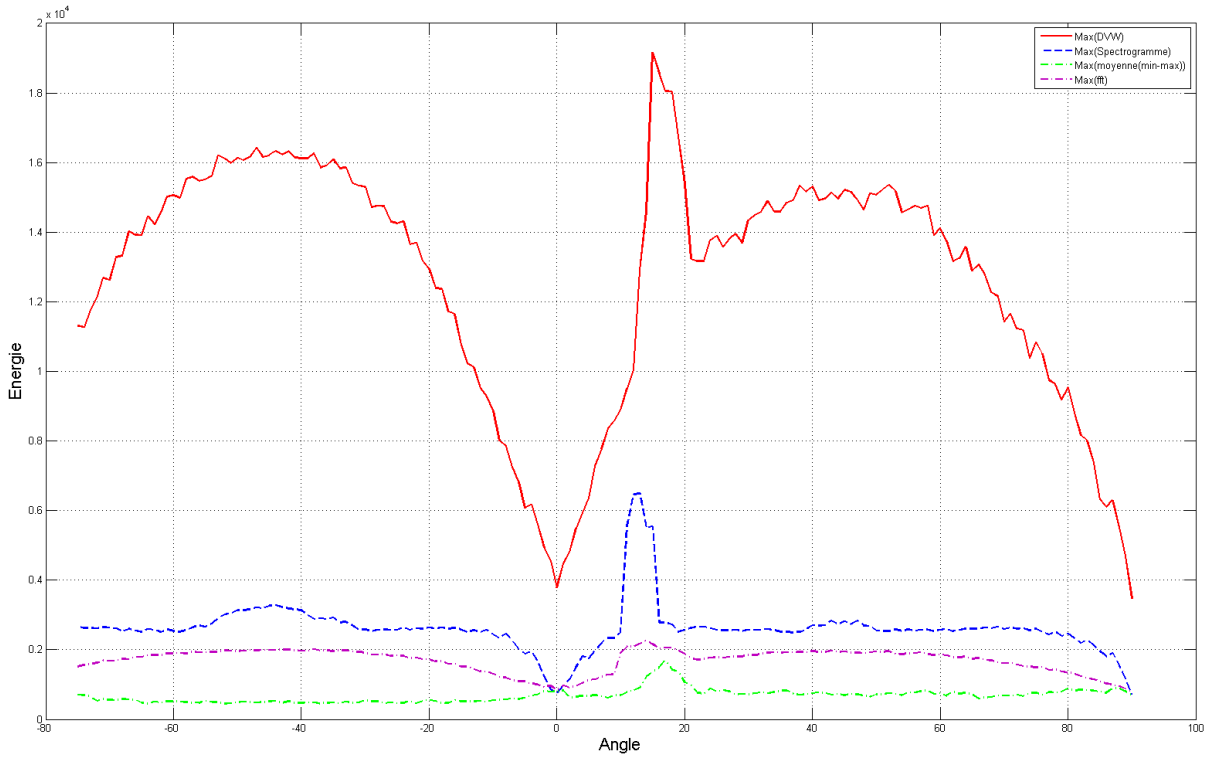


FIGURE 2.8: Courbes d'énergie : maximum des distributions de Wigner-Ville (traits pleins en rouge), de spectrogramme (traits pointillés en bleu) et de la transformée de Fourier discrète (traits mixtes en violet) de la racine carrée des histogrammes de projection du document de la figure 2.7 et maximum de la moyenne de différence entre maxima et minima (traits mixtes en vert) des histogrammes de projection du même document, en fonction de l'angle de projection.

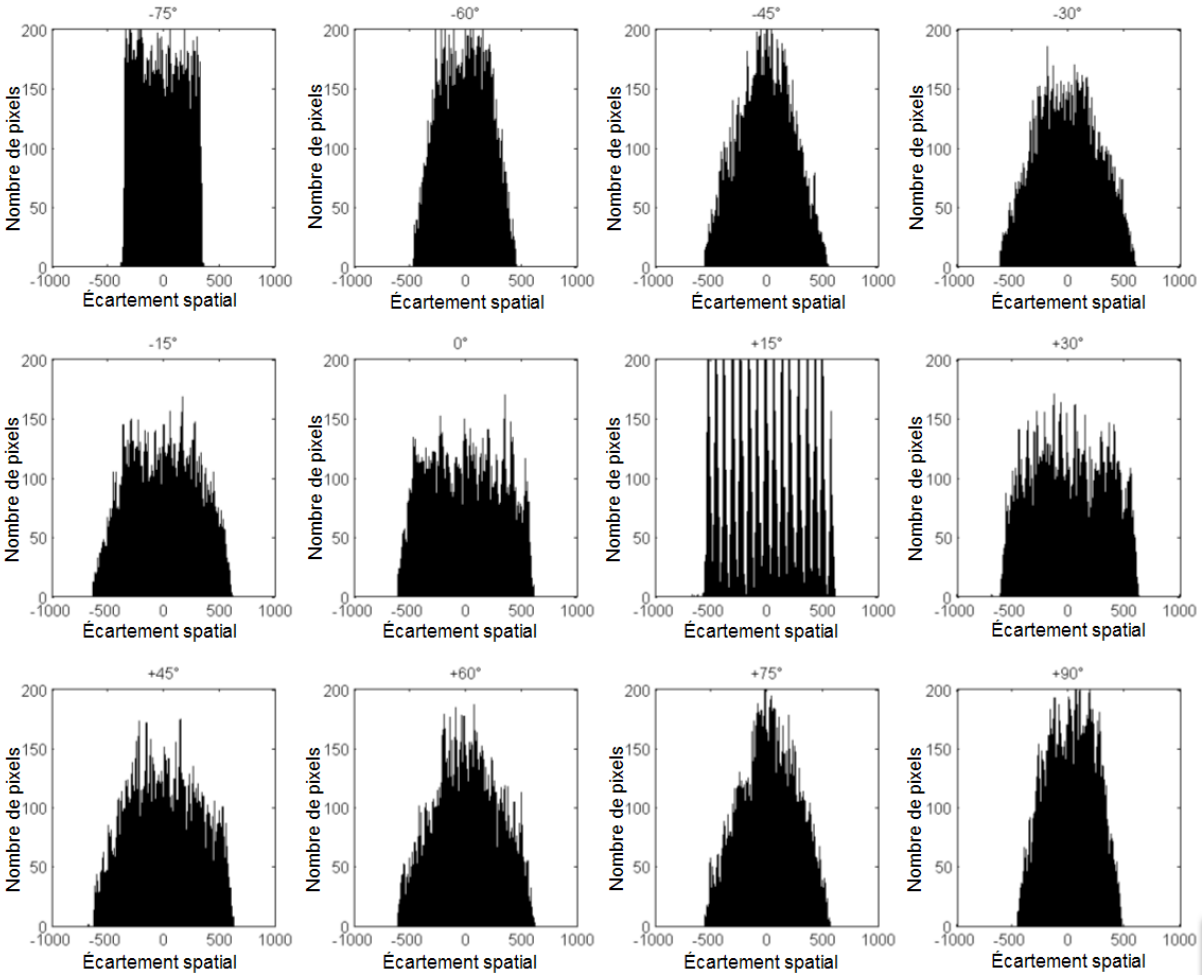


FIGURE 2.9: Quelques histogrammes de projection du document de la figure 2.7.

Distribution	**Angle estimé**
Wigner-Ville	$+14.7\,^\circ$
Pseudo Wigner-Ville	$+13.1\,^\circ$
Pseudo Wigner-Ville lissée	$+14.3\,^\circ$
Margenau-Hill	$+13.7\,^\circ$
Pseudo Margenau-Hill	$+15.1\,^\circ$
Spectrogramme	$+12.9\,^\circ$
Choi-Williams	$+14.2\,^\circ$
Born-Jordan	$+12.8\,^\circ$
Zhao-Atlas-Marks	$+14.6\,^\circ$
Histogrammes de projection	$+14.9\,^\circ$
Transformée de Fourier	$+13.8\,^\circ$

TABLEAU 2.3: Estimations de l'angle d'inclinaison du document de la figure 2.7 obtenues avec les représentations temps-fréquence retenues dans cette étude.

Distribution	**taux d'estimation (%)**	**temps moyen (s)**
Wigner-Ville	100	4.2
Pseudo Wigner-Ville	92	7.1
Pseudo Wigner-Ville lissée	97	6.9
Margenau-Hill	97	7.2
Pseudo Margenau-Hill	98	8.1
Spectrogramme	92	7.9
Choi-Williams	97	7.6
Born-Jordan	90	7.8
Zhao-Atlas-Marks	94	8.5
Histogrammes de projection	93	6.4
Transformée de Fourier	89	4.0

TABLEAU 2.4: Taux d'estimation correcte de l'inclinaison obtenue avec les représentations temps-fréquence retenues dans cette étude. Le temps moyen correspond à la durée moyenne de traitement d'une image de 500×500 pixels.

Fourier de la racine carrée de l'histogramme de projection [Peake et Tan, 1997, Yin, 2001] est moins bonne que les estimations obtenues avec cinq des neuf représentations temps-fréquence retenues. Le tableau 2.5 illustre l'efficacité de notre algorithme sur un échantillon de 6 documents représentatifs (voir figure 2.11) choisis parmi les 864 documents traités.

Le tableau 2.4 présente la synthèse des résultats obtenus sur l'ensemble des 864 documents. Le taux d'estimation correct correspond au pourcentage de cas où l'inclinaison estimée est à $\pm 0.5\,^\circ$ de l'inclinaison exacte, intervalle correspondant aux besoins réels des algorithmes d'extraction de lignes et de reconnaissance automatique de caractères. Le temps moyen, mesuré en secondes, correspond à la durée de traitement d'une image de 500×500 pixels. Il dépend de la taille du document et de la représentation utilisée. Comme on peut le voir dans ce tableau, le meilleur taux d'estimation correcte est obtenu avec la distribution de Wigner-Ville et est égal à 100%. Les résultats obtenus avec les autres représentations temps-fréquence sont plus faibles, soit parce que celles-ci ne vérifient pas les propriétés qui nous semblent nécessaires pour cette application, comme les propriétés d'invariance par changement d'échelle et de conservation des supports, soit parce que la localisation des composantes du signal qu'elles fournissent n'est pas suffisante pour délivrer une estimation correcte de l'inclinaison par leur valeur maximale, soit parce que les paramètres de réglage de ces représentations ne sont pas adaptés. Ces paramètres sont

a) la fenêtre de lissage temporel $g(t)$;

b) la fenêtre de lissage fréquentiel $h(\tau)$;

c) la largeur σ du noyau de la distribution Choi-Williams.

Pour simplifier les comparaisons, toutes les fenêtres de lissage ont été choisies du type Kaiser-Bessel [James et Ronald, 1980]. Les meilleurs résultats ont alors été obtenus pour des longueurs de fenêtres de lissage temporel de 19 points et des longueurs des fenêtres de lissage fréquentiel de 91 points. De meilleurs résultats pourraient peut-être être obtenus en adaptant ces longueurs à la dimension des histogrammes, qui varie en fonction de l'angle de projection. La valeur du paramètre de réglage de la représentation de Choï-Williams qui a fourni les meilleurs résultats est $\sigma = 3.6$.

Le tableau 2.4 permet également de voir que sur les neuf représentations étudiées, six fournissent de meilleurs résultats que la méthode d'analyse des histogrammes de projection. Ceci montre que l'analyse temps-fréquence apporte bien une solution aux problèmes dûs à la présence de maximas locaux dans les histogrammes de projection. Le faible résultat obtenu avec le module carré de la transformée de Fourier semble montrer qu'il est important de s'appuyer sur le caractère non-stationnaire des histogrammes. Cette méthode convient sans doute mieux à des documents latins et/ou imprimés, pour lesquels l'inclinaison est beaucoup plus facile à estimer.

Nous avons également pu mettre en évidence que l'algorithme proposé fournit une estimation aberrante lorsque la hauteur de l'image est plus grande que sa largeur. Les histogrammes de projection dénombrent alors davantage de pixels dans le sens de la

hauteur que dans le sens de la largeur. La figure 2.12 montre que lorsque la largeur et la hauteur sont égales, l'estimation obtenue est correcte.

2.6 Conclusion

Nous avons proposé dans ce chapitre une approche originale qui permet d'estimer de manière robuste l'inclinaison d'un document ou d'une fenêtre d'un document arabe manuscrit numérisé. À cause des interlignes variables, des connexions et de la présence de hampes et de jambages dans les documents arabes manuscrits, ce problème d'estimation est difficile, et les techniques classiques conçues et utilisées principalement pour les caractères latins ne fournissent pas de résultats satisfaisants. L'approche proposée est basée sur la maximisation de la représentation temps-fréquence du signal analytique de la racine carrée des histogrammes de projection. Ceci permet de distribuer dans un plan "temps-fréquence" le nombre de pixels noirs de l'image et de chercher pour quel angle de projection cette population est la plus concentrée dans ce plan. En nous basant sur une liste de propriétés pertinentes pour ce problème, nous avons retenu neuf représentations temps-fréquence de la classe de Cohen. La distribution de Wigner-Ville s'est avérée fournir les meilleurs résultats. Ceci s'explique sans doute par le fait que cette représentation ne possède aucun paramètre de réglage, qu'elle fournit une localisation extrêmement concentrée des composantes des signaux et qu'elle vérifie deux propriétés importantes pour cette application : l'invariance par changement d'échelle et la conservation des supports.

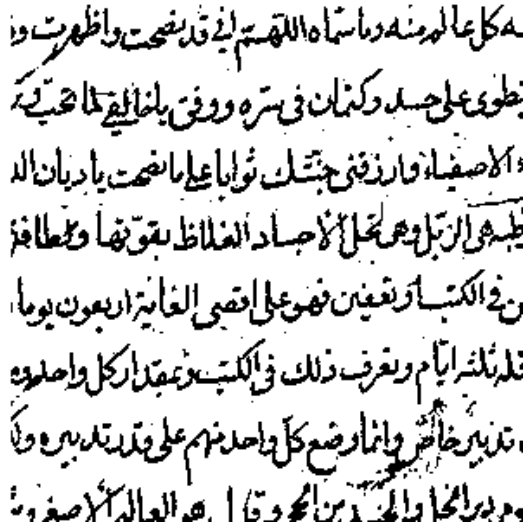


FIGURE 2.10: Partie d'un document arabe manuscrit. La largeur et la hauteur de cette image sont de 300 pixels et son inclinaison est de $0\,^{\circ}$.

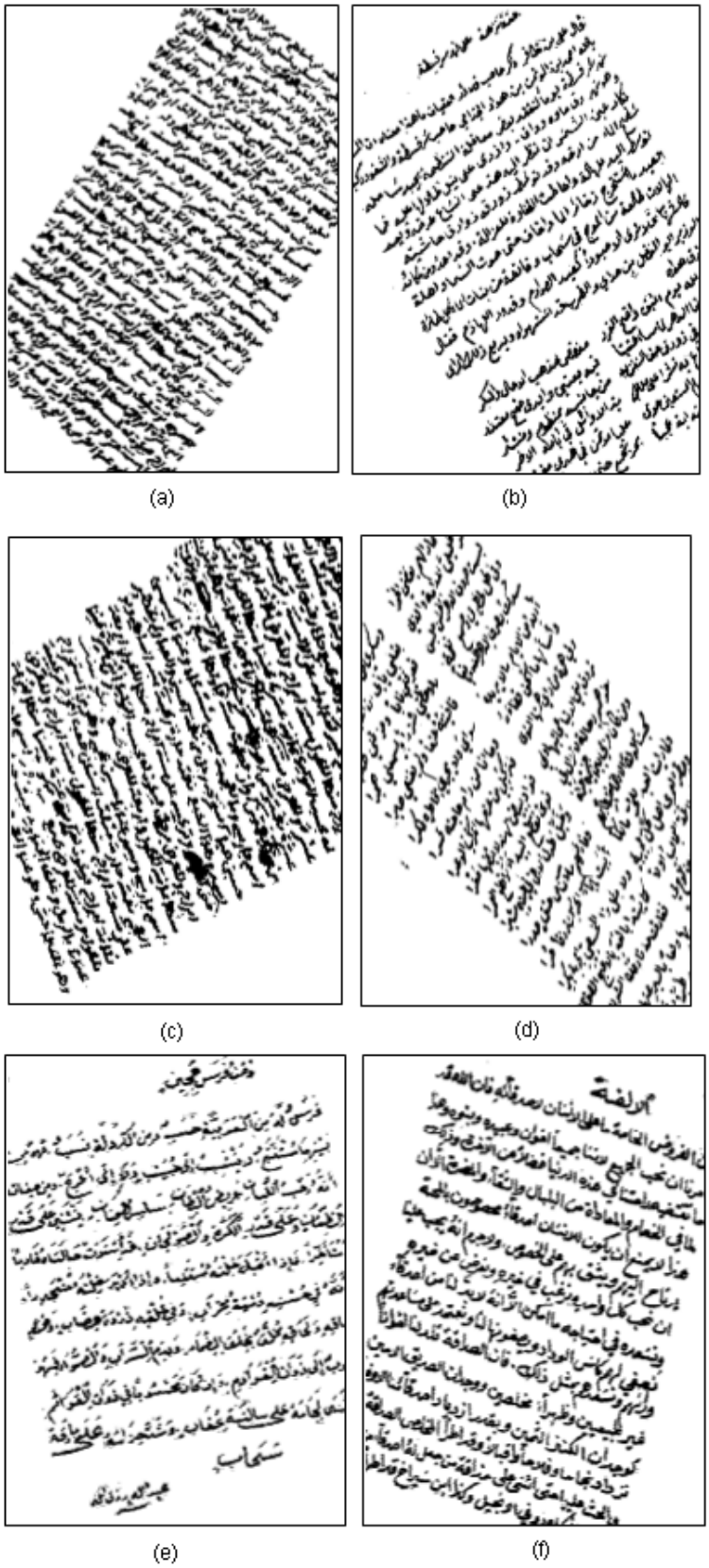


FIGURE 2.11: Échantillon de 6 documents parmi les 864 documents traités.

Documents de la figure 2.11		(a)	(b)	(c)	(d)	(e)	(f)
Taille du document (largeur × hauteur), en pixels		521*565	553*621	825*749	573*461	417*577	569*577
Résolution de la numérisation (dpi)		96	72	72	72	100	100
Angle exact		−36.0 °	+30.0 °	−67.0 °	+55.0 °	+10.0 °	−17.0 °
Wigner-Ville	angle estimé	−36.0 °	+30.0 °	−67.0 °	+55.0 °	+10.0 °	−17.0 °
	temps d'exécution (s)	5.1	5.9	10.0	6.0	4.6	5.3
Pseudo Wigner-Ville	angle estimé	−34.5 °	+26.0 °	−69.0 °	+57.0 °	+13.0 °	−19.0 °
	temps d'exécution (s)	7.3	7.9	15.0	8.0	7.4	7.6
Pseudo Wigner-Ville lissée	angle estimé	−37.0 °	+30.0 °	−67.0 °	+55.0 °	+11.0 °	−15.5 °
	temps d'exécution (s)	7.1	7.4	13.1	7.6	6.9	7.2
Margenau-Hill	angle estimé	−36.0 °	+30.0 °	−66.0 °	+55.0 °	+11.0 °	−17.0 °
	temps d'exécution (s)	7.6	8.1	12.1	8.5	7.1	8.5
Pseudo Margenau-Hill	angle estimé	−36.0 °	+29.5 °	−66.5 °	+55.5 °	+10.0 °	−17.5 °
	temps d'exécution (s)	8.4	7.1	16.1	7.1	8.3	8.3
Spectrogramme	angle estimé	−37.0 °	+30.5 °	−67.0 °	+51.0 °	+12.0 °	−15.0 °
	temps d'exécution (s)	8.6	8.0	14.3	7.9	8.9	8.5
Choi-Williams	angle estimé	−36.0 °	+30.0 °	−65.0 °	+55.0 °	+11.0 °	−17.0 °
	temps d'exécution (s)	7.9	8.6	13.2	8.5	8.2	7.8
Born-Jordan	angle estimé	−35.5 °	+28 °	−66.5 °	+53.5 °	+12.5 °	−19.5 °
	temps d'exécution (s)	8.4	8.8	15.2	8.4	8.4	8.1
Zhao-Atlas-Marks	angle estimé	−35 °	+30.5 °	−67.5 °	+55.5 °	+10.5 °	−17.5 °
	temps d'exécution (s)	9.1	9.6	16.3	9.7	8.9	9.4

TABLEAU 2.5: Estimation de l'inclinaison des six documents de la figure 2.11 obtenues à l'aide des neuf représentations temps-fréquence retenues dans cette étude.

