Université de La Rochelle — École doctorale S2i

Thèse

présentée pour obtenir le titre de DOCTEUR en
Informatique et Applications

Contribution à l'analyse complexe de documents anciens Application aux lettrines

Mickaël Coustaty
mickael.coustaty@univ-lr.fr

Soutenue publiquement le 20/10/2011 devant un jury composé de :

Président	Nicole Vincent	Université Paris Descartes
Rapporteurs	Florence Sèdes	Université de Toulouse
	Josep Llados	Universitat Autonoma de Barcelona
Examinateurs	Marie-Luce Demonet	CESR
	Karell Bertet	Université de La Rochelle
	Wenyin Liu	City University of Hong-Kong
	Salvatore-Antoine Tabbone	Université de Nancy 2
Directeur de thèse	Jean-Marc Ogier	Université de La Rochelle

Laboratoire L3i

Thèse réalisée au Laboratoire Informatique, Image, Interaction
Pôle Sciences & Technologies, Université de La Rochelle
Avenue M. Crépeau
17042 La Rochelle cedex 01

Tél : +33 5 46 45 82 62
Fax : +33 5 46 45 82 42

Web : http ://l3i.univ-larochelle.fr

Sous la direction de Jean-Marc Ogier jean-marc.ogier@univ-lr.fr

Financement Cette thèse a été financée par l'Agence Nationale de la Recherche



Remerciements

Je tiens tout d'abord à remercier Florence Sèdes et Josep Llados d'avoir accepté d'être les rapporteurs de cette thèse. Leurs commentaires et les échanges menés pendant le processus de relecture m'ont permis de faire évoluer ce manuscrit et de préparer au mieux ma soutenance pour aboutir à cette version finale.

Je tiens également à remercier Nicole Vincent, Marie-luce Demonet, Karell Bertet, Wenyin Liu, et Antoine Tabbone d'avoir accepté d'être présents dans mon jury pour examiner mes travaux. Je voudrais tout particulièrement remercier Nicole et Antoine qui m'ont accueilli dans leurs laboratoires pour des séjours et des échanges très enrichissants pour moi. Je voudrais également remercier chaleureusement Karell pour m'avoir donné envie de continuer sur une thèse au laboratoire après mon stage de Master, pour ses conseils et ses discussions, ses explications, et sa gentillesse.

Une thèse est une étape très importante dans la vie d'un jeune chercheur, et je voudrais donc adresser un très grand merci à Jean-Marc sans qui rien de tout cela n'aurait pu être possible. Je voudrais le remercier pour son encadrement scientifique, aussi bien pendant mon stage de master que pendant ma thèse, mais également pour son ouverture d'esprit, sa gentillesse, son talent, et son dynamisme. Il est de ces personnes qui, quand vous ne savez plus vraiment où aller, vous guide et vous redonne envie pour toujours avancer et vous dépasser (et tout cela avec le sourire et de l'humour).

Une thèse, c'est également un bout de vie que l'on partage avec des personnes qui vous marquent. Tout d'abord, je voudrais remercier Vincent Courboulay pour ses discussions (parfois philosophiques :-)), ses échanges, sa gentillesse et son humour. J'ai adoré faire de "la science" avec toi, et j'espère pouvoir continuer. Je voudrais également remercier Pascal Estraillier et Michel Menard pour leurs explications, leur gentillesse et leurs conseils, et l'équipe Ontologie et Lettrines (Karell, Alain, Georges, Thang, et Norbert) qui m'a initié aux ontologies (et ce n'était pas une mince affaire). Les échanges entre "nos deux mondes" resteront des moments très enrichissants pour moi. Je voudrais également remercier tous les stagiaires (Giap, Nishant, Huyen, et Thang) et l'ingénieur (Damien) avec qui j'ai eu l'occasion de travailler dans la bonne humeur pendant ces quatre années. Enfin, je tiens à remercier Stéphanie et Nicholas, deux anciens thésards

du laboratoire qui m'ont fait découvrir la vie de thésard alors que j'étais étudiant. C'est de leur rencontre qu'est née mon envie de faire une thèse et je les en remercie.

Je voudrais également remercier la direction du L3i pour m'avoir accueilli pendant tout ce temps, et les administratifs (Nathalie, Kathy, Dom, Christelle) qui m'ont rendu la vie plus facile dans les différentes démarches administratives. J'associe également les personnes avec qui j'ai pu découvrir le métier d'enseignant (Vincent, Anthony, Frédéric, Bertrand, Jean-Luc), qui m'ont aidée à éclaircir mes idées et à apprendre à rédiger de manière plus claire.

J'adresse une mention spéciale à la 121 bis team (Matthieu, Nicolas et Nathalie) pour ces trois années de bonheur passées dans le même bureau. Ok je parle trop... mais j'ai adoré échanger avec vous sur différents sujets. Et ces trois années auraient été nettement plus monotones sans cette équipe !

Enfin, je termine ces remerciements par ma famille, et tout particulièrement mes parents qui m'ont toujours soutenu pendant toutes ces années d'études et ces 26 ans d'existences. Merci.

à mes parents

“Ranger une bibliothèque, c’est exercer de façon silencieuse l’art de la critique”Jorge Luis Borges, Borges Verbal. Argentine, XXe siècle

Table des matières

Remerciements i

Table des matières v

Table des figures ix

Liste des tableaux xv

Introduction 1

I Indexation d'images par le contenu / État de l'art 7

1 Description et recherche des images par leur contenu 9
1.1 Description des images et Association de mots-clés 9
1.1.1 Recherche par mots-clés 10
1.1.2 Annotation semi-automatique des images 11
1.2 Images étudiées dans le cadre de cette thèse 12
1.3 Présentation générale de la recherche d'images par le contenu 13
1.4 Attributs pour décrire un contenu 15
1.4.1 Approches par description globale 16
1.4.2 Approches par description spatiale 35
1.4.3 Approches par description locale 37
1.5 Les projets déjà existants 38

2 État de l'art de l'analyse des documents anciens 47
2.1 Analyse des documents anciens par les historiens 47
2.1.1 Annotation manuelle et thesaurus 47
2.1.2 Problèmes de l'annotation manuelle d'images 50

2.2 Outils informatiques d'analyse des documents anciens - État de l'art . . . 51
2.2.1 Pourquoi les méthodes usuelles ne fonctionnent pas sur les documents anciens ? . . . 52
2.2.2 Caractérisation de la structure et de l'écriture dans les documents anciens . . . 55
2.2.3 Analyse et reconnaissance d'images graphiques de documents anciens 58
2.3 Discussions sur l'état de l'art . . . 63
2.3.1 Démarche proposée : une approche complexe d'analyse d'images de documents anciens . . . 64

II Approche proposée 67

3 Représentation des connaissances d'experts d'un domaine 69
3.1 Problématique liée à la représentation des connaissances d'experts d'un domaine . . . 69
3.2 Approche proposée : Vue d'ensemble . . . 72
3.2.1 Modélisation Connaissance historiens . . . 73
3.2.2 Cas des connaissances des informaticiens . . . 74
3.2.3 Combinaison de nos modèles . . . 76

4 Modèle d'analyse de documents anciens à partir d'une ontologie 77
4.1 Ontologies, outils et généralités . . . 78
4.2 Logique de description et Ontologie . . . 78
4.2.1 Logique descriptive . . . 78
4.2.2 Ontologie . . . 80
4.2.3 Les Descriptions d'ontologie . . . 82
4.2.4 L'inférence ou comment extraire de la connaissance de la base . . . 84
4.3 Ontologie du domaine historien . . . 88
4.3.1 Définition et Description des lettrines par l'approche des historiens 89
4.4 Ontologie des traitements d'images . . . 92
4.4.1 Lettrines, images complexes - approche informatique . . . 92
4.4.2 Modélisation de lettrines : concepts et propriétés de l'ontologie . . . 93

III Réalisations pratiques et résultats 97

5 Étapes pour la réalisation du modèle d'analyse 99
5.1 Images de documents anciens - description des lettrines . . . 100
5.2 Simplification des images en couches . . . 103
5.2.1 Différentes approches . . . 104

6 Cas de la couche forme **115**
6.1 Extraction d'information à partir des formes de l'image 115
6.1.1 Extraction des motifs de l'image - Loi de Zipf 117
6.1.2 Sac de motifs et Recherche par le contenu - Approche statistique . 120
6.1.3 Segmentation de l'image et extraction de couches - Approche topologique 125
6.2 Validation de l'approche sur la couche forme 133
6.2.1 Comparaison avec une vérité terrain 133
6.2.2 Évaluation des mots visuels obtenus sur les formes de l'image . . . 136
6.2.3 Évaluation de la description obtenue à l'aide des graphes de régions 139
6.2.4 Complexification des descriptions statistiques et structurelles . . . 141

7 Cas de la couche texture **145**
7.1 Extraction d'information à partir des traits présents dans les images . . . 145
7.1.1 Extraction des traits : Un bref état des lieux 146
7.1.2 Extraction des traits : Approche proposée 147
7.1.3 Approche par sacs de traits 152
7.1.4 Extraction de zones de traits et construction de graphes 154
7.2 Validation de l'approche sur la couche traits 156
7.2.1 Validation de l'extraction des traits 157
7.2.2 Évaluation des sacs de traits 164
7.2.3 Évaluation de la description par graphes topologiques 165
7.2.4 Complexification des descriptions statistiques et structurelles . . . 166
7.3 Complexification de la description - mise en relation des informations sur les formes et des informations sur les traits 168

8 Ontologie finale et règles d'inférences **173**
8.1 Ontologie finale 173
8.2 Les outils de modélisation 174
8.2.1 Protégé 174
8.2.2 SWOOP 174
8.2.3 Jena 174
8.2.4 Oracle 175
8.2.5 Jess 175
8.3 Peuplement des ontologies 175
8.4 Règles d'inférences 178
8.4.1 Règles mises en places 179
8.4.2 Règles pour la validation du modèle du domaine historien 179
8.4.3 Règles pour la validation de l'ontologie des traitements d'images . 181
8.4.4 Règles pour la validation de l'ontologie finale - réduction du fossé sémantique 187

Conclusion **193**
Discussions et perspectives 195

Annexes **197**

A Physique du papier **199**
A.1 Le papier 199
A.2 L'encre 200

B Caractéristiques des lettrines **201**

Publications **209**

Bibliographie **211**

Table des figures

0.0.1 Image de matrice et de tampons utilisés dans l'imprimerie. Les premiers exemplaires, créés au début de l'imprimerie, étaient fabriqués en bois. Ce n'est qu'à partir du $XVIII_{ème}$ siècle, avec l'évolution de l'imprimerie, que les tampons et matrices ont été fabriqués en métal 2

0.0.2 Schéma global de la thèse 4

1.1.1 Schéma de fonctionnement général des moteurs de recherche par mots-clés 11

1.1.2 Exemple d'images correspondantes au mot-clé **avocat** 12

1.1.3 Exemple d'images correspondantes au mot-clé **roue** 12

1.3.1 Schéma d'indexation spatiale 14

1.4.1 Pavage temps-fréquence obtenu lors d'une transformée en ondelettes. La résolution temps-fréquence dépend de l'échelle d'analyse. 18

1.4.2 Exemple d'histogramme de projection vertical obtenu sur une image contenant des caractères pour les segmenter 19

1.4.3 Profils haut, bas, gauche et droite du caractère 3 20

1.4.4 Exemple de calcul de l'histogramme de forces utilisé dans la F-Signature [Wendling 02] 20

1.4.5 Calcul et recherche de correspondance du Shape context. a) et b) Exemples de contours discrétisés de deux formes. c) Quadrillage log-polaire utilisé pour dénombrer les points appartenant à chaque cadran. d), e), and f) Exemples de shape context, sous forme matricielle, obtenus avec les images a) et b) en partant des points de références indiqués sur les images. Enfin, e) présente la mise en correspondance des deux images obtenues à partir de l'approche présentée dans [Belongie 02] 24

1.4.6 Exemples d'images CSS obtenues avec leur maxima sur trois formes similaires en utilisant la méthode présentée dans [Mokhtarian 96]. A gauche, contour de l'image d'origine avec l'origine de l'algorithme indiqué par une croix noire. Au milieu, courbe CSS obtenue, et à droite graphique des maxima des courbes. Ce sont ces maxima qui sont utilisés pour mettre deux images en correspondance 25