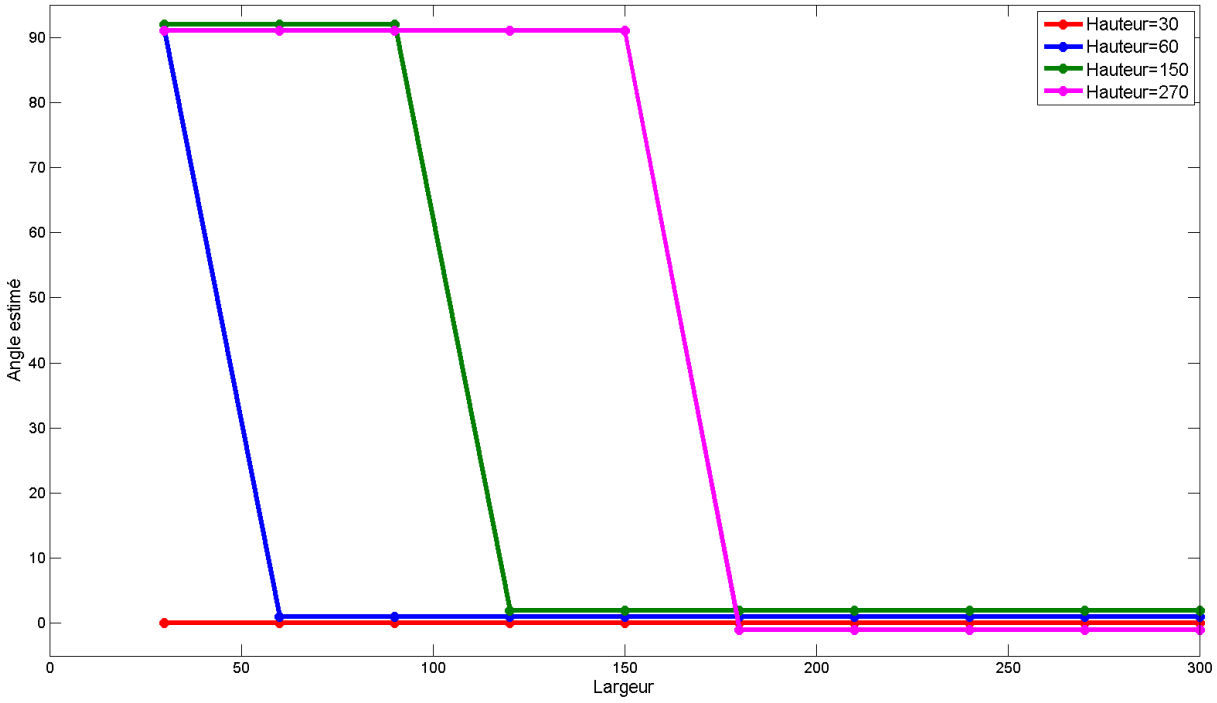


FIGURE 2.12: Évolution de l'estimation de l'angle d'inclinaison du document de la figure 2.10 en fonction de la largeur du document, pour quatre valeurs différentes de la hauteur. L'inclinaison est estimée en utilisant la distribution de Wigner-Ville.

Chapitre 3

Extension et correction des zones d'orientation

Ce chapitre est consacré à la détermination de zones globales d'orientation par fusion-fission de fenêtres d'orientation locales. Avant d'appliquer la méthode de fusion-fission, une étape de détection et de correction des orientations erronées dans certaines fenêtres est appliquée. Puis, la méthode est détaillée en montrant ses règles d'extension de l'orientation. Enfin, nous montrerons les résultats obtenus et nous terminerons ce chapitre par une conclusion.

Sommaire

3.1 Introduction 73
3.2 Détection de l'orientation erronée 74
3.3 Extension de l'orientation 75
3.4 Correction du maillage 76
3.5 Expérimentations 77
3.6 Conclusion 78

3.1 Introduction

Deux cas peuvent se produire après l'estimation de l'orientation. Le premier cas se présente lorsque l'orientation au sein d'une fenêtre n'est pas franche (orientation erronée c'est-à-dire la fenêtre contient plusieurs orientations), alors la fenêtre est découpée suivant la vallée de projection la plus profonde en deux fenêtres. Le deuxième cas se produit lorsque l'orientation est correcte.

Ensuite, chaque orientation de la fenêtre est revue en fonction de l'orientation des fenêtres voisines, prises suivant les règles d'écriture de l'Arabe : Est-Ouest (voir figure 3.1.a), Est-Sud-Ouest (voir figure 3.1.b), Sud-Nord et Nord-Sud (voir figure 3.1.c), Est-Nord-Ouest (voir figure 3.1.d). Par exemple, dans le cas c) où l'écriture est oblique, on exa-

minera les angles d'orientation dans les fenêtres θ_1, θ_2 et θ_3 qui correspondent à la suite de l'écriture Arabe. Soit, l'orientation est confirmée et étendue aux deux fenêtres, soit elle est gardée identique dans chaque fenêtre. Cette opération est suivie d'une étape de correction du fenêtrage qui consiste à décaler les limites des fenêtres suivant l'orientation principale, de manière à ne pas intercepter les lignes d'écriture.

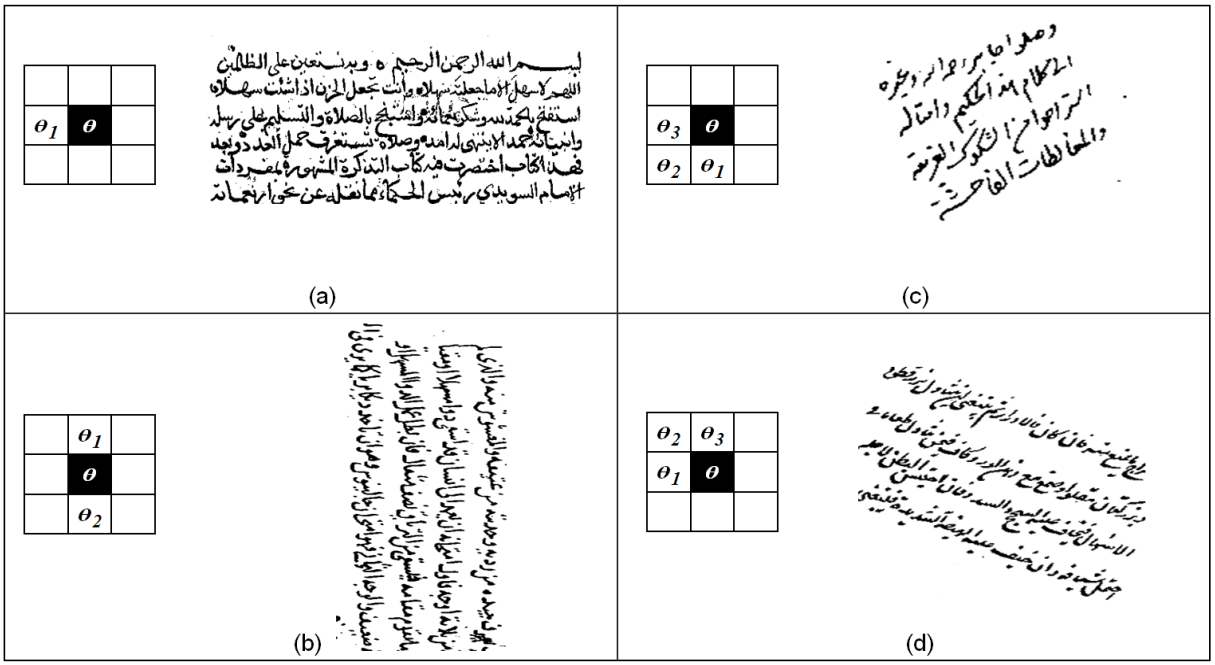


FIGURE 3.1: Les quatres règles d'extension de l'orientation vers les fenêtres voisines : (a) Est-Ouest, (b) Est-Sud-Ouest, (c) Sud-Nord et Nord-Sud et (d) Est-Nord-Ouest.

3.2 Détection de l'orientation erronée

Lorsqu'une fenêtre contient plusieurs écritures dans différentes orientations, l'estimation de l'orientation de la fenêtre θ sera erronée (voir figure 3.2.a). Pour détecter ce phénomène, on observe l'orientation θ_1 des fenêtres situées avant dans le sens de l'écriture (resp. en haut) et après θ_2 (resp. en bas).

L'orientation θ est erronée lorsque θ_1 est différent de θ, θ_2 est différent de θ et θ_1 est différent de θ_2. Dans la figure 3.2.a, $\theta = 60\,^\circ$, $\theta_1 = 0\,^\circ$ et $\theta_2 = 90\,^\circ$. Donc, l'orientation θ est erronée parce qu'elle vérifie les conditions ci-avant. Si ce phénomène survient pour une écriture où les fenêtres qui l'englobent étudiées situées avant et après, nous utilisons le profil de projection verticale pour corriger ce problème. Si ce cas arrive à côté d'une écriture où les fenêtres qui l'englobent étudiées situées en haut et en bas, nous utilisons le profil de projection horizontale pour résoudre ce problème. Pour cela, nous cherchons la valeur minimale dans le profil de projection qui représente la fin de la première inclinaison

(I_m : indice de la valeur minimale). Ensuite, nous divisons en I_m la fenêtre $w(\theta)$ en deux fenêtres w_{gauche} (resp. w_{haute}) et w_{droite} (resp. w_{basse}) (voir figure 3.2.b). Enfin, nous attribuons w_{droite} (resp. w_{basse}) à $w(\theta_1)$ et w_{gauche} (resp. w_{haute}) à $w(\theta_2)$. L'algorithme 3 résume l'application de la détection et la correction de l'orientation erronée.

Algorithme 3: Suivi des lignes dans chaque fenêtre.

Données : Orientation $\theta(i,j)$ de la fenêtre $w(i,j)$ où i est l'indice de la ligne et j l'indice de la colonne

Résultat : Deux fenêtres w_{gauche} et w_{droite}

si $\theta(i,j-1) \neq \theta(i,j)$ & $\theta(i,j+1) \neq \theta(i,j)$ & $\theta(i,j-1) \neq \theta(i,j+1)$ **alors**
- Calculer le profil de projection horizontale de $w(i,j)$;
- Calculer la valeur minimale I_m ;
- Diviser $w(i,j)$ en I_m en w_{gauche} et w_{droite} ;
- Fusionner w_{gauche} avec $w(i,j-1)$;
- Fusionner w_{droite} avec $w(i,j+1)$;

fin

si $\theta(i-1,j) \neq \theta(i,j)$ & $\theta(i+1,j) \neq \theta(i,j)$ & $\theta(i-1,j) \neq \theta(i+1,j)$ **alors**
- Calculer le profil de projection verticale de $w(i,j)$;
- Calculer la valeur minimale I_m ;
- Diviser $w(i,j)$ en I_m en w_{haute} et w_{basse} ;
- Fusionner w_{haute} avec $w(i-1,j)$;
- Fusionner w_{basse} avec $w(i+1,j)$;

fin

3.3 Extension de l'orientation

L'orientation de chaque fenêtre est étendue à ses fenêtres voisines afin de globaliser l'orientation dans des zones plus importantes que les fenêtres. Pour cela, nous avons proposé une méthode qui utilise quatre règles d'extension correspondant aux sens d'écriture vus précédemment. L'extension est faite de la droite vers la gauche en respectant le sens de l'écriture Arabe. Nous partons de la fenêtre en haut à droite vers la fenêtre en bas à gauche de la zone étudiée. Pour chaque fenêtre, les orientations de ses fenêtres voisines sont observées. Les deux fenêtres sont fusionnées en une seule, puis l'orientation de cette nouvelle fenêtre est réestimée par WVD sur le profil de projection. Si l'orientation de la nouvelle fenêtre est égale à l'orientation de l'une des deux premières fenêtres, la fusion est conservée. Dans le cas contraire, nous conservons l'état initial des deux fenêtres et nous continuons le travail sur les autres paires de fenêtres. Après cette étape, les zones d'orientation seront déterminées. La figure 3.3 montre le résultat final d'extraction des zones d'orientation. On retrouve les quatre zones d'orientation qui correspondent à la zone

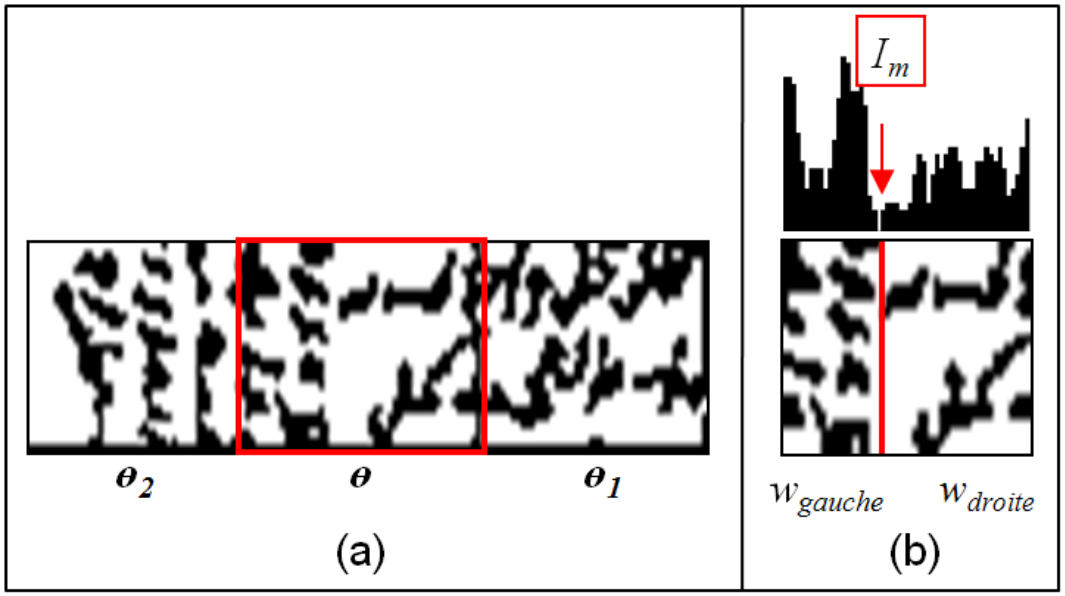


FIGURE 3.2: Algorithme de correction de l'orientation : (a) fenêtre avec une orientation erronée et (b) correction de l'orientation.

principale et à celles des annotations dans les marges.

3.4 Correction du maillage

Comme le maillage est appliqué automatiquement, les bords du premier maillage peuvent traverser les composantes connexes ce qui crée un problème (faux maxima) lors de l'étape du suivi des lignes. Comme le maillage utilisé est carré, le maillage incorrect arrive seulement dans les zones horizontales et verticales (voir la figure 3.4.avant). Nous corrigeons ce problème en procédant à un décalage des bords horizontaux ou verticaux par fenêtre, afin que la totalité des composantes connexes se trouvent dans la même fenêtre. Dans la zone horizontale (resp. verticale), le bord qui divise deux lignes (resp. colonnes) consécutives est déplacé vers la position la plus proche dans ces lignes (resp. colonnes) où le profil de projection horizontale (resp. verticale) pour chacune de ses fenêtres consécutives atteint sa valeur minimale (voir figure 3.4 et figure 3.5).

Il va de soi que cette opération reste approximative car nous ne pourrons l'effectuer avec exactitude vu l'irrégularité des lignes. Elle permet de corriger un peu le chevauchement quand il est réellement flagrant.

3.5 Expérimentations

Pour étudier l'efficacité de cette approche, nous l'avons testée sur la même base de 36 documents. Le taux d'extraction des zones dépend de trois algorithmes : l'algorithme de détection et de correction des orientations erronées, l'algorithme de fusion-fission et l'algorithme de correction de maillage automatique.

Pour l'algorithme de détection et de correction des orientations erronées, le taux d'erreur provient des cas où la division de fenêtres en fenêtres plus petites n'est pas faite au bon endroit. Cela peut arriver lorsque la valeur minimale de l'histogramme de projection arrive loin de la frontière de deux orientations.

Concernant l'algorithme de fission-fusion, on n'a aucune erreur parce qu'il traite toutes les orientations possibles dans les documents arabes.

Le taux d'erreur de l'algorithme de correction de maillage automatique est dû au fait que seuls les bords horizontaux et verticaux sont déplacés. Dans le cas où l'orientation est oblique, ce type de correction est moins efficace.

Le tableau 3.1, représente les taux d'extraction de zones d'orientation pour chaque classe. On remarque que ce taux augmente lorsque la complexité des documents diminue. Dans l'annexe B (tableau B.1), on présente des résultats sur un échantillon de 8 documents.

Classe	**Nombre de zones d'orientation**	**Nombre de zones détectées**	**Taux d'extraction %**
Documents simples	25	24	96.67 %
Documents multi-orientés	46	44	96.44%
Documents multi-tailles	8	7	91.67%
Documents avec problèmes au centre	9	8	91.67%

TABLEAU 3.1: Résultats de l'extraction des zones d'orientation.

Dans la figure 3.6, on présente un échantillon de 4 documents arabes très complexes où la détection des zones est erronée. Dans le document (a), les zones d'orientations s'interfèrent ce qui donne des fenêtres avec des orientations multiples et rend l'estimation de l'orientation très difficile. Pour le document (b), les lignes de texte ne sont pas bien alignées. On voit un nuage des composantes connexes dans toutes les directions et pas des zones des lignes. Dans le document (c), l'orientation dans le centre de document n'est pas horizontale, ce qui crée une limite pour l'algorithme du maillage automatique qui impose que l'écriture au centre soit horizontale. Le document (d), contient aux centres des lignes discontinues ce qui crée une limite pour l'algorithme du maillage. Ainsi, le document contient des lignes obliques (en bas à gauche), ce qui crée des fenêtres avec des orientations erronées. On rappelle que l'approche proposée ne convient pas pour l'écriture oblique.

3.6 Conclusion

Nous avons proposé dans ce chapitre une méthode pour l'extension et la correction de l'orientation dans les fenêtres du maillage. L'orientation dans chaque fenêtre est analysée. Si l'orientation est avérée correcte, on la garde et on passe à l'étape d'extension de l'orientation où plusieurs règles sont appliquées pour trouver les zones multi-orientées. Dans le cas contraire, nous appliquons un algorithme basé sur l'analyse de profil de projection (horizontal ou vertical) afin de diviser cette fenêtre. Après la détection des zones multi-orientées, nous corrigeons le maillage initial afin d'éviter le problème des faux maxima lors de l'étape de suivi des lignes du texte.

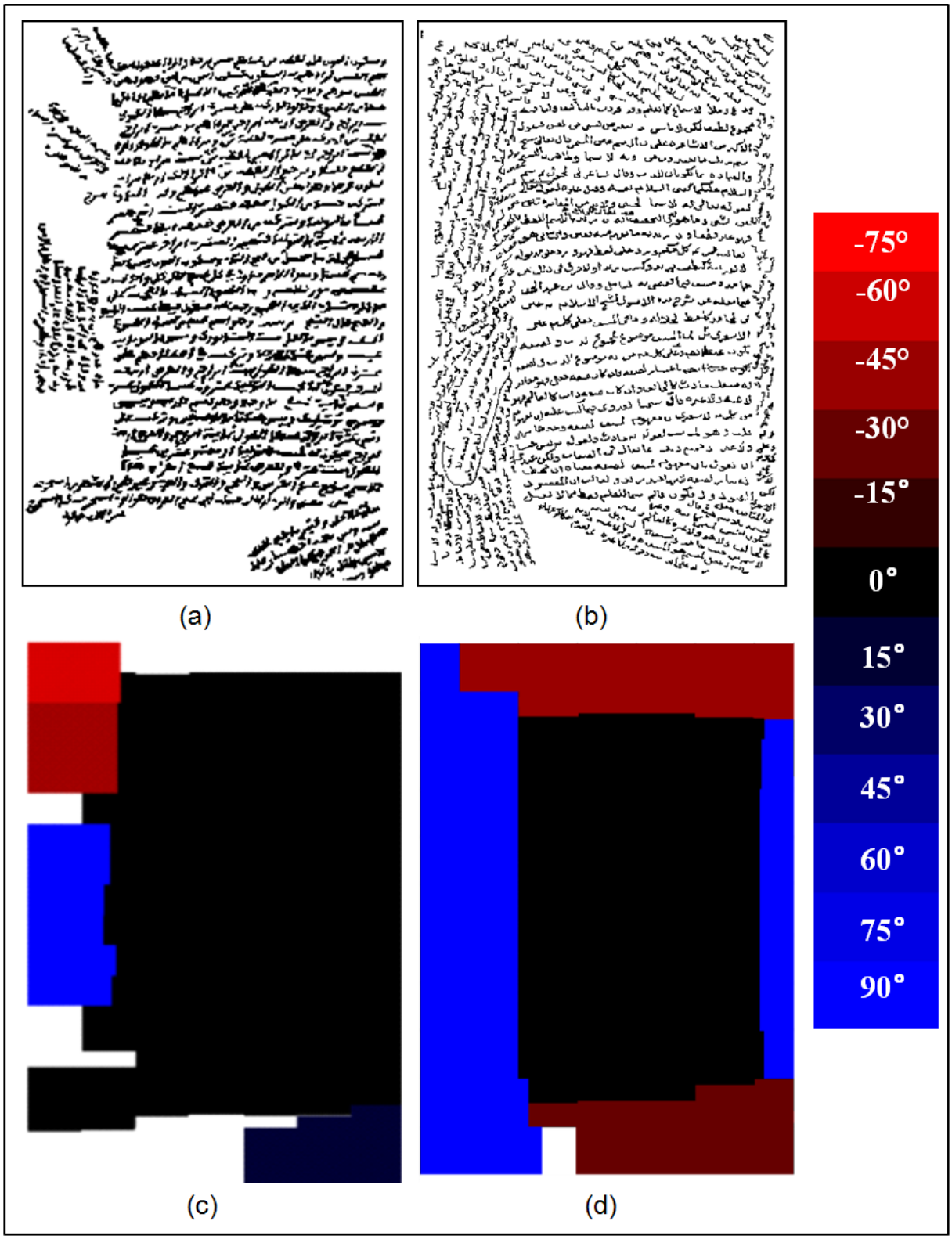


FIGURE 3.3: Résultats d'extension de l'orientation. Chaque couleur représente une zone d'orientation.

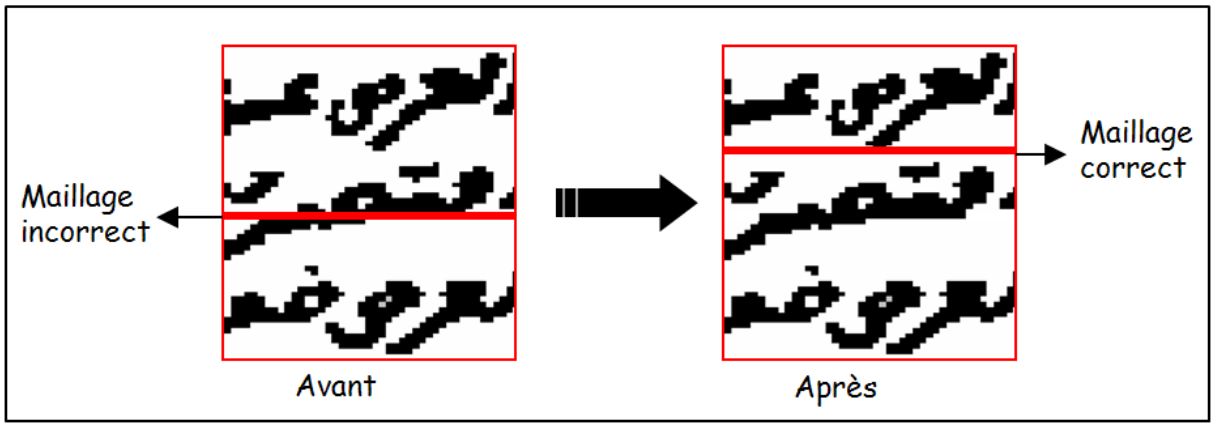


FIGURE 3.4: Exemple d'un maillage incorrect dans une zone horizontale.

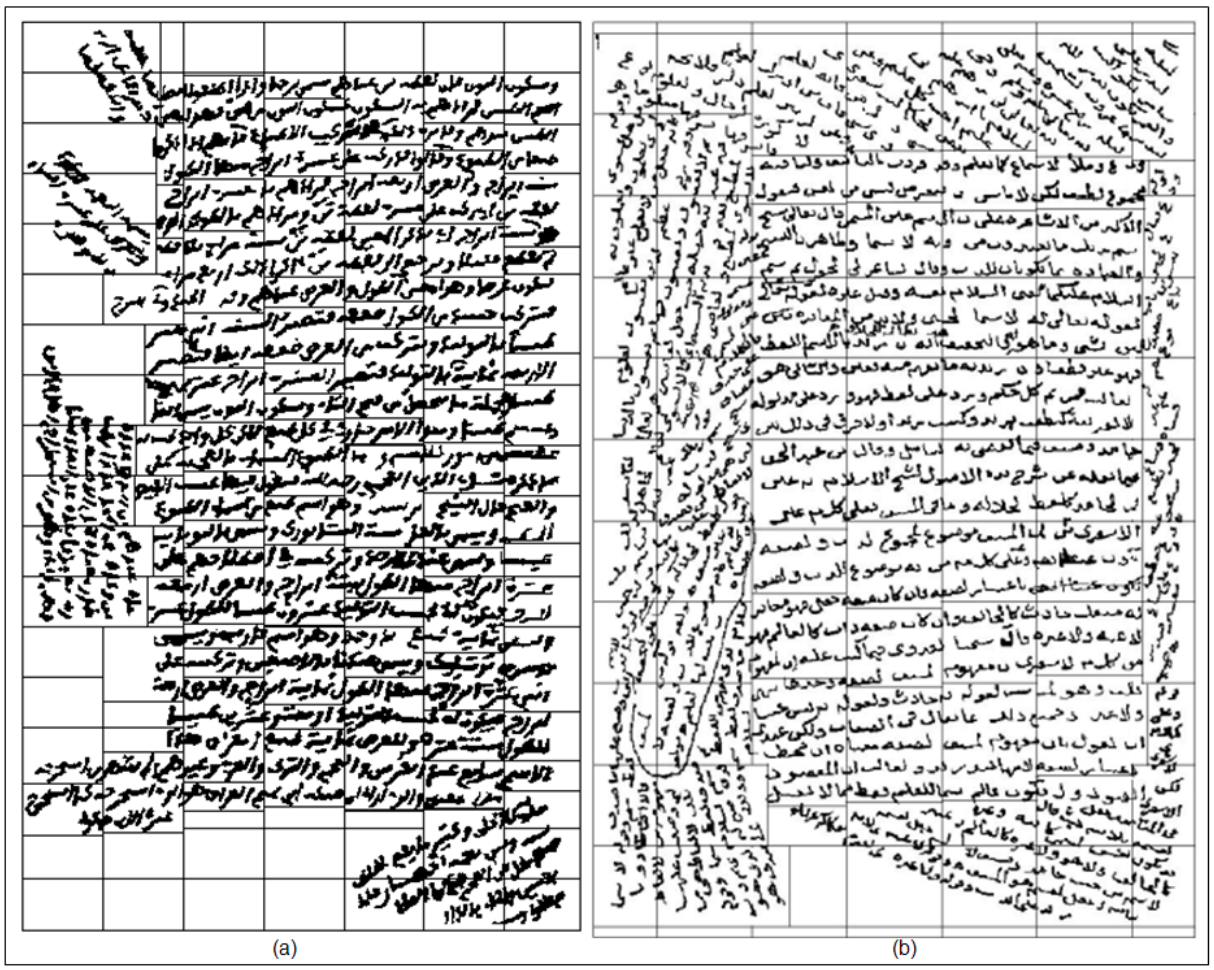


FIGURE 3.5: Résultats de la correction du maillage pour les documents représentés dans la figure 1.2.a et b.

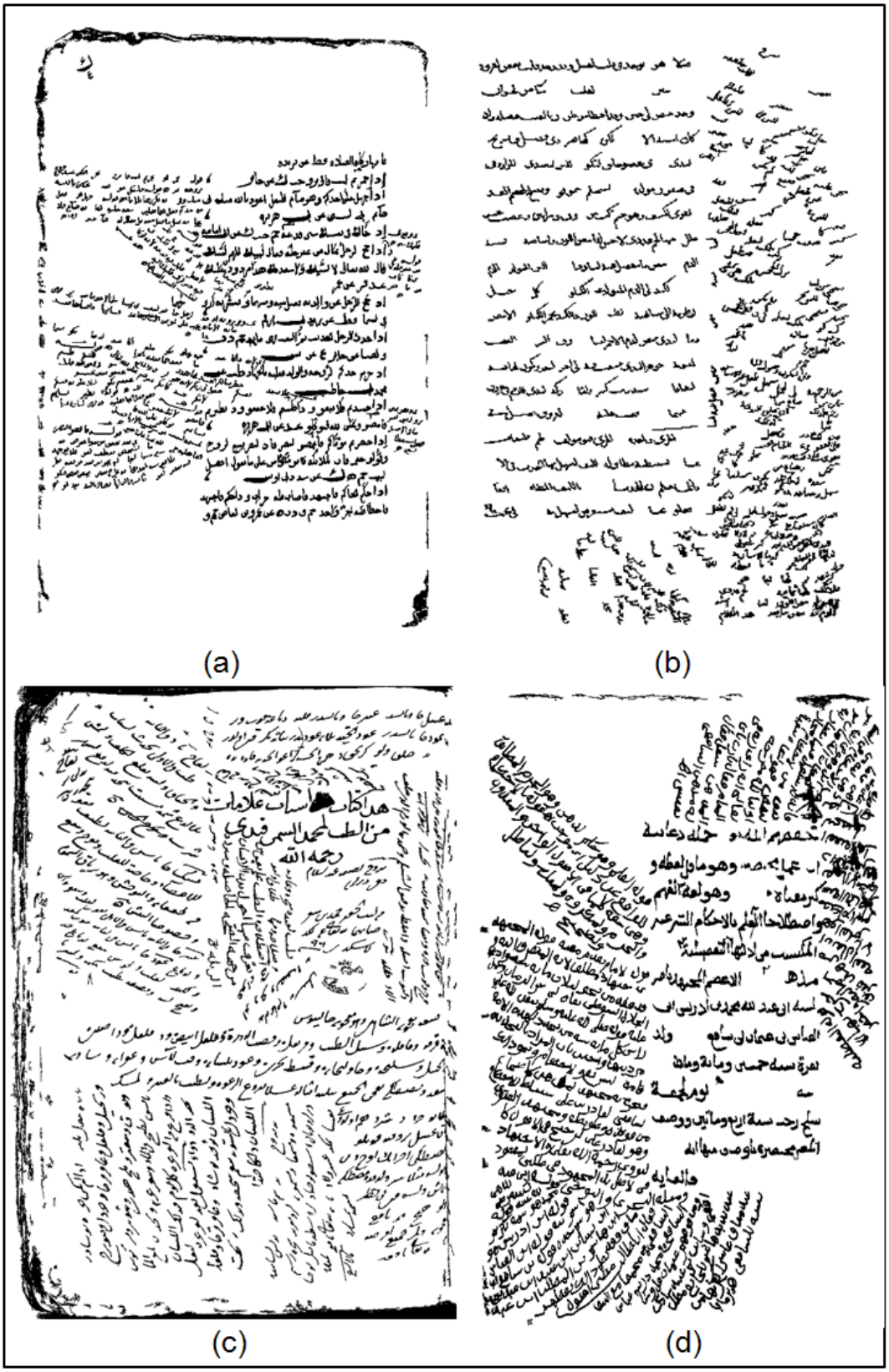


FIGURE 3.6: Échantillon de 4 documents très complexes où l'approche proposée trouve ses limites.

Chapitre 4

Suivi de lignes

Nous présentons dans ce chapitre, une méthode d'extraction des lignes. Tout d'abord, nous détaillons l'algorithme de suivi des composantes connexes qui s'appuie sur les orientations précédemment trouvées. Ensuite, nous discuterons les résultats obtenus lors des différentes expérimentations. Enfin, nous terminerons ce chapitre par une conclusion.

Sommaire

4.1 Introduction 83
4.2 Algorithme 84
4.3 Expérimentations 85
4.4 Conclusion 88

4.1 Introduction

La littérature abonde de méthodes d'extraction de lignes. Nous retrouvons ici les techniques basées sur l'analyse du profil de projection, la transformée de Hough, la transformée de Fourier, le suivi des composantes connexes, le lissage (Run-length), etc. Toutes ces techniques ont été sommairement décrites dans le chapitre 2 de la première partie. Dans notre cas, nous proposons une méthode adaptative basée sur l'analyse du profil de projection et le suivi de composantes connexes. D'abord, le profil de projection est déterminé pour chaque fenêtre selon l'orientation estimée dans les étapes précédentes. Puis, les lignes d'orientation sont calculées en vue de définir le chemin de suivi. Nous calculons alors les composantes connexes et nous les suivons afin de trouver les lignes. Une ligne contient toutes les composantes connexes se succédant suivant la même orientation. Après ce suivi, une méthode de correction est appliquée afin de trouver les composantes connexes qui n'appartiennent à aucune ligne. Cette méthode exploite la distance minimale entre les composantes connexes et les lignes d'orientation.

4.2 Algorithme

En nous basant sur l'orientation dans les fenêtres, nous calculons le profil de projection par rapport à cette orientation, puis on procède à la recherche de maxima (voir figure 4.1.a). Chaque maxima représente le point de départ d'une ligne Ps, à partir duquel nous suivons la ligne d'orientation blj en respectant l'angle d'orientation. Le suivi commence dans la première fenêtre, dans le coin droit-haut du document (pour l'arabe). Le point d'arrivé Pe de la ligne d'orientation est calculé en utilisant, l'angle, la largeur et la longueur de chaque fenêtre (voir figure 4.1.b). La ligne d'orientation blj est calculée en utilisant les deux points (Ps, Pe) et l'orientation de la fenêtre. Pendant ce suivi, les composantes connexes qui appartiennent à une ligne d'orientation blj sont recherchées pour former les lignes dans chaque fenêtre (voir figure 4.1.c).

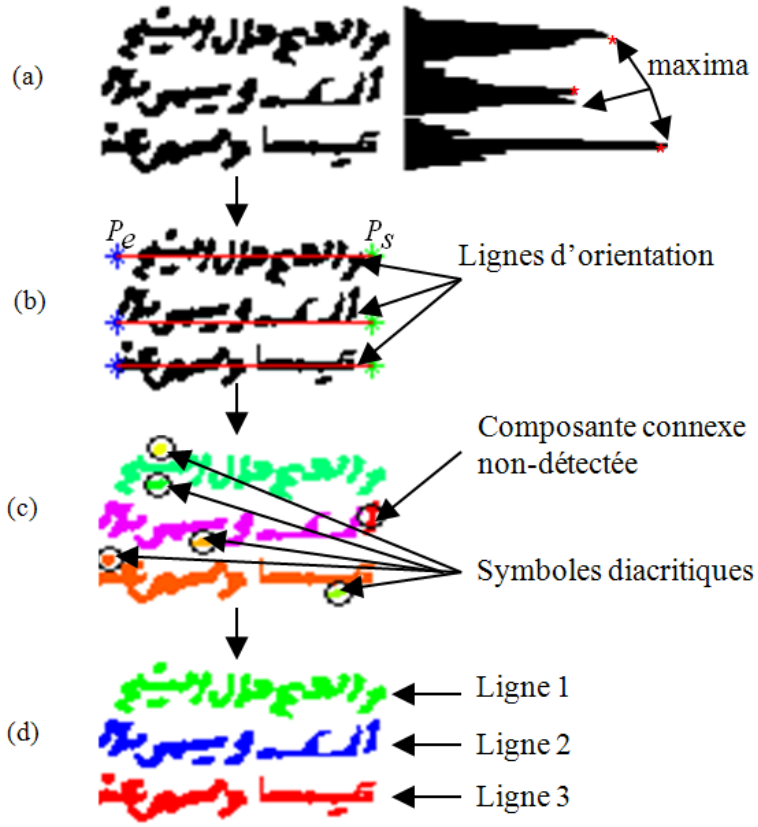


FIGURE 4.1: Étapes de détection de lignes dans une fenêtre.

Une étape de correction de la détection complète cette méthode. Elle attribue les composantes connexes non détectées et les symboles diacritiques à la ligne la plus proche. La distance est calculée entre le centre de gravité de la composante ou du symbole à traiter et la ligne candidate (voir figure 4.1.c et figure 4.1.d). Les étapes de suivi de lignes sont présentées dans l'algorithme 4.

Pour chaque zone, les lignes sont regroupées et nous regardons si elles sont bien détachées les unes des autres (voir figure 4.3).

Algorithme 4: Suivi des lignes dans chaque fenêtre.

Données : Orientation θ_i de la fenêtre w_i
Résultat : Composantes connexes de chaque ligne

Calculer les composantes connexes (CCXs) de w_i ;
Calculer le profil de projection de w_i par rapport à θ_i ;
Calculer les maxima de ce profil ;
j←1 ;
pour *chaque maxima* **faire**
  Calculer les points de départ P_s et d'arrivé P_e ;
  Calculer la ligne d'orientation bl_j ;
  j←j+1 ;
fin
j←1 ;
pour *chaque ligne* bl_j **faire**
  Suivre les pixels de la ligne bl_j ;
  Trouver les composantes connexes qui passent par cette ligne ;
  Stocker ces composantes dans le vecteur l_j ;
  j←j+1 ;
fin
pour *chaque CCX* **faire**
  si *CCX* $\notin$ *à* l **alors**
    Calculer les distances d_{c_j,l_j} et d_{c_j,l_j+1} ;
    si $d_{c_j,l_j} < d_{c_j,l_j+1}$ **alors**
      CCX $\in$ à l_j ;
      sinon
        CCX $\in$ à l_{j+1} ;
      fin
    fin
  fin
fin

Pendant cette étape, nous rencontrons le problème de connexion entre les lignes de texte adjacentes (voir figure 4.3).