1.4.7 Exemple de symbole utilisé dans [Coustaty 11]. De gauche à droite sont présentés, d'abord le symbole original, puis au milieu l'image obtenue à l'aide d'une squelettisation, et enfin à droite le résultat obtenu à l'aide d'une transformée de Hough adaptée 26
1.4.8 Exemple de polygonisation obtenue à l'aide de l'algorithme GOAL. L'image d'origine (à gauche) est simplifiée à l'aide de polygones isothétiques (1ère étape de simplification au milieu, résultat final à droite) 27
1.4.9 Exemple de construction d'un filtre de Gabor, avec la représentation imagée d'une gaussienne à gauche, d'un sinus au centre, et enfin du filtre obtenu à droite 30
1.4.10 Exemple de roses des vents calculées sur différentes images. a) objet graphique vertical, b) partie d'un squelette, c) quelques lignes de textes (titre), d) texte écrit de manière penchée 31
1.4.11 Exemple de construction du pavage de Voronoï à l'aide d'un algorithme par balayage : l'algorithme de Fortune [Fortune 86] 32
1.4.12 Système visuel humain - Image extraite de Wikipedia 35

2.2.1 Exemples de lettrines composées de traits 54
2.2.2 Exemples de lettrines en gros plan pour mettre en évidence les spécificités des images imprimées avec des tampons en bois 54
2.2.3 Les différentes structures d'un document, telle que définies dans [Journet 06] 57
2.2.4 Exemples d'annotation automatique de formulaires militaires en utilisant l'approche DMOS [Coüasnon 03]. A gauche, un exemple de formulaire militaire sur lequel a été ajouté une feuille, et à droite la structure reconnue par l'approche 58
2.2.5 Schéma de décomposition d'une image selon des radiogrammes. Dans cet exemple, l'image d'origine (le rectangle) est décomposée selon 6 orientations (sous-figure de gauche) et des histogrammes sont obtenus à partir des projections (sous-figure de droite) selon un principe similaire à la transformée de Hough [Duda 72] 59
2.2.6 Exemples de roses des directions obtenues sur différentes images : a) dessin vertical, b) un crâne, c) quatre lignes de texte, d) texte penché . 60
2.2.7 Impressions médiévales et leurs cartes de dissimilarités globales telles que définies dans [Baudrier 08]. Dans cet exemple, les cartes représentent les dissimilarités entre la première image (a) et les autres 61

3.2.1 Boucle d'analyse automatique d'un document ancien 73
3.2.2 Schéma global de structuration de la connaissance des traiteurs d'images 75

4.2.1 Exemple d'une ontologie sur les animaux. Elle illustre les relations de « sous-classe » et de « parent ». Les classes inférieures dans la hiérarchie représentent les relations subClassOf. 80

4.2.2 Exemple de triplet RDF 82
4.2.3 Relations spatiales utilisées pour observer l'agencement des formes dans l'image. Ces relations sont heritées de l'algèbre RCC-8. 88
4.3.1 Superposition de couches d'informations dans les documents anciens : exemple d'une lettrine 90
4.3.2 Hiérarchie de concepts utilisés dans l'ontologie du domaine Historien . . 91
4.4.1 Hiérarchie de concepts utilisés dans l'ontologie du domaine traitement des images 94
4.4.2 Description globale de l'analyse des documents anciens 95

5.1.1 Superposition de couches d'informations dans les documents anciens : exemple d'une lettrine 101
5.1.2 Exemples de fond, de motif, et de couleur de lettre dans les lettrines . . 102
5.1.3 Étapes d'extraction et de description du modèle proposé 104
5.2.1 Ensemble des couches obtenues avec la décomposition de Meyer. La sous-figure (b) présente les niveaux de gris de l'image de la sous-figure (a) en 3 dimensions (le niveau de gris correspond à la troisième dimension). La couche forme (U) apparaît en bas à gauche, tandis que la couche texture (V) apparaît en haut à droite. On voit clairement que les informations importantes des lettrines sont séparées au cours de cette décomposition 112
5.2.2 Exemples de couches obtenues sur des lettrines via la décomposition de Meyer 113

6.1.1 Graphique log/log de la fréquence d'apparition des mots par rapport à leur rang dans "Ulysses" de James Joyce (image issue de wikimedia) . . 118
6.1.2 Exemple d'images de la couche forme de Meyer en 256 niveaux de gris (1ère ligne) et en 3 niveaux de gris (2nde ligne). Dans le cas de la seconde ligne, les 3 niveaux ont été représentés par les couleurs rouge, verte et noire pour améliorer le contraste entre les niveaux. 121
6.1.3 Exemple de courbe TF-IDF obtenue sur une image de document ancien 123
6.1.4 Exemple de courbe associée à la Loi de Zipf de l'image présentée ci-contre 126
6.1.5 Exemple de couches obtenues avec la décomposition de Zipf 127
6.1.6 Exemple de résultat obtenu en appliquant la décomposition de Zipf soit sur l'image d'origine, soit sur la couche forme de Meyer. On peut observer que les composantes connexes extraites à partir de la couche forme de Meyer sont plus régulières et plus grandes. Ceci permet d'obtenir une meilleure compréhension de l'image du point de vue visuel 129
6.1.7 Exemples d'images et de représentations des différents traitements : a) Image originale ; b) Couche forme de Meyer (zones avec de faibles variations de niveaux de gris) ; c) Formes segmentées en utilisant la loi de Zipf (les formes apparaissent en bleue dans cette image) ; d) Plus grandes composantes connexes retenues 130

6.2.1 Exemple de lettrines avec les masques obtenus après saisie de la vérité terrain. Les masques (tâches noires) correspondent aux régions de formes dans l'image 134
6.2.2 Exemple d'image ou l'extraction automatique ne permet pas de couvrir l'ensemble de la surface des zones annotées dans la vérité terrain 135
6.2.3 Exemples de lettrines pour chaque style présent dans la base d'images utilisées par [Pareti 08]. Le style numéro 1 correspond aux lettrines composées de personnages dans une fond hachuré, le style numéro 2 est composé des lettrines avec un fond blanc et décoratif, le style numéro 3 aux lettrines avec un fond noir et des personnages, et enfin le dernier style contient toutes les lettrines qui n'appartiennent pas aux trois premiers styles 136
6.2.4 Exemples d'images de documents anciens composées de formes, appartenant aux 9 classes utilisées dans la seconde évaluation 138
6.2.5 Courbe rappel-précision obtenue sur la base d'images diverses de documents anciens 139