4.3 Expérimentations

Pour étudier l'efficacité de notre approche, nous l'avons testée sur la même base de 36 documents arabes manuscrits anciens contenant 932 lignes.

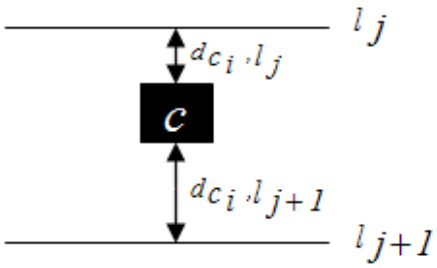


FIGURE 4.2: Attribution des composantes connexes et des symboles diacritiques à leur ligne.

Nous avons étudié deux types de résultats : les résultats qualitatifs et les résultats quantitatifs. Les résultats qualitatifs sont mesurés selon deux critères : la sur-segmentation lorsque la ligne extraite contient plus de composantes connexes que ses composantes réelles et la sous-segmentations lorsque la ligne extraite ne contient pas toutes les composantes réelles qui la forment. Les résultats quantitatifs seront mesurées par : 100×(nombre de lignes extraites) /(nombre de lignes réelles trouvées dans le document).

Concernant la qualité d'extraction, nous avons obtenu un taux d'extraction moyen (pour les 4 classes) de 3.6% de sur-segmentation et 2.1% de sous-segmentation. Le taux de sur-segmentation est dû à la fluctuation des lignes par rapport à la ligne d'orientation. Les lignes sont très proches, ce qui implique que pendant l'étape de suivi des lignes, certaines composantes connexes qui appartiennent à la ligne l_{i+1} ou à la ligne l_{i-1} sont affectés incorrectement à la ligne l_i. La sous-segmentation vient de la présence massive de symboles diacritiques dont les positions sont loin de la ligne directrice de l'écriture. La détection de ces symboles est difficile et dans la plupart des cas, nous utilisons une méthode de distance pour les attribuer aux lignes. Cette méthode de distance a toujours des effets négatifs parce qu'elle dépend de paramètres empiriques.

Concernant les résultats quantitatifs, les taux d'extraction sont présentés dans le tableau 4.1. À cause du maillage et de l'algorithme d'extension de l'orientation, nous arrivons dans certains documents à manquer certaines lignes qui n'appartiennent pas à une zone d'orientation.

Classe	**Nombre de lignes**	**Nombre de lignes extraites**	**Taux d'extraction %**
Documents simples	302	297	98.35 %
Documents multi-orientés	490	471	96.24%
Documents multi-tailles	80	52	65.07%
Documents avec problèmes au centre	60	53	91.97%

TABLEAU 4.1: Résultats de l'extraction des zones d'orientation.

Ces taux d'extraction montrent l'efficacité et la performance de l'approche proposée. L'Annexe B (tableau B.2) présente des résultats sur un échantillon de 8 documents.

FIGURE 4.3: Résultats de l'algorithme d'extraction de lignes (les cercles bleus indiquent les régions de connexion).

À cause de l'absence dans la littérature de méthode qui traite des bases similaires de documents, nous avons procédé à une synthèse des résultats obtenus par notre méthode et cells trouvées dans la littérature (voir tableau 4.2).

4.4 Conclusion

Nous avons proposé dans ce chapitre une méthode pour l'extraction des lignes orientées. Cette méthode est basée sur l'analyse du profil de projections orientée et le suivi des composantes connexes. L'orientation estimée de chaque fenêtre est utilisée comme l'élément de base pour construire les profils de projection orientées. Pendant ce suivi, nous remarquons qu'il manque des composantes connexes qui n'appartiennent pas à une ligne. Ces composantes connexes sont attribuées à la ligne d'orientation la plus proche. Cette méthode est performante et rapide parce qu'elle ne nécessite pas beaucoup de paramètres ou de calcul.

Auteurs	**Base de test**	**Taux d'extraction**
[Nicolaou et Gatos, 2009]	20 documents contenant 476 lignes et 80 documents de 1771 lignes	98.6%
[Bennasri *et al.*, 1999]	100 documents manuscrits arabes mono-orientés contenant 2000 lignes	98%
[Shapiro *et al.*, 1993]	200 documents manuscrits latins	n.c.
[Antonacopoulos et Karatzas, 2004]	Registres de la seconde guerre mondiale	n.c.
[Nagy *et al.*, 1988]	Articles latins imprimés	n.c.
[Lu *et al.*, 2004]	904 documents latins imprimés	99%
[Likforman-Sulem et Faure, 1994]	56 documents (35 brouillons de texte et 21 enveloppes postales)	82% et 96.3%
[Zahour *et al.*, 2004]	une dizaine de documents manuscrits arabes	n. c.
[Zahour *et al.*, 2007]	100 anciens documents manuscrits arabes	96%
[Zahour *et al.*, 2008]	160 anciens documents manuscrits arabes (120 avec une petite écriture et 40 avec une grande écriture)	96.6% et 100%
[Feldbach et Tönnies, 2001]	7 documents anciens contenant 61 paragraphes et datant des années 1649 (2 documents), 1727, 1741, 1812, 1817 et 1838	97%
[Likforman-Sulem *et al.*, 1995]	plusieurs documents, y compris des brouillons, des blocs d'adresses et des lettres	n.c.
[Pu et Shi, 1998]	plus de 100 documents manuscrits latins	n.c.
[Malleron *et al.*, 2009]	200 documents manuscrits (lettres, textes imprimés, notes, dictionnaires, recueillies)	88%
[Louloudis *et al.*, 2009a]	80 documents contenant 1773 lignes et 40 anciens documents manuscrits contenant 1095 lignes	97.4%
[Shi et Govindaraju, 2004]	1864 documents des colis USPS	93%
[Bourgeois *et al.*, 2001]	20 livres latins imprimés du $16^{ème}$ siècle	n.c.
[Oztop *et al.*, 1999]	plusieurs documents anciens arabes (Ottman) et latins	n.c.
[F. Yin, 2008]	plusieurs documents manuscrits chinois comportant 8664 lignes de texte de la base des données HIT-MW [Su *et al.*, 2007]	95.02%
[Nicolas *et al.*, 2004]	plusieurs documents manuscrits de Gustave FLAUBERT	n.c.
[Bukhari *et al.*, 2009]	80 documents contenant 1770 lignes	96.3%
[Li *et al.*, 2008]	2691 lignes de texte arabe, 1672 lignes de texte chinois, 1365 lignes de texte indien et 1932 lignes de texte coréen	84.6% (textes arabe), 92% (textes chinois), 95% (textes indien) et 96% (textes coréen).
[Du *et al.*, 2009]	100 documents chinois, 96 documents Indien et 100 documents Coréens appartiennent à l'Université de Maryland	98%(documents chinois), 98% (documents indien) et 96% (documents coréen).
Approche proposée	**36 anciens documents manuscrits arabes contenant 932 lignes : 15 documents simples contenant 302 lignes, 15 documents multi-orientés contenant 490 lignes, 3 documents multi-tailles contenant 80 lignes et 3 documents avec problèmes au centre contenant 60 lignes.**	**98.35% (documents simples), 96.24% (documents multi-orientés), 65.07% (documents multi-tailles) et 91.97% (documents avec problèmes au centre**

TABLEAU 4.2: Synthèse des résultats obtenus par les méthodes proposées dans la littérature pour la segmentation en lignes de documents. (n.c. : non connu).

Chapitre 5

Séparation des lignes connectées

Ce chapitre présente la méthode de séparation des lignes adjacentes connectées par considération de la morphologie des lettres terminales des PAWs. Tout d'abord, les types de connexion traités sont énumérés. Ensuite, la morphologie de l'écriture Arabe est discutée et notre approche de séparation de lignes est détaillée. Enfin, les expérimentations sont présentées et une conclusion termine ce chapitre.

Sommaire

5.1 Introduction 91
5.2 Les types de connexion 93
5.3 L'analyse de la morphologie Arabe 93
5.4 Direction correcte des lettres connectées 94
5.5 Application 95
5.6 Expérimentations 98
5.7 Conclusion 99

5.1 Introduction

L'extraction des lignes et des mots d'un document manuscrit, est considérée comme une tâche difficile. La difficulté vient des caractéristiques des documents manuscrits en particulier lorsqu'ils sont anciens. Ces documents présentent des espacements irréguliers entre les lignes du texte. Les lignes peuvent se connecter lorsque les ascendants et descendants touchent les lignes voisines (voir figure 5.1).

Dans la littérature peu de méthodes ont été proposées pour la séparation de mots appartenant à deux lignes adjacentes. Une méthode de segmentation par analyse en composantes indépendantes (ACI) est proposée par Chen et al. dans [Chen et Leedham, 2005]. [Hyvarinen, 1999] l'ACI convertit les mots originaux connectés dans une matrice aveugle de sources et calcule la matrice de la valeur pondérée avant que les valeurs soient réévaluées en utilisant un modèle rapide (ACI rapide). La valeur de la matrice réévaluée est

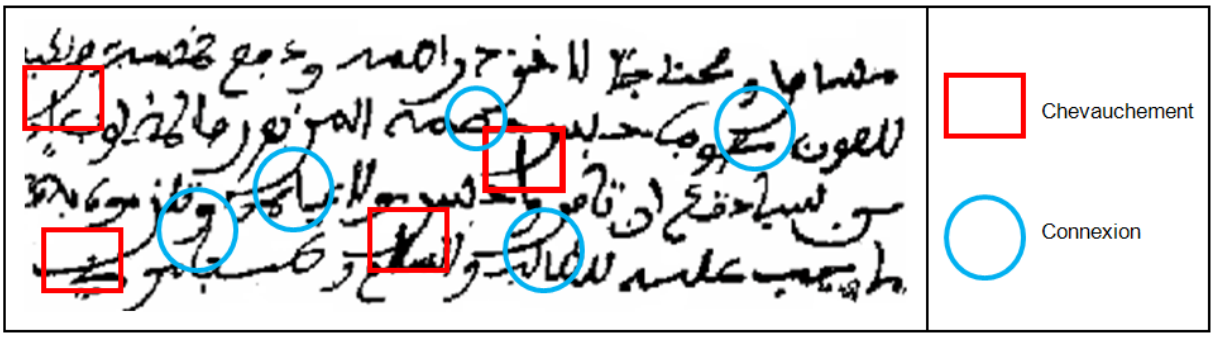


FIGURE 5.1: Extrait d'un document représentant quelques lignes qui se connectent.

appliquée à la matrice de source aveugle afin de séparer les mots.

Louloudis et al. ont proposé, dans [Louloudis *et al.*, 2009b] un algorithme basé sur l'analyse du squelette des mots connectés. Cet algorithme recherche les points d'intersection par un algorithme de suivi selon la morphologie de l'écriture latine.

Dans [k. Takru et Leedham, 2002], Takru et Leedham proposent une technique qui utilise les connaissances structurelles de l'écriture où la connexion est fréquemment observée. Cette approche généralise les cas de connexion, puis compare les descendants des composantes connexes connectées avec ceux des cas possibles. L'approche sépare ensuite les composantes connexes si les lignes sont identiques aux cas déjà définis.

Toutes ces approches sont applicables sur le Latin mais elles ne sont pas facilement adaptables pour l'arabe en raison de la morphologie spécifique de l'écriture arabe. La forme des lettres terminales arabe n'est pas la même que celle du Latin et elle peut varier d'une écriture à une autre (voir figure 5.2). Cela empêche des méthodes comme le matching (ACI) et le suivi de squelette d'être appliquées ici.

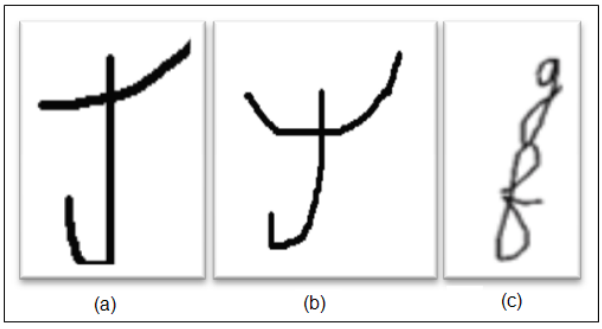


FIGURE 5.2: Cas de connexion entre deux lettres : (a) Arabe, (b) même lettres que (a) avec une autre écriture (c) Latin.

5.2 Les types de connexion

L'alphabet arabe est composé de 28 lettres. Parmi elles, 21 ont un ascendant (comme alif ا, Ta ط, DHa ظ, kaf ك, lam ل, heh ه), un descendant droit (comme ra ر, zïn ز, sïn س, shïn ش, Sad ص, Dad ض, qaf ق, mïm م, nün ن, waw و, yeh ي) ou un descendant gauche (comme jïm ج, ḥa ح, kha خ, ‘ain ع, ghain غ) provoquant la connexion entre les lignes. Ces descendants se dessinent soit de la droite vers la gauche soit de la gauche vers la droite.

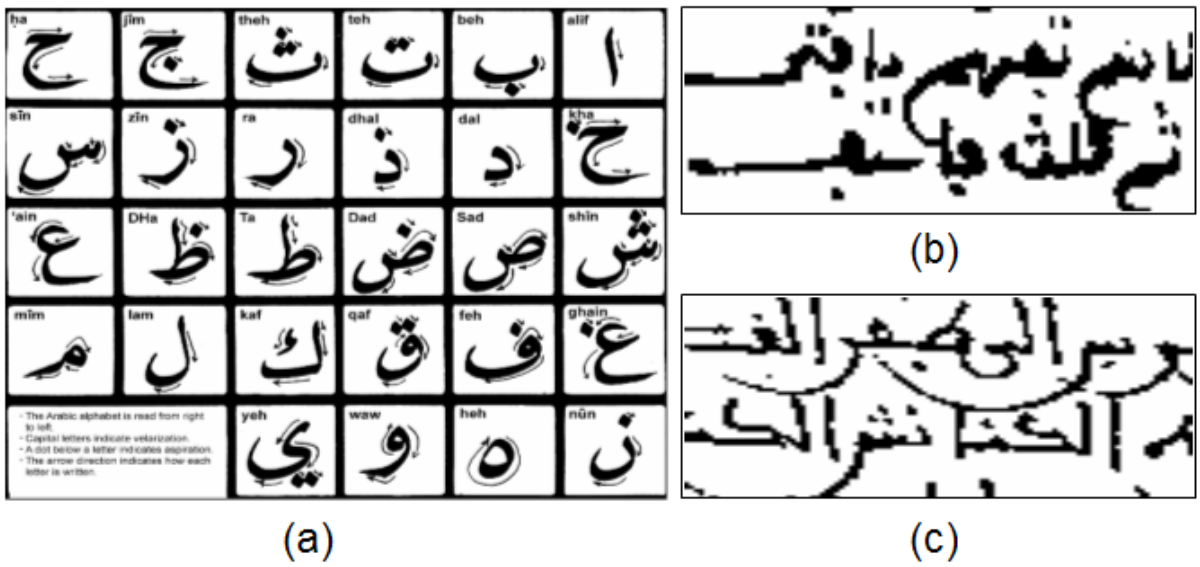


FIGURE 5.3: (a) L'alphabet Arabe (la flèche indique la direction de l'écriture), (b) exemple de connexion et (c) connexion causée par le type du script (Diwani).

Parmi les cas de connexion existants, nous en avons retenu quatre dont nous pourrions automatiser la résolution. Concernant les autres cas, la compréhension du texte est nécessaire pour les défaire. Le premier cas (voir tableau 5.1.a), correspond aux lettres avec descendants courbes orientés vers la droite. Le deuxième cas (voir tableau 5.1.b), correspond aux lettres avec descendants courbe orientés vers la gauche. Dans le troisième cas (voir tableau 5.1.c), un descendant droit se connecte avec un ascendant courbe. Dans le quatrième cas (voir tableau 5.1.d), un descendant courbe gauche se connecte avec un ascendant courbe.

5.3 L'analyse de la morphologie Arabe

Dans tous les cas de connexion recensés, nous remarquons la présence d'un descendant connectant un ascendant (voir tableau 5.1, colonne zone de connexion). Les descendants

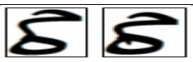
Type	Lettre terminale	Zone de connexion	Exemple
a	Haut : ر ,ز ,س ,ش ,ص ,ض ,ن ,ق ,و ,ي		
	Bas : ا ,ط ,ظ ,ك ,ل		
b	Haut : ج ,ح ,خ ,ع ,غ		
	Bas : ا ,ط ,ظ ,ك ,ل		
c	Haut : ر ,ز ,م ,و		
	Bas : ص ,ض ,ه		
d	Haut : ج ,ح ,خ ,ع ,غ		
	Bas : ه		

TABLEAU 5.1: Les 4 types de connexion observés dans les documents Arabe manuscrits.

sont regroupés en deux catégories : (a, c) lorsque le descendant démarre de la droite et (b, d) lorsque le descendant débute de la gauche.

La méthode consiste à d'abord identifier la zone de connexion, de repérer les points potentiels de connexion, les lettres ou les parties de lettres connectées dans la zone, et enfin de suivre ces lettres ou parties de lettres pour les détacher les unes des autres.

On définit la zone de connexion comme un rectangle autour du point d'intersection S_p de deux composantes connexes. La taille de cette zone doit être assez grande pour contenir les lignes de base des lignes adjacentes d'où démarrent les connexions (voir section 5.4). Le point de départ de la ligature supérieure B_p est le point le plus élevé dans la zone la plus proche de la ligne de base. La direction du descendant est déterminée en fonction de B_p par rapport à S_p (voir section 5.4).

Considérant ces caractéristiques, notre idée consiste à suivre les pixels du squelette des composantes connexes dans la zone en utilisant le point de départ B_p et la direction du descendant droit. Le suivi traverse le point d'intersection S_p et continue dans la bonne direction que nous devons le déterminer.

5.4 Direction correcte des lettres connectées

Pour séparer les composantes connexes, l'idée est de démarrer depuis le point haut de la ligature (proche de la ligne de base) et de suivre le contour de la lettre jusqu'à la fin, en traversant les points d'intersection. On se sert de la courbure du contour qui doit être progressivement croissante tout au long du suivi. Ce critère permettra d'éviter de suivre d'autres bouts de contours appartenant à d'autres lettres.

On se sert du squelette pour le suivi et de la variation angulaire pour déterminer un tracé. Comme le montre la figure 5.4.a, le suivi commence à partir du point de départ de la connexion, B_p, et se poursuit jusqu'au point d'intersection S_p. À ce point S_p, plusieurs directions sont possibles. Le choix est effectué en fonction de la courbe qui représente le descendant de la lettre terminale. Ceci s'explique par le fait que les terminales des lettres arabes ont la même orientation le long de la courbe du descendant.

Variance angulaire : la variance angulaire représente la dispersion des angles d'orientation le long de la courbe. Elle est estimée en utilisant la formule de la variance statistique :

$$Var(\Theta) = \sum_{i=1}^{n} (\theta_i - \mu)^2 \tag{5.1}$$

où Θ est le vecteur de variation des angles de la courbe et μ est la moyenne de Θ.

Le vecteur de variation des angles θ de la courbe est estimé en utilisant un algorithme itératif. Cet algorithme calcule l'angle d'orientation θ_i entre deux pixels p_i et p_{i+2} à l'aide de la formule suivante (voir figure 5.4.b). Le choix de $i + 2$ est fait expérimentalement parce qu'il a donné les meilleurs résultats :

$$\theta_i = \left| \text{Arctan} \left(\frac{\text{dy}_{\text{i,i+2}}}{\text{dx}_{\text{i,i+2}}} \right) \right| \tag{5.2}$$

Par exemple, dans la figure 5.4, les variances angulaires sont : Var $(C_{1+2}) = 703,19$, $Var(C_{1+3}) = 299$, $Var(C_{1+4}) = 572,37$. Dans cet exemple, la variance angulaire minimale $Var(C_{1+3})$ est donnée par le descendant.

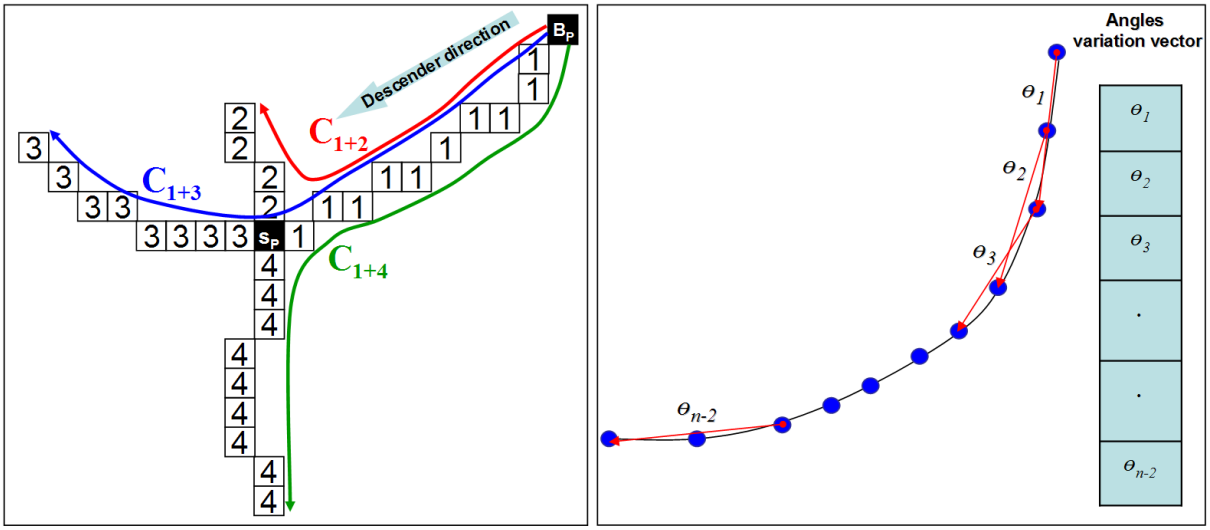


FIGURE 5.4: (a) Exemple de composantes connexes arabes se touchant (la lettre Arabe "ra ر" touche la lettre "alif ا") et (b) algorithme de calcul du vecteur des variations angulaires.

5.5 Application

L'application proposée se déroule en quatre étapes qui sont :

- **Étape 1 : détection des composantes connexes connectées :**
  Dans l'étape précédente de l'extraction des lignes, les composantes connexes qui forment chaque ligne sont extraites. Une composantes connexe qui appartient à deux lignes adjacentes est considérée comme cas de connexion (voir figure 5.5.a).
- **Étape 2 : détection de la courbe :**
  Pour détecter les courbes, le squelette est d'abord extrait en utilisant un algorithme d'éclaircissage décrit dans [Lam *et al.*, 1992, León *et al.*, 2004]. Ensuite, les points d'intersection de chaque composante connexe sont détectés (voir figure 5.5.b). Un point d'intersection est un pixel qui a au moins trois pixels voisins. Comme dans l'écriture Arabe, la connexion se produit en un seul point d'intersection S_p près de l'axe minimum (vallée entre deux lignes, voir la figure 5.5.c), le point S_p est donc le point le plus proche de l'axe minimum (voir figure 5.5.d).
  Une fois le point S_p trouvé, nous recherchons la zone des composantes connexes. Le centre de cette zone est S_p et sa largeur (resp. hauteur) est égale à $w_{ccx}/4$ (resp. $h_{ccx}/4$) où w_{ccx} (resp. h_{ccx}) est la largeur (resp. hauteur) de la composante connexe connectée (4 est déterminé expérimentalement). Étant donné que cette zone est extraite du document initial, elle est nettoyée en enlevant les petites composantes connexes. Pour faire cela, les composantes connexes sont marquées, et nous ne gardons que la composante connexe qui contient le point d'intersection S_p (voir figure 5.5.e). Après le pré-traitement, le squelette de la zone est extrait et l'algorithme de suivi de squelette est appliqué en utilisant le point de départ du descendant B_p (le pixel qui a $min(y_i)$ dans la zone) et la direction de suivi (de droite à gauche si $x(B_p) > x(S_p)$ et de gauche à droite si $x(B_p) < x(S_p)$, voir figure 5.5.f).
- **Étape 3 : estimation de la variance angulaire de la courbe :**
  La variance angulaire de chaque courbe est estimée en utilisant l'algorithme détaillé dans la section 5.4. Dans la figure 5.5.f, la première composante connectée a $V(C_1) = 538.2099$ et $V(C_2) = 754.2284$. La courbe C_1 ayant la variance angulaire minimale correspond au descendant. La deuxième composante chevauchée a $V(C_1) = 1160.8$, $V(C_2) = 438.4$ et $V(C_3) = 1208$. La courbe C_2 ayant la variance angulaire minimale est le descendant.
- **Étape 4 : attribution des pixels :**
  Dans la figure 5.5.g, il existe deux courbes et trois types différents de pixels (le point d'intersection "en rouge", la première composante connexe en "vert" et la deuxième composante connexe en "bleu"). Cette étape consiste à attribuer chaque pixel noir dans l'image initiale (voir figure 5.5.e) à sa courbe appropriée. Pour faire cela, l'image est balayée pixel par pixel et les 48-pixels connectés de chaque pixel de l'image p_i sont examinés dans le squelette de l'image. La valeur du pixel des branches les plus proches est attribuée au pixel de l'image initiale (voir figure 5.5.h). Enfin, l'attribution des pixels est appliquée au document initial afin d'obtenir les lignes séparées (voir figure 5.5.i).

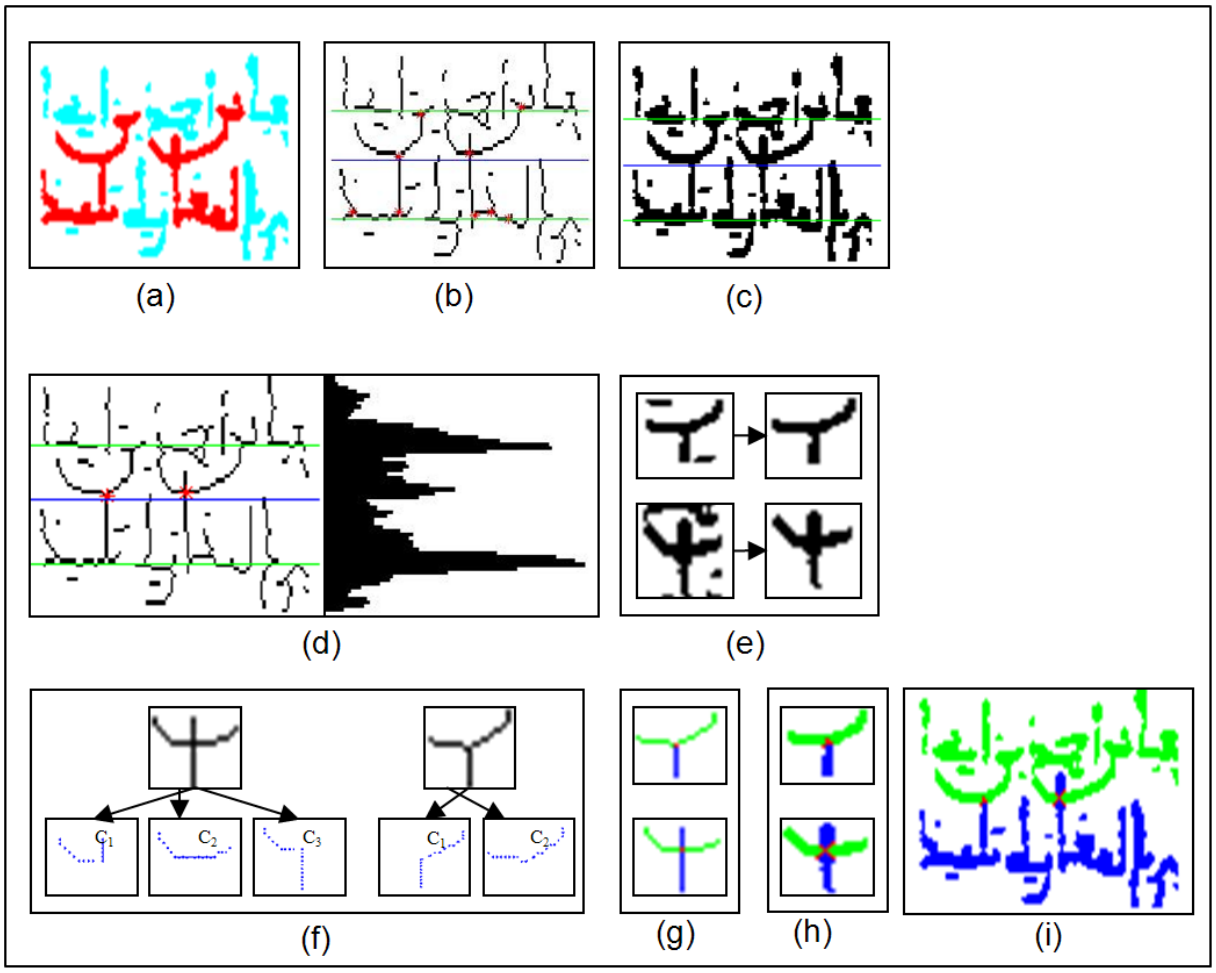


FIGURE 5.5: Les étapes de l'application de séparation des composantes connectées.

5.6 Expérimentations

L'approche a été appliquée sur nos 36 manuscrits arabes contenant 256 cas de connexion. Les tests ont été élaborés après une étape d'indexation manuelle des composantes connectées. Ensuite, ces composantes sont réparties dans les 4 classes (voir la section 5.2) : 103 occurrences de type (a), 77 occurrences de type (b), 46 occurrences de type (c), 30 occurrences de type (d). Le tableau 5.2 décrit les résultats pour chaque classe. La moyenne pondérée des taux de détection est égale à 94.90%. Le taux d'erreur de 5.10% est dû à l'algorithme de détection du point d'intersection parce que, dans certains cas, la connexion ne se tient pas près de l'axe des minima. Il est également dû au critère de variance angulaire qui, dans certains cas, fait que la variance angulaire minimale peut se produire pour une direction incorrecte (voir figure 5.6).

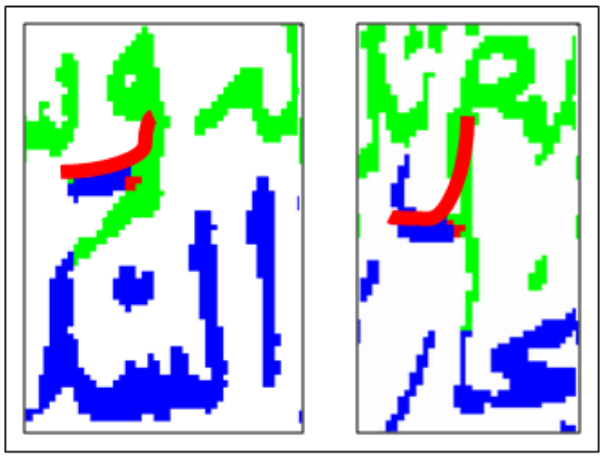


FIGURE 5.6: Cas de séparation incorrecte (le flèche rouge indique le vrai sens de suivi qui n'est pas atteint.

La figure 5.7 illustre l'efficacité de l'algorithme sur un échantillon représentatif de 12 composantes connexes choisies arbitrairement à partir de 256 occurrences.

Type de connexion	**Occurrence**	**Nombre de connexions manquées**	**Nombre de séparations non réalisées**	**Taux de séparation correcte%**
a	103	1	1	98.05%
b	77	3	2	93.50%
c	46	2	1	93.40%
d	30	1	2	90.00%

TABLEAU 5.2: Résultats de l'approche de la séparation des composantes connectées.

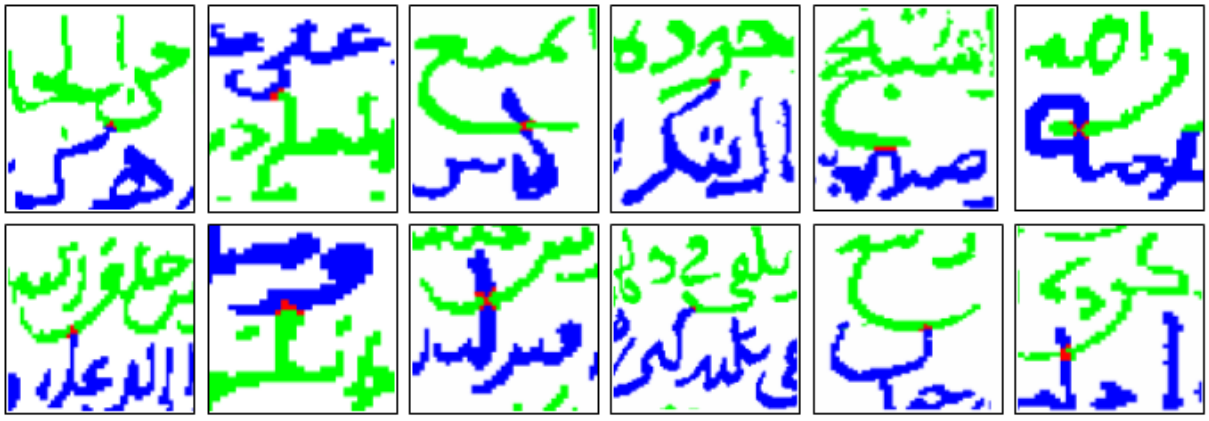


FIGURE 5.7: Exemples de résultats de séparation de composantes connectées.

5.7 Conclusion

Une méthode originale de séparation des composantes connexes connectées dans les lignes des documents arabes manuscrits a été proposée. La méthode utilisée tient compte de la calligraphe arabe conduisant à la connexion des lettres terminales. Des expériences ont montré l'efficacité et la performance de notre approche.

Conclusion

Nous avons présenté dans cette partie l'approche qui vise à extraire les lignes multi-orientées des documents arabes manuscrits anciens. D'abord, les zones multi-inclinées sont détectées en utilisant un maillage automatique du document. Après, l'orientation est estimée, corrigée et étendue afin de trouver toutes les orientations locales en utilisant la distribution de Wigner-Ville sur l'histogramme de projection. Ensuite, les lignes sont extraites en se basant sur l'orientation des lignes de base de chaque fenêtre. Enfin, les lignes adjacentes connectées sont séparées en utilisant des informations statistiques sur la morphologie des lettres terminales arabes.

Conclusion et perspectives

Nous avons abordé au cours de cette thèse le problème de la segmentation des documents manuscrits arabes multi-orientés. Le travail réalisé concerne l'extraction des lignes dans ces documents. La difficulté de cette tâche vient des caractéristiques des documents étudiés. Ces documents sont anciens et complexes. Ils contiennent beaucoup de bruit qui vient de l'ancienneté de ces documents et de la présence massive des symboles diacritiques. Ajoutons à ces problèmes la présence des annotations dans les marges. Ces annotations sont obliques à cause du manque d'espace. De plus, les lignes présentent des fluctuations par rapport à la ligne horizontale. Ces lignes peuvent se chevaucher et se connecter du fait du type d'écriture qui occasionne des réductions d'espace, et donc des artefacts de type sinuosité, multi-orientation et rattachement (e.g. Diwani).

Devant tous ces problèmes, nous avons proposé une méthode générique capable d'extraire toutes les orientations locales. À cause de la multi-orientation et la petite taille des lignes ajoutées, nous avons procédé à un pavage de l'image afin de mieux cerner les orientations locales et les localiser dans des fenêtres séparées. Ensuite, les lignes seront extraites dans chaque fenêtre et séparées.

La méthode proposée a été testée sur une base de données comportant 36 anciens documents arabes manuscrits complexes contenant 932 lignes. Nous n'avons pas pu faire la comparaison avec une autre méthode à cause de l'absence dans la littérature d'une méthode qui étudie ce type des documents. Notre méthode est complètement automatique et ne demande aucune intervention lors de la segmentation. Elle exige seulement que les documents soient binaires et nettoyés du bruit de numérisation.La méthode proposée nécessite le déploiement de plusieurs techniques. Après l'indexation de nos documents, nous avons étudié deux types de résultats : les résultats qualitatifs et les résultats quantitatifs. Par rapport à la qualité d'extraction, nous avons obtenu un taux d'extraction moyen de 3.6% de sur-segmentation et 2.1% de sous-segmentation. Par rapport à la quantité, le taux d'extraction atteint varie entre 65.07 et 98.35 %. Ces taux d'extraction montrent l'efficacité et la performance de l'approche proposée.

Les taux d'extraction qui dépassent pour certaines classes 96% nous encouragent à continuer le travail sur les points suivants :

– Essayer de trouver d'autres bases de données et plus de documents pour tester l'efficacité de l'approche proposée.

- Segmenter les lignes extraites en mots. Cette tâche nous paraît un peu difficile et demande beaucoup de réflexion parce que les caractères qui appartiennent à deux mots successifs se chevauchent et se connectent entre eux.
- Tester l'approche proposée sur d'autres types de scripts comme le Latin, l'Urdu, le Farsi et le Chinois.
- Appliquer l'approche sur des documents hétérogènes qui contiennent plusieurs types de scripts.
- Orienter l'approche vers la segmentation de documents qui contiennent un mélange de texte et d'image. L'idée du maillage automatique peut être adaptée pour la séparation des images de texte.

Annexe A

Outils de développement

Dans cette thèse, nous avons utilisé l'environnement de développement "Matlab" parce qu'il offre aux chercheurs scientifiques des outils en traitement d'image et du signal. Cet environnement m'a fait gagner beaucoup de temps et m'a permis de réaliser mes tests d'une manière très large et très rapide. Aussi, j'ai développé une interface graphique nommée "OuwayedImaPro" qui contient toutes les fonctions de bases de traitement d'image et mes algorithmes de segmentation des documents. Cette annexe a pour objectif de présenter "Matlab", mon interface graphique "OuwayedImaPro" et l'outil temps-fréquence "TFTB".

A.1 Matlab

[MAT, 2009] MATLAB est un logiciel commercial de calcul numérique, de visualisation et de programmation très performant et convivial développé par la société The MathWorks. Attention : ce n'est cependant pas un logiciel de calcul algébrique ou symbolique (pour cela, voir les logiciels commerciaux Mathematica ou Maple, ou le logiciel libre Maxima).

Le nom de MATLAB vient de MATrix LABoratory, les éléments de données de base manipulés par MATLAB étant des matrices (pouvant bien évidemment se réduire à des vecteurs et des scalaires) qui ne nécessitent ni dimensionnement ni déclaration de type. Contrairement aux langages de programmation classiques (scalaires et à compiler), les opérateurs et fonctions MATLAB permettent de manipuler directement et interactivement ces données matricielles, rendant ainsi MATLAB particulièrement efficace en calcul numérique, analyse et visualisation de données en particulier.

Mais MATLAB est aussi un environnement de développement "progiciel" à part entière : son langage d'assez haut niveau, doté notamment de structures de contrôles, fonctions d'entrée-sortie et de visualisation 2D et 3D, éditeur/debugger, outils de construction d'interface utilisateur graphique (GUI)... permet à l'utilisateur d'élaborer ses propres fonctions ainsi que de véritables programmes "M-files" appelés scripts vu le caractère

interprété de ce langage.