7.1.1 Exemples d'images de la couche texture obtenue à partir de deux lettrines. La première ligne présente l'image originale, la seconde la couche texture obtenue via la décomposition de Meyer, la troisième l'image binarisée, et enfin la dernière ligne présenté l'image débruitée 149
7.1.2 Orientation Principale (OP) obtenue pour chaque image 151
7.1.3 Exemple de squelette composé de plusieurs branches, et détection de la branche principale 151
7.1.4 Schéma général de construction du dictionnaire des mots visuels de traits dans les lettrines (partie supérieure), et de la comparaison avec une image requête (partie inférieure) 153
7.1.5 La première ligne présente l'image originale et la seconde présente les images segmentées où chaque couleur représente un rassemblement de traits appartenant à la même classe et proches spatialement (la même couleur n'est pas associée à la même classe dans toutes les images) . . . 155
7.1.6 Exemple d'extraction et de regroupement de traits dans des lettrines . 156
7.2.1 Exemples de lettrines sous et sur-segmentées 159
7.2.2 Exemple de lettrines avec les masques obtenus après saisie de la vérité terrain. Les masques (tâches noires) correspondent aux zones de traits dans l'image 160
7.2.3 Comparaison entre les segmentations obtenues avec l'approche proposée et une segmentation avec des filtres de Gabor. Chaque couleur correspond à une classe de trait dans l'image (une même couleur dans deux images différentes ne représente pas forcément la même classe) 162

7.2.4 Exemple d'image où la segmentation à l'aide de filtres de Gabor donne des résultats inférieurs. On peut voir clairement qu'une partie des régions de traits est omise dans la phase d'extraction 163
7.2.5 Exemple d'images de régions de traits obtenus à l'aide de l'approche proposée et de la segmentation basée sur des filtres de Gabor. À gauche on peut trouver l'image d'origine, au centre les traits extraits avec notre Cette étape approche, et enfin à droite les traits extraits avec les filtres de Gabor 163

8.3.1 Interface utilisée par les historiens pour saisir leur connaissance sur les images de lettrines 176
8.3.2 Exemples de régions extraites automatiquement en utilisant les traitements présentés dans la partie 5. A gauche, image d'origine, et à droite les régions obtenues (une couleur par région) 177
8.3.3 Exemple de découpage d'une image en neuf sous-parties pour le calcul des relations spatiales entre les régions et le masque défini. Chaque couleur correspond à une région extraite automatique par les algorithmes présentés dans la section 6.1 178
8.4.1 Sous-parties de l'image utilisées pour localiser les formes dans l'image (régions au centre, ou plutôt sur les bords) 183
8.4.2 Comparaison entre l'extraction de la lettre à l'aide de la binarisation à partir du critère de Otsu [Otsu 79], et de la segmentation obtenue à partir de la loi de Zipf. On peut remarquer que la principale différence repose sur le fait que la méthode développée par Otsu suppose que les formes sont blanches sur fond noir 185
8.4.3 Comparaison entre les résultats de binarisation obtenus avec le critère d'Otsu et la segmentation basée sur la Loi de Zipf. La binarisation obtenue avec le critère définit par Otsu classe chaque pixel comme appartenant au fond ou à la forme à partir de sa valeur. La segmentation obtenue par une loi de Zipf, quant à elle, repose sur l'analyse de la fréquence des motifs qui composent l'image. Les grandes zones comme la lettre peuvent ainsi être extraites même si elle est connectée au fond (les liens correspondent à des motifs moins fréquents) 186
8.4.4 Exemples de résultats d'extraction de la lettre. La première colonne présente les images simplifiées, la seconde correspond aux images obtenues avec l'algorithme associé à la Loi de Zipf. Enfin, la troisième colonne correspond à la lettre extraite. Chaque ligne présente les résultats obtenus pour chaque approche testée et les différences entre les deux résultats justifient les différences dans les taux de reconnaissance 188
8.4.5 Exemples d'images très difficiles à traiter pour l'extraction de la lettre. La première ligne présente l'image d'origineT, tandis que la seconde présente la forme qui a été considérée comme la lettre par notre système . . 189

8.4.6 Un exemple d'image ou les zones de formes ont été annotées avec les règles *isLetter* et *isBody* 190
8.4.7 Exemple d'image pour laquelle la partie du corps n'a pas été correctement extraite 191

Liste des tableaux

1.1 Comparatif de différentes méthodes de caractérisation de formes 23
1.2 Comparatif de différentes méthodes de la littérature pour l'analyse et la description des contours 29
1.3 Comparatif entre les 5 grandes catégories de méthodes d'extraction et de description de textures 34
1.5 Tableau récapitulatif des différents projets de recherche d'images par le contenu en fonction de leur domaine d'application 45

2.1 Tableau récapitulatif des méthodes de description d'images de documents anciens 62

5.1 Tableau comparatif des méthodes présentées 107

6.1 Rappel et précision obtenus avec notre méthode d'extraction de traits sur un sous-ensemble de 45 lettrines accompagnées de leur vérité terrain. Le critère de rappel permet de mettre en avant la capacité de notre approche à extraire les traits marqués dans la vérité terrain, pendant que la précision permet de connaître le pourcentage de surface des régions de la vérité terrain qui est couverte par les traits extraits automatiquement 135
6.2 Taux de reconnaissances obtenus avec notre méthode sur la base de lettrine utilisée dans [Pareti 06a] pour différentes valeurs de K dans l'algorithme des K-Plus Proches Voisins 137
6.3 Taux de reconnaissances obtenus sur la base d'images diverses de documents anciens 138
6.4 Taux de reconnaissances obtenus sur la base de lettrines utilisées dans [Pareti 06a] en utilisant une signature structurelle sur les zones de formes, et une distance hybride entre graphes 141
6.5 Taux de reconnaissance des lettrines à partir de la couche forme en utilisant une description recomplexifiée des lettrines 142

7.1 Exemples de résultats de sélection des zones de traits en utilisant différents critères. Les régions sélectionnées correspondent au sous-ensemble de région minimaliste qui contiennent des traits 157