MATLAB est disponible sur tous les systèmes d'exploitation standards (Windows, Unix/Linux, MacOS X...) et son architecture est relativement ouverte. Le champ d'utilisation de MATLAB peut être étendu aux systèmes non linéaires et aux problèmes associés de simulation avec le produit complémentaire SIMULINK. Les capacités de MATLAB peuvent en outre être enrichies par des fonctions spécialisées regroupées au sein de dizaines de "toolboxes" (boîtes à outils qui sont des collections de "M-files") couvrant des domaines très variés tels que :

- analyse de données,
- statistiques,
- mathématiques symboliques (accès au noyau Maple V),
- analyse numérique (accès aux routines NAG),
- traitement d'image, cartographie,
- traitement de signaux (et du son en particulier),
- acquisition de données et contrôle de processus (gestion ports série/parallèle, cartes d'acquisition, réseau TCP ou UDP), instrumentation,
- logique floue,
- finance,
- etc...

Une interface de programmation applicative (API) rend finalement possible l'interaction entre MATLAB et les environnements de développement classiques (exécution de routines C ou Fortran depuis MATLAB, ou accès aux fonctions MATLAB depuis des programmes C ou Fortran). MATLAB permet en outre de déployer de véritables applications à l'aide des outils de conversion optionnels suivants :

- MATLAB → code C/C++, avec le MATLAB Compiler.
- MATLAB → Excel add-ins, avec le MATLAB Excel Builder.
- MATLAB → objets COM Windows, avec le MATLAB COM Builder.
- etc...

Toutes ces caractéristiques font aujourd'hui de MATLAB un standard incontournable en milieu académique, dans les différents domaines de l'ingénieur et la recherche scientifique.

A.2 Notre système

OuwayedImaPro est un système pour le traitement d'images, la segmentation des documents et la reconnaissance de formes. Il contient toutes les fonctions de bases pour le traitement d'images (lecture, écriture, binarization, extraction des composantes connexes, rotation, etc..), les fonctions de reconnaissance de formes (modèle de contour actif : snake, descripteur de Fourier, transformée de Hough, transformée de Fourier etc..) ainsi que les fonctions nécessaires pour la segmentation de documents (estimation de l'orientation

par les distributions de la classe de Cohen, maillage automatique en utilisant le snake, extraction des zones multi-orientées, suivi des lignes, séparation des lignes connectées, variance angulaire, suivi des pixels, etc..). La figure A.1 montre notre interface graphique.

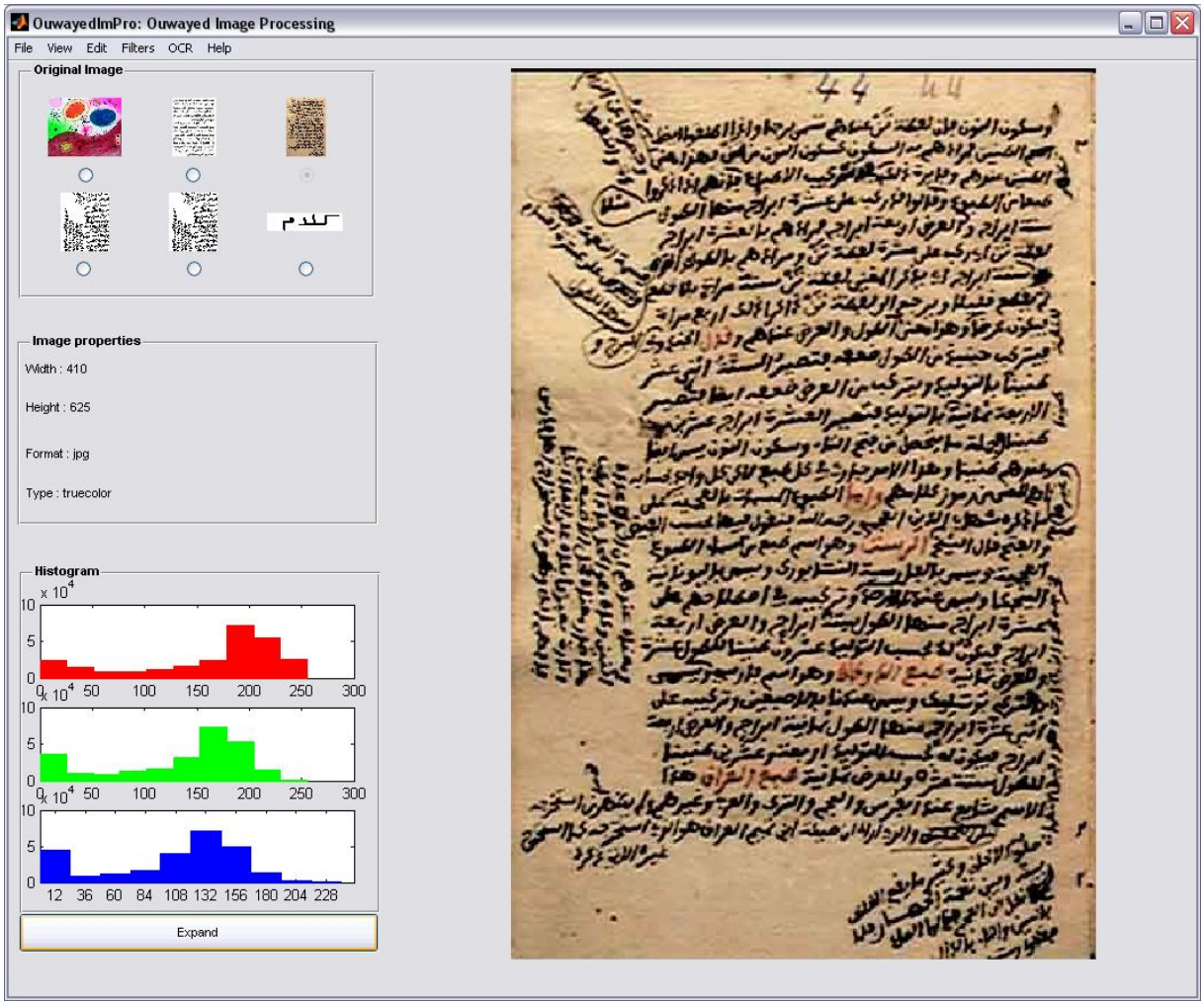


FIGURE A.1: Notre système de traitement d'image et de segmentation des documents "OuwayedImaPro" pour "**Ouwayed Im**age **Pro**cessing".

A.3 Outil temps-fréquence "TFTB"

Nous avons intégré dans notre système **OuwayedImaPro** la boîte à outils d'analyse temps-fréquence "TFTB" **??** afin de pouvoir faire les tests pour notre algorithme d'estimation de l'orientation.

[Auger *et al.*, 1996] La boîte à outils temps-fréquence (Time-Frequency Toolbox en anglais, TFTB) est une collection de quelques 100 scripts pour GNU Octave et Matlab

développée pour l'analyse de signaux non-stationnaires à l'aide de distributions temps-fréquence. Elle est destinée à des chercheurs, des ingénieurs et des étudiants ayant des connaissances de base en traitement du signal.

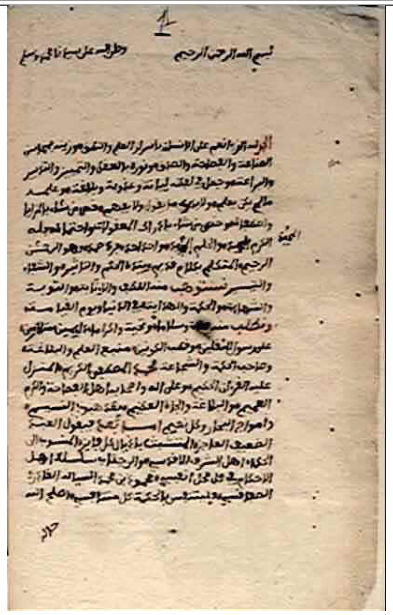
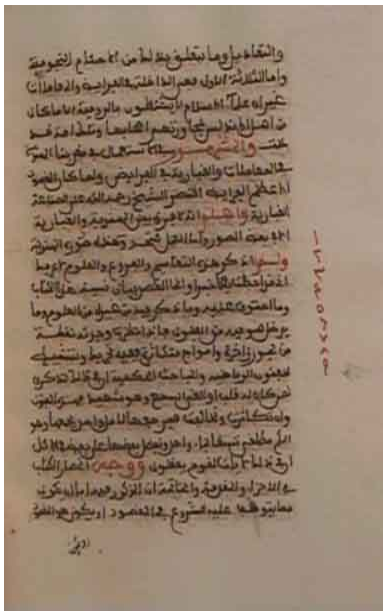
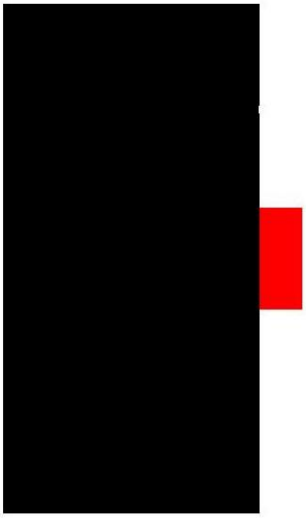
La boîte à outils contient de nombreux algorithmes calculant divers types d'analyse temps-fréquence, en particulier les distributions quadratiques de la classe de Cohen et de la classe affine ainsi que leur version améliorée par la méthode de réallocation. La boîte à outils inclut également des procédures de simulation des signaux et de traitements et post-traitements (pour l'affichage) et quelques démonstrations illustrent son fonctionnement. La TFTB est distribuée sous les conditions définies par la Licence Publique Générale GNU.

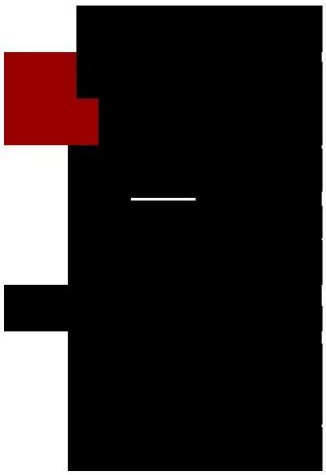
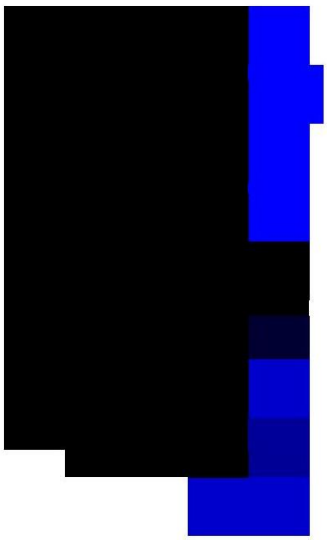
Cette boîte a été développée par François Auger, Olivier Lemoine, Paulo Gonçalvès et Patrick Flandrin sous les auspices du CNRS (Centre National de la Recherche Scientifique) et de son GdR Information, Signal et Images. Des parties de la boîte à outils ont été écrites au Département d'Ingéniérie Electrique et Informatique de l' Université de Rice (USA) avec le support de la NSF.

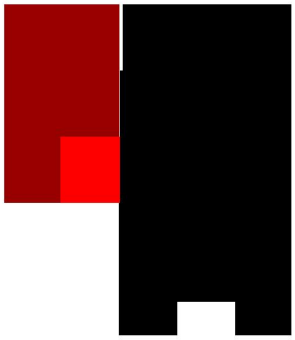
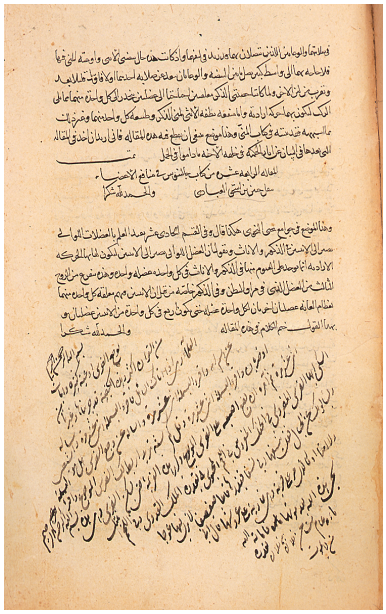
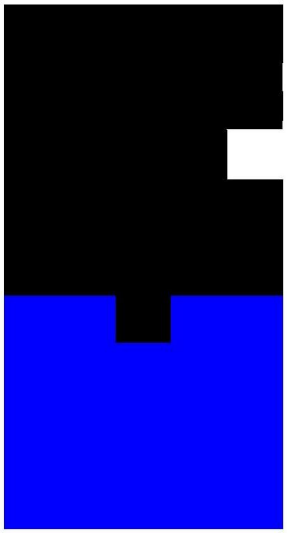
Annexe B

Expérimentations

TABLEAU B.1: Résultats de la détection de zones pour un échantillon de 8 documents appartenant à 4 classes différentes.

Classe	Image originale	Zones d'orientation	Taux de détection%
Documents simples			100%
			100%

Classe	Image originale	Zones d'orientation	Taux de détection%
Documents multi-orientés			100%
			100%

Classe	Image originale	Zones d'orientation	Taux de détection%
Documents multi-tailles			66.67%
			90%

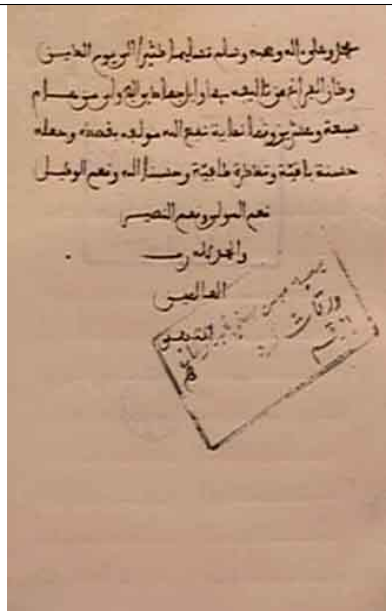
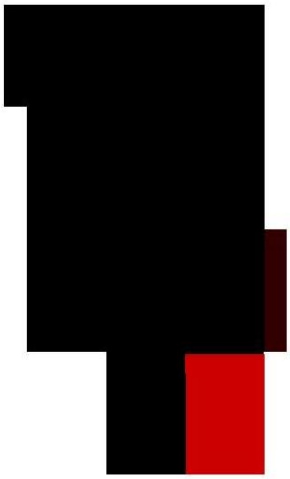
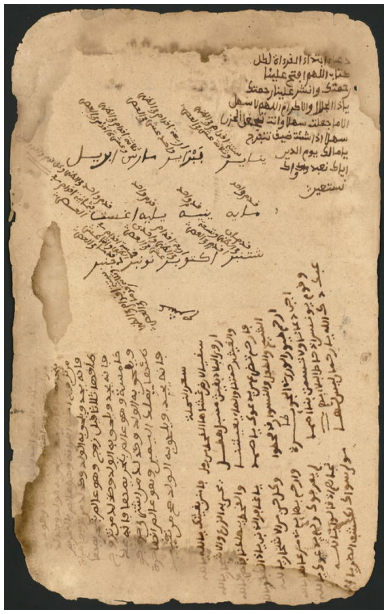
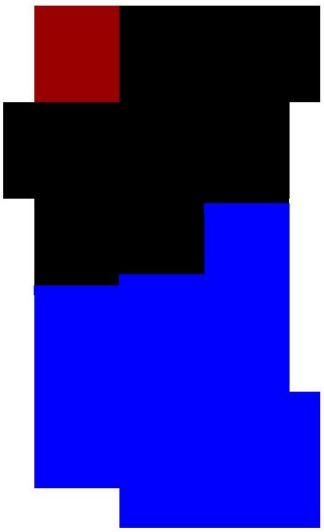
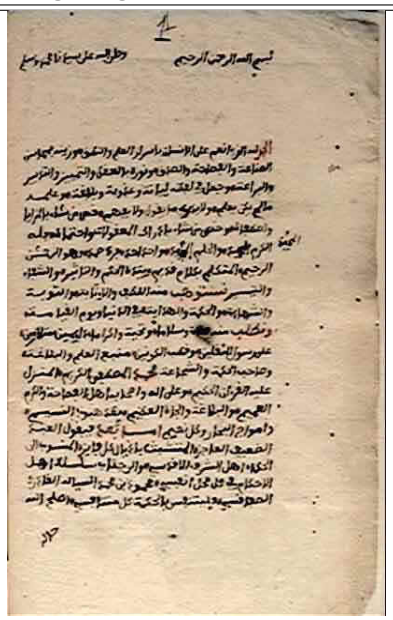
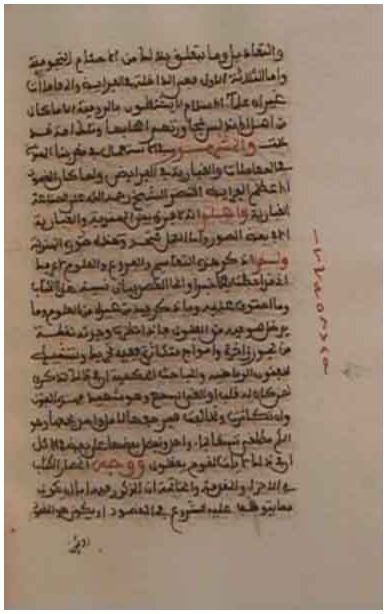
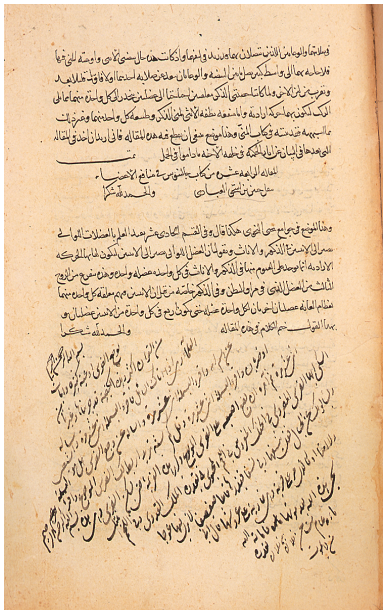
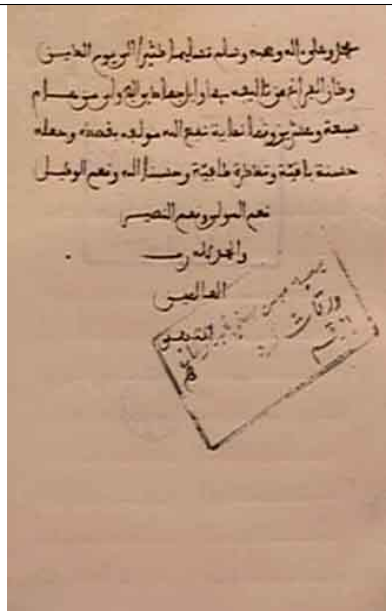
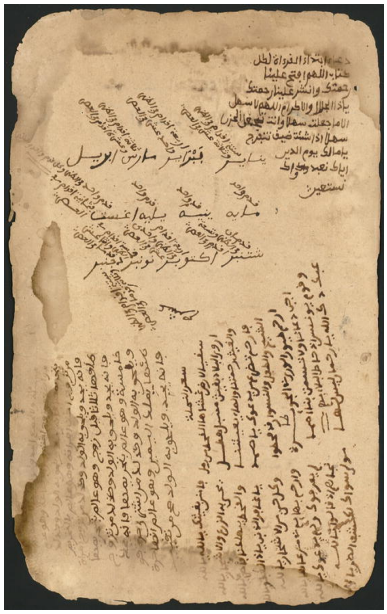
Classe	Image originale	Zones d'orientation	Taux de détection%
Documents avec problèmes au centre			60%
			50%

TABLEAU B.2: Résultats de la détection de lignes pour un échantillon de 8 documents appartenant à 4 classes différentes.

Classe	Image originale	Lignes détectées	Taux de détection%
Documents simples			100%
			100%

Classe	Image originale	Lignes détectées	Taux de détection%
Documents multi-orientés			96.29%
			86.30%

Classe	Image originale	Lignes détectées	Taux de détection%
Documents multi-tailles			85.18%
			50%

Classe	Image originale	Lignes détectées	Taux de détection%
Documents multi-tailles			80.64%
			60%

Annexe C

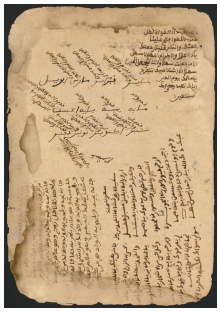
Base de documents

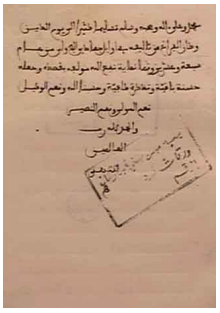
C.1 Documents simples

C.2 Documents multi-orientés

C.3 Documents multi-tailles

C.4 Documents avec problèmes au centre





Annexe D

Publications

D.1 Revues nationales avec comité de lecture

[1] N. Ouwayed, A. Belaïd et F. Auger, "Estimation de l'inclinaison d'un document arabe manuscrit numérisé par analyse temps-fréquence des histogrammes de projection", Journal Traitement du Signal (TS), vol. 26, N°4, 2009.

D.2 Conférences internationales avec actes et comité de lecture

[1] N. Ouwayed, A. Belaïd, "Détection et Séparation de Lignes Connectées des Anciens Documents Manuscrits Multi-orientés", Colloque International Francophone sur l'Ecrit et le Document (CIFED'2010), Sousse, Tunisie, 18-20 mars 2010.

[2] N. Ouwayed, A. Belaïd and F. Auger, "General text line extraction approach based on locally orientation estimation", International Conference on Document Recognition and Retrieval XVII (DRR'2010), California, USA, January 2010.

[3] N. Ouwayed et A. Belaïd, "Une Approache Générale pour l'Extraction des Lignes des Documents Arabes Anciens Multi-orientées", 12ème Colloque International sur le Document Electronique (CIDE.12), Canada, Octobre 2009.

[4] N. Ouwayed and A. Belaïd, "Separation of Overlapping and Touching Lines within Handwritten Arabic Documents", the 13th International Conference on Computer Analysis of Images and Patterns (CAIP'2009), pages 237-244, Germany, September 2009.

[5] N. Ouwayed, A. Belaïd and François Auger, "Cohen's Class Distributions for Skew Angle Estimation in Noisy Ancient Arabic Documents", the 10th International Conference on Document Analysis and Recognition (ICDAR'2009), Third Workshop on Analytics for Noisy Unstructured Text Data (AND'2009), pages 41-46, Spain, July 2009.

[6] N. Ouwayed and A. Belaïd, "Multi-Oriented Text Line Extraction from Handwritten Arabic Documents", The Eighth IAPR Workshop on Document Analysis Systems

(DAS'2008), pages 339-346, Japan, September 17-19, 2008.

D.3 Talks

[1] N. Ouwayed et A. Belaïd, "Extraction des lignes de manuscrit arabes anciens mutli-orientées", Journée GRCE'2009, Paris,Juin 2009.

Bibliographie

[BNT, 2009] (2009). http ://www.bibliotheque.nat.tn/.

[NLM, 2009] (2009). http ://www.nlm.nih.gov/hmd/arabic/welcome.html.

[WIK, 2009] (2009). http ://fr.wikipedia.org/.

[TSI, 2009] (2009). http ://www.tsi.enst.fr/.

[sna, 2009] (2009). http ://vims.cis.udel.edu/ chandra/689/Spring07/snakesUD1.ppt.

[gvf, 2009] (2009). http ://www.tsi.enst.fr/tsi/enseignement/ressources/mti/gvf/gvf2.htm.

[MAT, 2009] (2009). http ://iste.epfl.ch/.

[Akiyama et Hagita, 1990] Akiyama, T. et Hagita, N. (1990). Automatic entry system for printed documents. *Pattern Recognition*, 23:1141–1154.

[Allili et Ziou, 2007] Allili, M. et Ziou, D. (2007). Adaptive appearance model for object contour tracking in videos. *Computer and Robot Vision, Canadian Conference*, 0:510–518.

[Antonacopoulos et Karatzas, 2004] Antonacopoulos, A. et Karatzas, D. (2004). Document image analysis for world war ii personal records. *International Workshop on Document Image Analysis for Libraries*, 0:336–343.

[Aradhya *et al.*, 2006] Aradhya, V. N. M., Kumar, G. H. et Shivakumara, P. (2006). Skew detection technique for binary document images based on hough transform. *International Journal of Information Technology*, 3.

[Auger et Doncarli, 1992] Auger, F. et Doncarli, C. (1992). Quelques commentaires sur des représentations temps-fréquence proposées récemment. *Traitement du Signal*, 9(1):3–25.

[Auger *et al.*, 1996] Auger, F., Lemoine, O., Goncalves, P. et P.Flandrin (1996). *Time-frequency toolbox for Matlab, user's guide and reference guide.* http ://tftb.nongnu.org.

[Babu *et al.*, 2006] Babu, D. R. R., Kumat, P. M. et Dhannawat, M. D. (2006). Skew angle estimation and correction of hand written, textual and large areas of non-textual document images : A novel approach. *In IPCV*, pages 510–515.

[Bascle, 1994] Bascle, B. (1994). Contributions et applications des modèles déformables en vision par ordinateur. *Thèse de doctorat de l'Université de Nice-Sophia Antipolis*, page 200.

[Belaid et Belaid, 1992] Belaid, A. et Belaid, Y. (1992). *Reconnaissance des formes : méthodes et applications*. InterEditions.

[Bennasri *et al.*, 1999] Bennasri, A., Zahour, A. et Taconet, B. (1999). Extraction des lignes d'un texte manuscrit arabe. *Vision Interface'99*, pages 42–48.

[Blake et Yuille, 1992] Blake, A. et Yuille, A. (1992). *Active Vision*. MIT Press.

[Blu et Lebrun, 2005] Blu, T. et Lebrun, J. (2005). Analyse temps-fréquence linéaire ii : représentations de type ondelettes. *In* Hlawatsch, F. et Auger, F., éditeurs : *Temps-fréquence, concept et outils*, pages 101–138. Hermès, Traité IC2.

[Boashash, 1992] Boashash, B. (1992). Estimating and interpreting the instantaneous frequency of a signal – part i : Fundamentals. *Proc. IEEE*, 80(4):519–538.

[Boashash *et al.*, 1987] Boashash, B., Lovell, B. et White, L. (1987). Time frequency analysis and pattern recognition using singular value decomposition of the wigner-ville distribution. *Adv. Algorithms Architect. Signal Process., Proc. SPIE 828*, pages 104–114.

[Boles et Boashash, 1988] Boles, P. et Boashash, B. (1988). The cross wigner-ville distribution-a two dimensional analysis method for the processing of vibrosis seismic signals. *Proc. IEEE ICASP 87*, pages 904–907.

[Born et Jordan, 1925] Born, M. et Jordan, P. (1925). Zur quantenmechanik. *Z. Phys*, 34:858–888.

[Bourgeois, 1997] Bourgeois, F. L. (1997). Robust multifont ocr system from gray level images. *In Proceedings of the Fourth International Conference on Document Analysis and Recognition*, volume 1, pages 1–5.

[Bourgeois *et al.*, 2001] Bourgeois, F. L., Emptoz, H., Trinh, E. et Duong, J. (2001). Networking digital document images. *International Conference on Document Analysis and Recognition*, 0:379–384.

[Brunet *et al.*, 1992] Brunet, R., Ferras, R. et Théry, H. (1992). Les mots de la géographie. *dictionnaire critique, Reclus-Documentation Française*, pages 286–287.

[Bukhari *et al.*, 2009] Bukhari, S. S., Shafait, F. et Breuel, T. M. (2009). Script-independent handwritten textlines segmentation using active contours. *In ICDAR09*, pages 91–108.

[Caselles *et al.*, 1995] Caselles, V., Kimmel, R. et Sapiro, G. (1995). Geodesic active contours. *International Conference on Computer Vision*, pages 694–699.

[Chaudhuri et Chaudhuri, 1997] Chaudhuri, A. et Chaudhuri, S. (1997). Robust detection of skew in document images. *IEEE Trans. on Image Processing*, 6(2):344–349.

[Chen et Haralick, 1994] Chen, S. et Haralick, R. M. (1994). An automatic algorithm for text skew estimation in document images using recursive morphological transforms. *Proc. Int. Conf. on Image Processing, Austin, USA*, 1:139–143.

[Chen et Leedham, 2005] Chen, Y. A. et Leedham, G. (2005). Independent component analysis segmentation algorithm. *In 8th International Conference on Document Analysis and Recognition*, pages 680–684.

[Choi et Williams, 1989] Choi, H. et Williams, W. J. (1989). Improved time-frequency representation of multi-component signals using exponential kernels. *IEEE Trans. Acoust., Speech, Signal Process.*, 37(6):862–871.

[Choi *et al.*, 2001] Choi, W.-P., Lam, K.-M. et Siu, W.-C. (2001). An adaptive active contour model for highly irregular boundaries. *Pattern Recognition*, 34(2):323–331.

[Classen et Mecklenbrauker, 1980] Classen, T. et Mecklenbrauker, W. (1980). The wigner distribution - a tool for time frequency analysis. *Parts I-III, Philips J. Res.*, 35, Part I : n°3, p. 217-250 ; Part II n°4/5, p. 372-389 ; Part III : n°6, p.372-389.

[Cohen, 1966] Cohen, L. (1966). Generalized phase-space distribution functions. *J. Math. Phys.*, 7(5):781–786.

[Cohen, 1989] Cohen, L. (1989). Time-frequency distributions – a review. *Proc. IEEE*, 77(7):941–948.

[Cristobal *et al.*, 1986] Cristobal, G., Bescos, J. et Santamaria, J. (1986). Application of wigner distribution for image representation and analysis. *Proc. IEEE Eighth Int. Conf. Pattern Recogn.*, 23:998–1000.

[Das et Chanda, 2001] Das, A. K. et Chanda, B. (2001). A fast algorithm for skew detection of document images using morphology. *Int. Journal on Document Analysis and Recognition*, 4(2):109–114.

[Dhandra *et al.*, 2006] Dhandra, B. V., Malemath, V. S., H., M. et Hegadi, R. (2006). Skew detection in binary image documents based on image dilation and region labeling approach. *International Conference on Pattern Recognition*, 2:954–957.

[Du *et al.*, 2009] Du, X., Pan, W. et Bui, T. D. (2009). Text line segmentation in handwritten documents using mumford-shah model. *Pattern Recogn.*, 42(12):3136–3145.

[Duda et Hart, 1972] Duda, R. O. et Hart, P. E. (1972). Use of the hough transformation to detect lines and curves in pictures. *Commun. ACM*, 15(1):11–15.

[Escudié et Gréa, 1976] Escudié, B. et Gréa, J. (1976). Sur une formulation générale de la représentation en temps et en fréquence dans l'analyse des signaux d'énergie finie. *CR. Acad. Sci. Paris*, 283:1049–1051.

[F. Yin, 2008] F. Yin, C.-L. L. (2008). Handwritten text line segmentation by clustering with distance metric learning. *In Proc. 11th ICFHR*, pages 229–234.

[Feldbach et Tönnies, 2001] Feldbach, M. et Tönnies, K. D. (2001). Line detection and segmentation in historical church registers. *In ICDAR '01 : Proceedings of the Sixth International Conference on Document Analysis and Recognition*, pages 743–748, Washington, DC, USA. IEEE Computer Society.

[Flandrin, 1993] Flandrin, P. (1993). *Temps-fréquence*. Hermes, Paris.

[Flandrin, 1999] Flandrin, P. (1999). *Time-Frequency/Time-Scale Analysis*. Academic Press, San Diego, CA.

[Flandrin *et al.*, 1992] Flandrin, P., Martin, N., Auger, F., Demars, C., Doligez, T., Lambert-Nebout, Mars, J., Molinaro, F., Ovarlez, J. et Rioul, O. (1992). Méthodes temps-fréquence : Fiches synthétiques. *numéro spécial de la revue Traitement du Signal*, 9(1):77–113.

[Flandrin et Martin, 1984] Flandrin, P. et Martin, W. (1984). A general class of estimators for the wigner-ville spectrum of nonstationary processes. *In* Bensoussan, A. et Lions, J.-L., éditeurs : *Systems Analysis and Optimization of Systems, Lecture Notes in Control and Information Sciences*, volume 62, pages 15–23. Springer, Berlin.

[Gribonval, 2005] Gribonval, R. (2005). Analyse temps-fréquence linéaire i : représentations de type fourier. *In* Hlawatsch, F. et Auger, F., éditeurs : *Temps-fréquence, concept et outils*, pages 69–100. Hermès, Traité IC2.

[Hippenstiel et Oliviera, 1990] Hippenstiel, R. et Oliviera, P. M. D. (1990). Time varying spectral estimation using the instantaneous power spectrum (ips). *IEEE Trans. Acoust., Speech, Signal Process.*, 38(10):1752–1759.

[Hlawatsch et Auger, 2005] Hlawatsch, F. et Auger, F. (2005). *Temps-fréquence : concepts et outils*. Hermes, Lavoisier, Paris.

[Hlawatsch et Boudreaux-Bartels, 1992] Hlawatsch, F. et Boudreaux-Bartels, G. F. (1992). Linear and quadratic time-frequency signal representation. *IEEE Signal Process. Mag*, 9(2):21–67.

[Hyvarinen, 1999] Hyvarinen, A. (1999). Survey on independent component analysis. *In Helsinki University of Technology*.

[Inglis, 1999] Inglis, S. (1999). Lossless document image compression. *Thèse de doctorat, Université de Waikato, New Zealand*.

[James et Ronald, 1980] James, F. K. et Ronald, W. S. (1980). On the use of the iosinh window for spectrum analysis. *IEEE Transactions on Acoustics, Speech and Signal Processing*, ASSP-28(1).

[k. Takru et Leedham, 2002] k. Takru et Leedham, G. (2002). Separation of touching and overlapping words in adjacent lines of handwritten text. *In International Workshop on Frontiers in Handwriting Recognition*, pages 496–501.

[Kass *et al.*, 1987] Kass, M., Witkin, A. et Terzopoulos, D. (1987). Snakes : Active contour models. *Proc. 1st ICCV*, pages 259–268.

[Kass *et al.*, 1988] Kass, M., Witkin, A. et Terzopoulos, D. (1988). Snakes : Active contour models. *International Journal of Computer Vision*, 55:321–331.

[Kavallieratou *et al.*, 1999] Kavallieratou, E., Fakotakis, N. et Kokkinakis, G. (1999). Skew angle estimation in document processing using cohen's class distributions. *Pattern Recogn. Lett.*, 20:11–13.

[Kavallieratou *et al.*, 2002] Kavallieratou, E., Fakotakis, N. et Kokkinakis, G. (2002). Skew angle estimation for printed and handwritten documents using the wigner-ville distribution. *Image and Vision Computing*, 20:813–824.

[Lam *et al.*, 1992] Lam, L., Lee, S.-W. et Suen, C. Y. (1992). Thinning methodologies-a comprehensive survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(9):869–885.

[Leitner et Cinquin, 1993] Leitner, F. et Cinquin, P. (1993). From splines and snakes to snake splines. *In Selected Papers from the Workshop on Geometric Reasoning for Perception and Action*, pages 264–281, London, UK. Springer-Verlag.

[León *et al.*, 2004] León, J. L. D., Yánez, C. et Guzmán, G. (2004). *Progress in Pattern Recognition, Image Analysis and Applications*. Springer.

[Li *et al.*, 2008] Li, Y., Zheng, Y., Doermann, D. et Jaeger, S. (2008). Script-independent text line segmentation in freestyle handwritten documents. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(8):1313–1329.

[Likforman-Sulem et Faure, 1994] Likforman-Sulem, L. et Faure, C. (1994). Une méthode de résolution des conflits d'alignements pour la segmentation des documents manuscrits. *Traitement de Signal (TS)*, 12(6):541–549.

[Likforman-Sulem *et al.*, 1995] Likforman-Sulem, L., Hanimyan, A. et Faure, C. (1995). A hough based algorithm for extracting text lines in handwritten document. *In Proc. of ICDAR'95*, pages 774–777.

[Louloudis *et al.*, 2009a] Louloudis, G., Gatos, B., Pratikakis, I. et Halatsis, C. (2009a). Text line and word segmentation of handwritten documents. *Pattern Recognition*, 42(12):3169 – 3183. New Frontiers in Handwriting Recognition.

[Louloudis *et al.*, 2009b] Louloudis, G., Gatos, B., Pratikakis, I. et Halatsis, C. (2009b). Text line and word segmentation of handwritten documents. *Pattern Recognition*, 42(12):3169–3183.

[Lu *et al.*, 2004] Lu, Y., Wang, Z. et Tan, C. L. (2004). Word grouping in document images based on voronoi tessellation. *In Document Analysis Systems*, pages 147–157.

[Mahadevan et Nagabushnam, 1995] Mahadevan, U. et Nagabushnam, R. C. (1995). Gap metrics for word separation in handwritten lines. *In ICDAR '95 : Proceedings of the Third International Conference on Document Analysis and Recognition (Volume 1)*, pages 124–127.

[Malleron *et al.*, 2009] Malleron, V., Eglin, V., Emptoz, H., Dord-Crouslé, S. et Régnier, P. (2009). Text lines and snippets extraction for 19th century handwriting documents layout analysis. *International Conference on Document Analysis and Recognition*, 0:1001–1005.

[Margenau et Hill, 1961] Margenau, H. et Hill, R. W. (1961). Correlation between measurements in quantum theory. *Proc. Theor. Phys.*, 26:772–738.

[Maître, 2000] Maître, H. (2000). *Le traitement des images*. ENST.

[Mcinerney et Terzopoulos, 1996] Mcinerney, T. et Terzopoulos, D. (1996). Deformable models in medical image analysis : A survey. *Medical Image Analysis*, 1:91–108.

[Montagnat *et al.*, 2001] Montagnat, J., Delingette, H. et Ayache, N. (2001). A review of deformable surfaces : topology, geometry and deformation. *Image and Vision Computing*, 19(14):1023–1040.

[Mumford et Shah, 1989] Mumford, D. et Shah, J. (1989). Optimal approximation by piecewise smooth functional and associated variational problems. *Commun. Pure Appl. Math.*, 42:577–685.

[Nagy *et al.*, 1988] Nagy, G., Kanai, J., Krishnamoorthy, M., Thomas, M. et Viswanathan, M. (1988). Two complementary techniques for digitized document analysis. *In DOCPROCS '88 : Proceedings of the ACM conference on Document processing systems*, pages 169–176, New York, NY, USA. ACM.

[Nicolaou et Gatos, 2009] Nicolaou, A. et Gatos, B. (2009). Handwritten text line segmentation by shredding text into its lines. *International Conference on Document Analysis and Recognition*, 0:626–630.

[Nicolas *et al.*, 2004] Nicolas, S., Paquet, T. et Heutte, L. (2004). Text line segmentation in handwritten document using a production system. *Frontiers in Handwriting Recognition, International Workshop on*, pages 245–250.