7.2 Nombre d'images bien, sous, ou sur segmentées sur un sous-ensemble de 228 images 158
7.3 Rappel et précision obtenus avec notre méthode d'extraction de traits sur un sous-ensemble de 45 lettrines accompagnées de leur vérité terrain. Le critère de rappel permet de mettre en avant la capacité de notre approche à extraire les traits marqués dans la vérité terrain, pendant que la précision permet de connaître le pourcentage de surface des régions de la vérité terrain qui est couverte par les traits extraits automatiquement 161
7.4 Comparaison quantitative globale entre notre approche et celle basée sur les filtres de Gabor utilisant les critères de rappel et de précision définis page **??** 164
7.5 Taux de reconnaissances obtenus avec l'approche statistique proposée pour décrire des images de documents anciens à partir des traits qui les composent. Pour chaque valeur de K utilisé dans l'algorithme des K plus proches voisins, nous présentons les taux de reconnaissances obtenus. Dans le cas ou K est supérieur à 1, la classe associée à chaque image correspond à la classe majoritaire parmi l'ensemble des résultats 165
7.6 Taux de reconnaissance obtenus sur la base de lettrines utilisées dans [Pareti 06a] en utilisant une signature structurelle sur les zones de traits, et une distance hybride entre graphes 166
7.7 Taux de reconnaissance des lettrines à partir de la couche texture en utilisant une description re-complexifiée des lettrines 167
7.8 Taux de reconnaissance en utilisant une description recomplexifiée des lettrines, qui mélange une description statistique issue de la couche forme, et une de la couche texture 168
7.9 Taux de reconnaissance en utilisant une description recomplexifiée des lettrines, qui mélange une description structurelle issue de la couche forme, et une de la couche texture 169
7.10 Taux de reconnaissance en utilisant une description recomplexifiée des lettrines, qui mélange une signature structurelle provenant de la couche texture, et une signature statistique provenant de la couche forme . . . 170
7.11 Taux de reconnaissance obtenus en utilisant une description recomplexifiée des lettrines qui mélange les résultats obtenus avec les graphes issus de la couche forme, et ceux de l'approche par sacs de traits 170
7.12 Taux de reconnaissance utilisant une description recomplexifiée des lettrines suite à un vote à la majorité de toutes les caractéristiques (fusion globale) 171

8.1 Taux de reconnaissance des lettres extraites dans les lettrines en utilisant deux approches d'extraction de la lettre différentes et deux OCR 184
8.2 Taux de reconnaissance de la lettre dans les lettrines en utilisant deux approches de simplification d'images différentes et un OCR 187

Introduction

La recherche d'image est très populaire sur internet. De nos jours, il est très courant de rechercher une image pour illustrer une présentation, un rapport, ou tout autre document. Cependant, la profusion d'images numériques disponibles sur internet impose de pouvoir les catégoriser et de les retrouver.

De manière analogue, les entreprises manipulent tous les jours des documents pour échanger entre leurs services. Ainsi, ce sont des millions de documents qui sont créés, imprimés et édités tous les jours sur la planète. Afin de rationnaliser leurs coûts, de nombreuses organisations publiques et privées lancent des campagnes de dématérialisation de leurs documents pour accélérer leurs flux de traitements. Parmi de nombreux exemples, nous pouvons citer le secteur médical qui s'oriente vers le dossier médical numérique accessible à tous les médecins, mais également les grandes entreprises et administrations qui souhaitent pouvoir réduire l'encombrement de leurs archives et y retrouver de l'information rapidement.

Cette dématérialisation requiert le développement d'outils stratégiques de comparaison de documents, d'aide à la décision, et de navigation dans ces corpus d'informations faiblement structurées. De plus, les documents sont spécifiques à chaque domaine et cas d'utilisation. Cette spécificité doit donc être nécessairement prise en compte dans les modalités de recherche, et permettre d'intégrer une connaissance liée à des experts du domaine. Selon la littérature, la recherche d'images est catégorisable selon trois procédés :

1. recherche par **navigation libre** dans la base d'image : l'ensemble des images de la base sont présentées à l'utilisateur qui la parcourt pour rechercher une image. Cette technique n'est que peu utilisée puisqu'elle impose de naviguer longtemps au sein de la base sans garantir de résultat à l'utilisateur. Cependant, dans certains cas d'utilisation précis, ce type de recherche est employé, comme la recherche de Cliparts sous différents logiciels d'édition de textes ou d'affiches.

2. recherche par **mots-clés** : le principe général consiste à associer des mots-clés aux images (manuellement ou (semi-)automatiquement) et à rechercher des images à partir de ces mots-clés. Cette technique est à la base des moteurs de recherche utilisés tous les jours par des millions de personnes. Nous reviendrons rapidement sur cette catégorie au début du chapitre 1.

3. recherche par le **contenu** : plus communément appelé Content-Based Image Retrieval (CBIR), cette technique utilise des méthodes qui décrivent le contenu des images. C'est un domaine de recherche très actif depuis les années 1990, qui propose des outils génériques pour représenter et rechercher des images de scènes naturelles. La section 1.3 présente un état de l'art des approches de description des images pour une indexation par le contenu. Malheureusement, ces méthodes sont trop généralisées et inopérantes pour les images de documents [1].

Dans le cadre de cette thèse, nous nous intéressons à la recherche par le contenu d'images graphiques de documents de la Renaissance ($XV_{ème}$ et $XVI_{ème}$ siècles), ayant été fournies par le Centre d'Études Supérieures de la Renaissance (CESR) de Tours. Cette période, qui inclut le début de l'imprimerie, a vu l'apparition de documents imprimés avec des caractéristiques particulières. Tout d'abord, il faut noter que les documents étaient imprimés à l'aide de tampons en bois. Ceux-ci étaient assemblés sur une matrice [2], encrés, puis pressés sur le papier pour y déposer les caractères. Un exemple de caractères assemblés sur une matrice peut-être observé dans la figure 0.0.1.



FIGURE 0.0.1: Image de matrice et de tampons utilisés dans l'imprimerie. Les premiers exemplaires, créés au début de l'imprimerie, étaient fabriqués en bois. Ce n'est qu'à partir du $XVIII_{ème}$ siècle, avec l'évolution de l'imprimerie, que les tampons et matrices ont été fabriqués en métal

Ces premières impressions se faisaient en noir et blanc, sans possibilité de créer de niveaux de gris (les variations de gris sont utiles pour créer des impressions de reliefs et de nuances). Afin de palier à ce problème, les imprimeurs de l'époque ont utilisé une

1. La non adéquation de ces méthodes avec les documents anciens sera présentée par la suite
2. support en bois, puis métallique à partir du $XVIII_{ème}$ siècle, qui servait de guide pour tenir les lettres ensembles

technique similaire au tramage utilisé actuellement en informatique [3]. Le tramage est une technique qui consiste à mélanger des points noir et blanc de manière assez rapprochée pour donner une impression de gris. Similairement, les créateurs de tampons de l'époque dessinaient des traits fins légèrement espacés. Ceci donnait une impression de relief et de variations de teintes.

Il faut également noter que les documents ont été créés sur du papier, qui évolue et se dégrade à travers le temps. Il présente des traces de vieillissement telles que le jaunissement du papier, la fragilisation, des déchirements, des arrachements ou l'amincissement des pages. Cependant, le principal agent de dégradation des documents reste l'homme, puisqu'à chaque manipulation d'un livre il peut risquer de l'endommager. Numériser ces documents présente donc l'intérêt d'éviter les contacts directs avec l'ouvrage et ainsi une meilleure conservation, et permet une diffusion de son contenu de manière plus large.

Les images de ces documents, mémoires de nos sociétés et de notre passé, sont utilisées par les historiens pour identifier l'origine des livres (date d'impression, lieu d'impression, ...) et retracer sa vie. De nombreux projets de numérisation sont actuellement menés en France et en Europe [Europeana , Google b, Bibliotheek 11, BLSB , Gallica , OCA , IMPACT] pour sauvegarder le contenu de ces dizaines de milliers d'ouvrages anciens. Numériser ces documents consiste à l'heure actuelle à sauvegarder leur contenu sous format image inerte, sans proposer réellement d'outils pour indexer et naviguer dans cette masse gigantesque de pages. Afin de proposer la consultation de ces documents au plus grand nombre, il est indispensable de développer des services qui permettent d'extraire automatiquement de l'information pour valoriser ces images.

Cette thèse s'inscrit dans le projet Navidomass (ANR-06-MDCA-012) qui a pour but de valoriser le patrimoine écrit français de la Renaissance. Cette valorisation propose d'extraire les éléments des pages, de les reconnaître, pour annoter automatiquement les images. Dans le cadre de cette thèse, nous nous sommes particulièrement intéressés aux images graphiques qui sont des objets complexes puisque composées de différentes couches d'informations (composées de traits et de formes). Les historiens, partenaires de ce projet, souhaite pouvoir réaliser deux types de requêtes sur les lettrines. La première catégorie de requête consiste à identifier une image en recherchant celles qui sont similaires dans la base (recherche par l'exemple). La seconde catégorie consiste à rechercher des éléments très particuliers dans les lettrines, pour les aider à identifier des concepts dans les images (image décorative, image composée de personnages, ...).

Ces deux types de requêtes relèvent de domaines très différents qui imposent d'être capable de comparer des images, et d'en extraire une information sémantique pour annoter les images. Nous commençons donc ce manuscrit par un état de l'art des différents types de recherche d'images qui existent, et de spécification des requêtes. Afin de traiter

3. A ne pas confondre avec le tramage utilisé en imprimerie qui illustre les différences de contraste qui apparaissent entre la visualisation sur écran et l'impression

de grands volumes de documents, il est indispensable de proposer des services qui structurent les documents à partir de leur contenu. Nous proposons donc, dans un deuxième temps, un état de l'art des méthodes existantes dans la littérature pour décrire le contenu des images. Enfin, nous présentons les méthodes de la littérature qui se sont spécialement intéressées aux documents anciens et à leur caractéristiques particulières.

Bien que de nombreuses approches se soient intéressées aux documents anciens, aucune ne propose réellement de s'adapter à leur complexité, et d'intégrer des connaissances propres aux historiens et propres au traiteurs d'images. Nous proposons dans la seconde partie de ce manuscrit une présentation des problèmes liés à la représentation de connaissances d'experts d'un domaine, et une présentation de notre modèle ontologique d'analyse complexe d'images de documents anciens. Ce modèle repose sur un paradigme de complexité énoncé par Edgar Morin [Morin 96], et permet d'intégrer dans une même base les connaissances propres aux historiens, et les connaissances extraites automatiquement par des algorithmes de traitements d'images. Le but de cette base de connaissance est de proposer à l'utilisateur des recherches en indiquant soit un exemple d'image, soit des caractéristiques issues de l'expertise des historiens. Nous présentons les deux ontologies créées pour représenter et hiérarchiser ces connaissances.

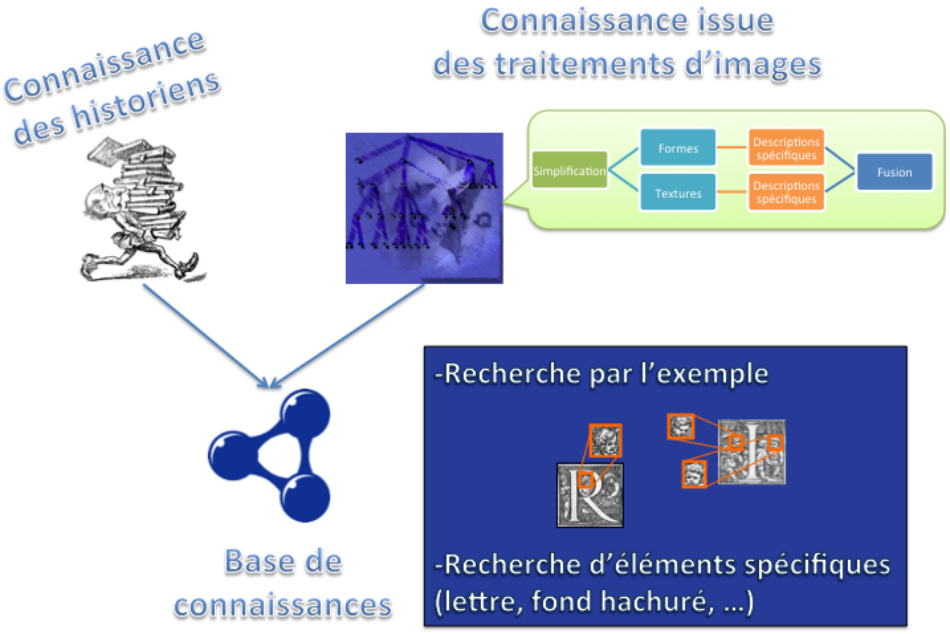


FIGURE 0.0.2: Schéma global de la thèse

La troisième partie de ce manuscrit présente en détails les détails des algorithmes qui ont été définis pour décrire de manière complexe et adaptée les images de documents anciens. Étant donné la nature particulière des images traitées (documents créés par l'homme pour l'homme), il nous a paru intéressant de décrire chaque couche d'information selon deux approches. Une première statistique, qui recherche des motifs pertinents pour décrire soit les formes, soit les traits, et une seconde structurelle, visant à extraire des régions d'intérêts dans les images et à décrire leur agencement spatial. La dernière étape dans la description des images consiste à re-complexifier la description en fusionnant les différentes modalités énoncées pour obtenir une description complexe des images.

La dernière section de ce manuscrit est dédiée à l'ontologie finale qui permet de rassembler au sein d'une même base la connaissance des historiens et celle obtenue via les traitements des images. Nous présentons donc comment nous déduisons de la connaissance des algorithmes de traitements d'images pour permettre des requêtes complexes, et ainsi réduire le fossé sémantique entre les deux domaines. Le schéma de la figure 0.0.2 présente ces différentes étapes.