[Osher et Paragios, 2003] Osher, S. et Paragios, N. (2003). *Geometric Level Set Methods in Imaging,Vision,and Graphics*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.

[Ovarlez *et al.*, 2005] Ovarlez, J.-P., Gonçalvès, P. et Baraniuk, R. (2005). Analyse temps-fréquence quadratique iii : la classe affine et autres classes covariantes. *In* Hlawatsch, F. et Auger, F., éditeurs : *Temps-fréquence, concept et outils*, pages 201–236. Hermès, Traité IC2.

[Oztop *et al.*, 1999] Oztop, E., Mulayim, A. Y., Atalay, V. et Vural, F. Y. (1999). Repulsive attractive network for baseline extraction on document images. *Signal Processing*, 75:1–10.

[Pal et Chaudhuri, 1996] Pal, U. et Chaudhuri, B. B. (1996). An improved document skew angle estimation technique. *Pattern Recognition Letters*, 17:899–904.

[Pal et Chaudhuri, 1997] Pal, U. et Chaudhuri, B. B. (1997). Skew angle detection of digitized indian script documents. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19(2):182–186.

[Pal et Datta, 2003] Pal, U. et Datta, S. (2003). Segmentation of bangla unconstrained handwritten text. *International Conference on Document Analysis and Recognition*, 2:11–28.

[Pavlidis et Zhou, 1992] Pavlidis, T. et Zhou, J. (1992). Page segmentation and classification. *Computer Vision Graphics and Image Processing*, 54(2):484–496.

[Peake et Tan, 1997] Peake, G. et Tan, T. (1997). A general algorithm for document skew angle estimation. *IEEE Int. Conf. Image Process.*, 2:230–233.

[Pham *et al.*, 2000] Pham, D. L., Xu, C. et Prince, J. L. (2000). A survey of current methods in medical image segmentation. *In Annual Review of Biomedical Engineering*, volume 2, pages 315–338.

[Pluempitiwiriyawej *et al.*, 2005] Pluempitiwiriyawej, C., Moura, J., Wu, Y. J. L. et Ho, C. (2005). Stacs : new active contour scheme for cardiac mr image segmentation. *Medical Imaging, IEEE Transactions on*, 24(5):593–603.

[Postl, 1986] Postl, W. (1986). Detection of linear oblique structures and skew scan in digitized documents. *Proceedings of the Eighth International Conference on Pattern Recognition, IEEE CS Press, Los Alamitos, CA*, pages 687–689.

[Pu et Shi, 1998] Pu, Y. et Shi, Z. (1998). A natural learning algorithm based on hough transform for text lines extraction in handwritten document. *In Proc. 6th International Workshop on Frontiers in Handwriting Recognition*, pages 637–646.

[Ramlau et Ring, 2007] Ramlau, R. et Ring, W. (2007). A mumford-shah level-set approach for the inversion and segmentation of x-ray tomography data. *J. Comput. Phys.*, 221(2):539–557.

[Rihaczek, 1992] Rihaczek, A. W. (1992). *Principles of High-Resolution Radar*. Artech House, Norwood, MA.

[Rousselle, 2003] Rousselle, J. (2003). Les contours actifs, une méthode de segmentation. application à l'imagerie médicale. *Thèse de doctorat de l'Université de François Rabelais de Tours*.

[Sarfraz *et al.*, 2007] Sarfraz, M., Mahmoud, S. A. et Rasheed, Z. (2007). On skew estimation and correction of text. *Computer Graphics, Imaging and Visualization, International Conference on*, 0:308–313.

[Sauvola *et al.*, 1997] Sauvola, J., Doermann, D. et Pietikainen, M. (1997). Locally adaptive document skew detection. *Proc. SPIE Document Recognition and Retrieval IV*, 3027:96–108.

[Sethian, 1985] Sethian, J. A. (1985). Curvature and the evolution of fronts. *Communications in Mathematical Physics*, 101(4):487–499.

[Shapiro *et al.*, 1993] Shapiro, V., Gluchev, G. et Sgurev, V. (1993). Handwritten document image segmentation and analysis. *Pattern Recogn. Lett.*, 14(1):71–78.

[Sharif et Movahhedinia, 2008] Sharif, A. E. et Movahhedinia, N. (2008). On skew estimation of persian/arabic printed documents. *Journal of Applied Sciences*, 8(12): 175–180.

[Shi et Govindaraju, 2004] Shi, Z. et Govindaraju, V. (2004). Line separation for complex document images using fuzzy run length. *In Int. Workshop on Document Image Analysis for Libraries*.

[Srihari et Govindraju, 1989] Srihari, S. N. et Govindraju, V. (1989). Analysis of textual image using the hough transform. *Machine Vision Applications*, 2:141–153.

[Stankovic, 1994] Stankovic, L. (1994). A method for time-frequency analysis. *IEEE Trans. Signal Process.*, 42(1):255–229.

[Su *et al.*, 2007] Su, T., Zhang, T. et Guan, D. (2007). Corpus-based hit-mw database for offline recognition of general-purpose chinese handwritten text. *Int. J. Doc. Anal. Recognit.*, 10(1):27–38.

[Sumengen, 2004] Sumengen, B. (2004). Variational image segmentation and curve evolution on natural imagest. *PhD dissertation*.

[Wigner, 1932] Wigner, E. P. (1932). On the quantum correction for thermodynamic equilibrium. *Phys. Rev.*, 40:749–759.

[Xu et Prince, 1997] Xu, C. et Prince, J. L. (1997). Gradient vector flow : A new external force for snakes. *Proc. IEEE Conf. on Comp. Vis. Patt. Recog. (CVPR)*, pages 66–71.

[Yin, 2001] Yin, P. Y. (2001). Skew detection and block classification of printed documents. *Image Vis. Comput. 19 (8)*, pages 567–579.

[Zahn et Roskies, 1972] Zahn, C. T. et Roskies, R. Z. (1972). Fourier descriptors for plane close curves. *IEEE Trans. Computers*, C-21(14):269–281.

[Zahour *et al.*, 2007] Zahour, A., Likforman-Sulem, L., Boussellaa, W. et Taconet, B. (2007). Text line segmentation of historical arabic documents. *In 9th Int. Conf. on Document Analysis and Recognition*, pages 138–142.

[Zahour *et al.*, 2008] Zahour, A., Taconet, B., Likforman-Sulem, L. et Boussellaa, W. (2008). Overlapping and multi-touching text-line segmentation by block covering analysis. *Pattern Analysis and Applications*, 12(4):335–351.

[Zahour *et al.*, 2004] Zahour, A., Taconet, B. et Ramdane, S. (2004). Contribution à la segmentation de textes manuscrits anciens. *In Conférence Internationale Francophone sur l'Ecrit et le Document (CIFED'04)*.

[Zhang et Hu, 2008] Zhang, J. et Hu, J. (2008). Image segmentation based on 2d otsu method with histogram analysis. *Computer Science and Software Engineering, International Conference on*, 6:105–108.

[Zhao *et al.*, 1990] Zhao, Y., Atlas, L. E. et Marks, R. J. (1990). The use of cone-shaped kernels for generalized time-frequency representations of nonstationary signals. *IEEE Trans. Acoust., Speech, Signal Process.*, 38(7):1084–1091.

[Zheng *et al.*, 2007] Zheng, X., Wang, Y. et Zhao, X. (2007). Fingerprint image segmentation using active contour model. *Image and Graphics, International Conference on*, 0:437–441.

Index des auteurs

Akiyama, T. 47, 49
Allili, M. 38
Antonacopoulos, A. 11, 29
Aradhya, V. N. Manjunath 48
Atalay, V. xii, 20
Atlas, L. E. 55
Auger, F. 51, 61
Ayache, N. 38

Babu, D. R. Ramesh 48
Baraniuk, R. 60
Bascle, B. 38
Belaid, A. 48
Belaid, Y. 48
Bennasri, A. xi, 9
Bescos, J. 51
Blake, A. 38
Blu, T. 60
Boashash, B. 51, 52
Boles, P. 51
Born, M. 55
Boudreaux-Bartels, G. F. 51
Bourgeois, F. Le 20, 29
Boussellaa, W. xi, 13
Breuel, T. M. xii, 23
Brunet, R. 35
Bui, T. D. xii, 26
Bukhari, S. S. xii, 23

Caselles, V. 22, 38
Chanda, B. 48
Chaudhuri, A. 48
Chaudhuri, B. B. 47, 48
Chaudhuri, S. 48
Chen, S. 48
Chen, Y. A. 91
Choi, H. 55
Choi, W-P. 38
Cinquin, P. 23
Classen, T.A.C.M. 51, 54
Cohen, L. 51
Cristobal, G. 51

Das, A. K. 48
Datta, Sagarika 36
Delingette, H. 38
Demars, C. 51
Dhandra, B. V. 48
Dhannawat, Mahesh D. 48
Doermann, D. xii,24–26,29,48,89
Doligez, T. 51
Doncarli, C. 51
Dord-Crouslé, S. xii, 17
Du, X. xii, 26
Duda, R. O. 15
Duong, J. 20, 29

Eglin, V. xii, 17
Emptoz, H. xii, 17
Escudié, B. 51

F. Yin, C.-L. Liu xii, 20
Fakotakis, N. 48, 60
Faure, C. xi,xii,6,13–15,17,18,29,89
Feldbach, M. xi, 14
Ferras, R. 35
Flandrin, P. 51, 54

Gatos, B. xi, xii
Gluchev, G. 11, 29
Goncalves, P. 61, 107
Gonçalvès, P. 60
Govindaraju, V. 19, 29
Govindraju, V. 48
Gréa, J. 51

Gribonval, R. 60
Guan, D. 20, 89
Guzmán, G. 96

H., Mallikarjun 48
Hagita, N. 47, 49
Halatsis, C. xii, 19
Hanimyan, A. xii, 15
Haralick, R. M. 48
Hart, P. E. 15
Hegadi, Ravindra 48
Heutte, L. xii, 20
Hill, R. W. 54
Hippenstiel, R.D 55
Hlawatsch, F. 51
Ho, C. 22
Hu, J. 47, 49
Hyvarinen, A. 91

Inglis, S.J. 47, 48

Jaeger, S. xii,24–26,29,89
James, F. Kaiser 68
Jordan, P. 55

k. Takru 92
Kanai, Junichi 12, 29
Karatzas, D. 11, 29
Kass, M. 22, 23
Kavallieratou, E. 48, 60
Kimmel, R. 22, 38
Kokkinakis, G. 48, 60
Krishnamoorthy, Mukkai 12, 29
Kumar, G. Hemantha 48
Kumat, Piyush M. 48

Lam, K-M. 38
Lam, L. 96
Lambert-Nebout 51
Lebrun, J. 60
Lee, S.-W. 96
Leedham, G. 91, 92
Leitner, F. 23
Lemoine, O. 61, 107
León, J. L. D. 96
Li, Y. xii,24–26,29,89
Likforman-Sulem, L. xi,xii,6,13–15,17,18,29,36,89
Louloudis, G. xii, 19
Lovell, B. 51
Lu, Yue 12, 29

Mahadevan, U. 37
Mahmoud, S. A. 47
Malemath, V. S. 48
Malleron, V. xii, 17
Margenau, H. 54
Marks, R. J. 55
Mars, J. 51
Martin, N. 51
Martin, W. 54
Maître, H. 41
Mcinerney, T. 22
Mecklenbrauker, W.F.G. 51, 54
Molinaro, F. 51
Montagnat, J. 38
Moura, J.M.F. 22
Movahhedinia, N. 47, 49
Mulayim, A. Y. xii, 20
Mumford, D. 26

Nagabushnam, R. C. 37
Nagy, George 12, 29
Nicolaou, A. xi, 10
Nicolas, S. xii, 20

Oliviera, P. M. De 55
Osher, S. 22
Ovarlez, J-P. 60
Ovarlez, J.P. 51
Oztop, E. xii, 20

Pal, U. 36, 47
Pan, W. xii, 26
Paquet, T. xii, 20
Paragios, N. 22
Pavlidis, T. 47, 49
Peake, G.S. 48, 68
P.Flandrin 61, 107
Pham, D. L. 38
Pietikainen, M. 48

Pluempitiwiriyawej, C. 22
Postl, W. 48
Pratikakis, I. xii, 19
Prince, J. L. xii, 38
Pu, Y. 17, 29

Ramdane, S. xi, 13
Ramlau, R. 23
Rasheed, Z. 47
Régnier, P. xii, 17
Rihaczek, A. W. 54
Ring, W. 23
Rioul, O. 51
Ronald, W. Schafer 68
Roskies, R. Z. 41
Rousselle, J. 38

Santamaria, J. 51
Sapiro, G. 22, 38
Sarfraz, M. 47
Sauvola, J. 48
Sethian, J. A. 22
Sgurev, V. 11, 29
Shafait, F. xii, 23
Shah, J. 26
Shapiro, V. 11, 29
Sharif, A. E. 47, 49
Shi, Z. 17, 19
Shivakumara, P. 48
Siu, W-C. 38
Srihari, S. N. 48
Stankovic, L. 55
Su, T. 20, 89
Suen, C. Y. 96
Sumengen, B. 23

Taconet, B. xi, 9
Tan, Chew Lim 12, 29
Tan, T.N. 48, 68
Terzopoulos, D. 22, 23
Thomas, Mathews 12, 29
Théry, H. 35
Tönnies, K. D. xi, 14
Trinh, E. 20, 29

Viswanathan, Mahesh 12, 29
Vural, F. Yarman xii, 20

Wang, Y. 38
Wang, Zhe 12, 29
White, L. 51
Wigner, E. P. 53
Williams, W. J. 55
Witkin, A. 22, 23
Wu, Y. J. L. 22

Xu, C. xii, 38

Yin, P. Y. 48, 68
Yánez, C. 96
Yuille, A. 38

Zahn, C. T. 41
Zahour, A. xi, 9
Zhang, J. 47, 49
Zhang, T. 20, 89
Zhao, X. 38
Zhao, Y. 55
Zheng, X. 38
Zheng, Y. xii,24–26,29,89
Zhou, J. 47, 49
Ziou, Djemel 38

Index

DEMOT, 34
E_{con} (énergie de contexte), 23, 40
E_{image} (énergie de l'image ou énergie extérieur), 23, 40
E_{int} (énergie interne), 23, 40
SIG, 35
élasticité, 39
érosion, 48

active contour model, i, 139
analyse temps-fréquence, 48
anciens manuscrits Grecque, 19
annotations, 6
approches descendantes, 9
Archives et Bibliothèque Nationales d'Egypte, 6

Bibliothèque Nationale de médecine des U.S.A, 5
Bibliothèque Nationale de Tunis, 5
binarization, 106
blocs, 13
Born-Jordan, 55

Choi-Williams, 55
classe de Cohen, 51
classification par k-ppv, 13
composantes connexes, 17
connexion, i, 139
Contour actif géométrique, 39
Contour actif paramétrique, 39
contours, 17
Correction du maillage, 76
courbe, 96

Détection de l'orientation erronée, 74
descendant, 94
dilatation, 48
distribution de Wigner-Ville, i, 48, 139
documents arabes manuscrits, i, 139
Documents hétérogènes, 5
documents manuscrits arabe multi-orientés, 5
documents manuscrits latins, 20
Documents mono-orientés, 5
Documents multi-orientés, 5
Documents multi-scripts, 5

espace de Hough, 15
Estimation de l'orientation, 47
Extension de l'orientation, 75
extraction de lignes, i, 139

filtre gaussien, 24
fission-fusion, 77

GVF : Gradient Vector Flow, 45
GVF : Gradient Vector Flow, 40

handwritten Arabic documents, i, 139

intelligence artificielle, 22

mécanisme de minimisation d'énergie, 38
maillage automatique, 41
maillage rectangulaire, 36
mailles, 35
Margenau-Hill, 54
Matlab, 105
maxima, 9
minima, 9, 11
modèle de contour actif, i, 38, 139
morphologie Arabe, 93

noyau anisotropique, 24

orientation, i, 103, 139
OuwayedImaPro, 106

PAWs, 49
point d'ancrage, 14
Pseudo Margenau-Hill, 55
Pseudo Wigner-Ville, 54
Pseudo Wigner-Ville lissée, 54

résolution, 6
registres imprimés de la seconde guerre mondiale, 11
rigidité, 39
RLS, 19

Sélection des fenêtres, 41
Séparation des lignes connectées, 91
snake, i, 23, 139
Spectrogramme , 55
Suivi de lignes, 83
système de production, 22

technique de lissage, 19
text line extraction, i, 139
TFTB, 107
topologie, 39
touching, i, 139
Transformée de Fourier, 67
transformée de Hough, 11, 15

variance angulaire, 96

Wigner-Ville distribution, i, 139

Zhao-Atlas-Marks, 55

Résumé

L'indexation de documents numérisés manuscrits pose le problème de la segmentation en lignes qui, si elle échoue, handicape les étapes suivantes d'extraction et de reconnaissance de mots. Dans les documents arabe anciens, s'ajoute à ce problème, la présence dans les marges, d'annotations souvent composées de lignes obliques. La détection de ces lignes est nécessaire et constitue un défi important pour l'indexation de ces documents. Ainsi, la segmentation visée dans ce travail de thèse concerne l'extraction de lignes multi-orientées. Pour ce problème, la bibliographie ne présente que des techniques rudimentaires basées essentiellement sur une projection directe de l'image du document suivant une seule direction et donc non applicable à du texte multi-orienté. Devant ce manque, nous avons proposé une approche adaptative permettant de localiser d'abord les zones d'orientation différentes, puis de s'appuyer sur chaque orientation locale pour extraire les lignes. Pendant ma thèse, j'ai développé les points suivants :

- Application d'un maillage automatique en utilisant le modèle de contour actif (snake).
- Préparation du signal de profil de projection en supprimant tous les pixels qui ne sont pas nécessaires dans le calcul de l'orientation. Ensuite, application de toutes les distributions d'énergie de la classe de Cohen sur le profil de projection pour trouver la meilleure distribution qui donne l'orientation.
- Application de quelques règles d'extension pour trouver les zones.
- Extraction des lignes en se basant sur un algorithme de suivi des composantes connexes.
- Séparation de lignes se chevauchant et se connectant en utilisant la morphologie des lettres terminales arabes.

Mots-clés: documents arabes manuscrits, extraction de lignes, estimation de l'orientation, modèle de contour actif (snake), distribution de Wigner-Ville, connexion de lignes.

Abstract

The indexing of handwritten scanned documents poses the problem of lines segmentation, if it fails, disabling the following steps of words extraction and recognition. In addition, the ancient Arabic documents contain annotations in the margins, often composed of lines obliquely oriented. The detection of these lines is important as the rest and is a major challenge for the indexing of these documents. Thus, the segmentation described in this thesis involves the extraction of multi-oriented lines. For this problem, the bibliography has only rudimentary techniques based essentially on the projection of the document image along one direction, which be failed in the case of multi-oriented documents. Given this lack, we have proposed an adaptive approach that first locates the different orientation zones, then based on each local orientation to extract the lines. During my thesis, i particularly invested on the following points :

- Applying an automatic paving using the active contour model (snake).
- Preparation the signal of the projection profile by removing all pixels that are not needed in the orientation estimation. Then, implementation of all energy distributions of Cohen's class on the projection profile to find the best distribution that gives the orientation.
- Applying some extension rules to find the oriented zones.
- Extraction of lines by using an connected components follow-up algorithm.
- Separation of overlapped and touched lines using the morphology of Arabic terminal letters.

Keywords: handwritten Arabic documents, text line extraction, orientation estimation, active contour model (snake), Wigner-Ville distribution, touching lines.