Première partie

Indexation d'images par le contenu / État de l'art

Chapitre 1

Description et recherche des images par leur contenu

Décrire des images par leur contenu n'est pas chose aisée. Cela consiste à résumer le contenu de l'image par des mots ou une signature, pour pouvoir comparer les images. Ainsi, deux images décrites par les mêmes mots, ou la même signature, seront considérées comme identiques. Ce chapitre présente un tour d'horizon des méthodes existantes pour décrire et comparer les images. Ainsi, dans une première section, nous présentons brièvement l'idée générale des approches par mots-clés, et les spécificités des images étudiées au cours de cette thèse. Puis, dans un second temps, nous nous attarderons plus particulièrement sur la recherche d'image par le contenu. Ainsi, nous présentons les différents attributs existants pour décrire une image et leur adéquation avec les spécificités des documents anciens. Enfin, la fin de ce chapitre présente une vue synthétique de nombreux projets existants en recherche d'image par le contenu.

1.1 Description des images et Association de mots-clés

De nombreux outils existent pour indexer et rechercher des documents textuels, des sites internet ou des images. On peut citer Google, Microsoft Live Search, Flickr ou encore Yahoo qui sont les plus connus. Tous ces outils indexent des textes en résumant leur contenu par des mots-clés. L'extraction de mots-clés repose généralement sur l'utilisation de statistiques (comme la recherche des mots les plus fréquents dans l'algorithme Term Frequency[Jones 72] ou des mots les plus discriminants dans un corpus avec l'algorithme Inverse Document Frequency[Croft 88], ou encore la recherche des mots et liens qui mènent à un document utilisé dans les algorithmes de Web Ranking[Sergey 06]).

De manière analogue, il est possible d'associer des mots clés aux images. Généralement, ces mots clés résument leur contenu, et servent d'index pour retrouver les images. Ainsi, rechercher une image consiste alors à identifier toutes les images résumées par ces mêmes mots clés. Cette technique présente l'avantage de pouvoir fonctionner via les moteurs de

recherches habituels. La difficulté majeure repose sur la sélection des mots-clés.

Dans le cadre de cette thèse, nous nous intéressons plus particulièrement à la recherche par le contenu d'images graphiques de documents de la Renaissance ($XV_{ème}$ siècle). Dans ce cas particulier d'images, les historiens choisissent des mots-clés dans un thesaurus, qui constitue une base de mots clés structurée permettant de résumer le contenu d'une image. L'association des mots clés reste cependant manuel.

Associer manuellement des mots clés aux images est une étape fastidieuse et coûteuse. Afin d'apporter des réponses à ce problème, des méthodes cherchent à associer automatiquement des mots-clés aux images pour les décrire. Nous les présentons ci-dessous. Dans un premier temps, nous allons présenter les moteurs de recherche connus qui associent des mots-clés aux images à partir de leur contexte (mots qui entourent l'image, méta-données associées aux images, ...). Puis, nous présenterons trois méthodes particulières (ALIPR, CHABOT, et BEHOLD), qui associent des mots-clés aux images en accord avec leur contenu.

1.1.1 Moteur de recherche par mots-clés

Ces moteurs de recherches sont très répandus et utilisés à travers le monde. Les plus connus (Google, Bing, Yahoo) représentent plus de 94% des recherches faites sur internet[Comscore, Inc] et proposent différentes options de recherche spécifiques aux images. Leur fonctionnement général est résumé dans la figure 1.1.1.

La recherche d'éléments sur internet se décompose en deux grandes étapes : l'indexation et la recherche. L'indexation consiste à organiser la base d'images pour pouvoir les parcourir et rechercher à l'intérieur aisément. L'indexation peut-être vue comme l'organisation d'une bibliothèque où chaque livre est étiqueté pour être retrouvé rapidement.

Il est donc important de bien choisir la manière dont sont construits les index, résumés des documents. Dans le cas des moteurs de recherche orientés textes, l'index est généralement conçu à partir du nom du fichier, des mots qui entourent l'image et des méta-données contenues dans le fichier lui-même (balise ALT en HTML ou méta-données EXIF ou IPTC [EXI , IPT]). Ces moteurs de recherche sont également basés sur des algorithmes de WebRanking[Sergey 06] et l'importance d'une image est alors liée aux thématiques des pages web qui y mènent et en proviennent.

Pour résumer, nous pouvons dire que les moteurs de recherche d'images annotent automatiquement les images à l'aide de mots-clés. Les mots-clés sont extraits du voisinage des images (mots qui entourent les images, méta-informations liées aux images, ...). Cependant, un problème de fossé sémantique apparaît rapidement puisque rien ne garantit la fiabilité des mots qui sont extraits, leur pertinence pour décrire une image, et les problèmes de polysémie des mots. Des exemples d'images annotées avec le même mot-clé sont présentés dans les Figures 1.1.2 et 1.1.3.

Les moteurs de recherches par mots-clés les plus célèbres sont Bing (développé par Microsoft Corporation), Google (fondé par Larry Page et Sergei Brin [Sergey 06]), et Yahoo (qui proposait initialement un annuaire de sites internet).

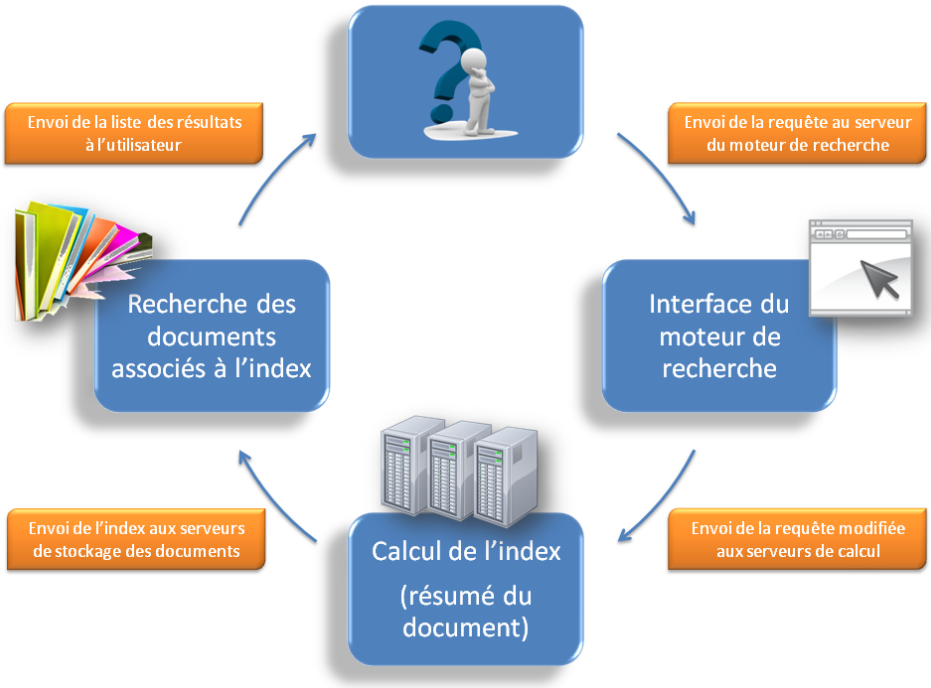


FIGURE 1.1.1: Schéma de fonctionnement général des moteurs de recherche par mots-clés

1.1.2 Annotation semi-automatique des images

Dans le monde des moteurs de recherche par mots-clés, de nombreuses approches proposent également d'extraire des mots-clés automatiquement à partir du contenu des images (une liste non exhaustive pourrait comporter les projets suivant : [Ogle 95, Heck 99, Yavlinsky 07, Li 08, Brut 09]). Ceux-ci proposent ainsi à l'utilisateur de le guider dans son annotation des images, en lui proposant des mots-clés déduits à partir des caractéristiques radiométriques calculées sur les images. De plus, ces approches tendent à réduire le fossé sémantique puisque le vocabulaire utilisé pour annoter est généralement restreint, et les requêtes consistent à choisir des mots-clés du vocabulaire pour rechercher les images qui s'y rattachent.

D'autres approches d'aide à l'annotation reposent sur des systèmes avec interactions utilisateurs (le lecteur intéressé par ce point pourra trouver de nombreuses références dans [Nguyen 11]). En effet, toute recherche via un système de recherche d'images par le contenu implique la formulation d'une requête. Cette requête permet d'indiquer plus ou moins précisément au système ce que recherche l'utilisateur. Cependant, ceci implique également que la requête soit formulable et que l'utilisateur soit capable de fournir une requête valide et appropriée. Proposer une interaction entre le système et l'utilisateur revient à modifier le comportement de la machine pour répondre plus finement aux



FIGURE 1.1.2: Exemple d'images correspondantes au mot-clé **avocat**



FIGURE 1.1.3: Exemple d'images correspondantes au mot-clé **roue**

attentes de l'utilisateur.

Le bouclage de pertinence est un mécanisme très répandu dans la recherche d'image avec interactions utilisateurs [Ishikawa 98, Ortega-Binderberger 04, Kim 05, Tao 06, Picard 08]. On peut ainsi citer des travaux qui visent à améliorer les résultats à court-terme, c'est à dire qui adaptent le système pour un utilisateur et une requête spécifique [Ishikawa 98, Ortega-Binderberger 04, Kim 05], et des approches plus orientées sur l'adaptation à long terme du système [Duygulu 06, Barnard 03, Jeon 03, Monay 03, Yavlinsky 07, Picard 10]

Enfin, afin de proposer des requêtes basées sur les mots clés et sur le contenu des images, des approches hybrides ont été proposées dans la littérature [Barrat 09, Nguyen 09]. Ces approches comparent des images sur la base de ces deux modalités (mots clés et descripteurs de leur contenu), et permettent ainsi de reconnaître des images en combinant des mots clés et des descripteurs d'images (un exemple de requête pourrait ressembler à « voiture avec beaucoup de bleu »). Google images [Google a] propose également de rechercher une image en indiquant des mots clés, ou en proposant un exemple d'image. Bien que cette méthode soit hybride, aucune information à ce jour n'a pu être trouvée sur le fonctionnement de leur système.

1.2 Images étudiées dans le cadre de cette thèse

Dans le cadre de cette thèse, nous nous intéressons plus particulièrement à la recherche par le contenu d'images graphiques de documents de la Renaissance ($XV_{ème}$ et $XVI_{ème}$

siècles). Cette période, qui inclut le début de l'imprimerie, a vu l'apparition de documents imprimés avec des caractéristiques particulières.

Tout d'abord, ces impressions étaient réalisées à l'aide de tampons en bois (pour les enluminures) ou métalliques (pour le texte), et se faisaient en noir et blanc (sans possibilité de créer de nuances de gris). Les créateurs de tampons de l'époque dessinaient donc des traits fins légèrement espacés pour donner une impression de relief et de variations de teintes. Enfin, les documents ont été créés sur du papier qui, en évoluant au cours du temps, s'est dégradé (jaunissement du papier, fragilisation et déchirements des feuilles, arrachements ou l'amincissement des pages, ...) et est source de bruits dans les images.

Toutes les images et tous les documents que nous étudions dans ce manuscrit nous ont été fournis par le Centre d'Études Supérieures de la Renaissance [BVH] de Tours, et présentent ces caractéristiques.

De nombreuses méthodes ont été développées dans la littérature pour décrire et résumer le contenu des images. Ces outils ont tout d'abord été développés pour analyser les images en général (des images de scènes naturelles, des photos, ...), qui présentent des caractéristiques bien différentes de celles énoncées ci-dessus. Le but de l'analyse présentée dans ce chapitre est d'étudier, en fonction du type d'attribut, l'adéquation des méthodes de la littérature avec les caractéristiques des documents anciens. Nous présentons les outils classiques de la littérature dans la suite de cette partie, et leurs caractéristiques principales. Ainsi, nous commençons par une présentation générale de la recherche d'image par le contenu. Puis, dans un second temps, nous présentons les différents types d'attributs utilisés pour décrire ces images.

1.3 Présentation générale de la recherche d'images par le contenu

La description d'une image par le contenu vise à résumer son contenu. On peut les retrouver dans la littérature sous le nom de CBIR, pour Content-Based Image Retrieval, ou bien recherche d'images par le contenu. Toutes ces techniques sont développées pour proposer quatre types de services aux utilisateurs :

- **Navigation dans des catégories pré-définies** : l'utilisateur navigue dans la base en choisissant des catégories pré-définies au moment de la construction de la base. Cette navigation impose de connaître les différentes catégories avant d'indexer la base et que toute la base soit indexée sur ces catégories.
- **Recherche par l'exemple** (également connue sous l'appellation QBE ou Query by Example en anglais) : l'utilisateur propose une image en exemple de ce qu'il recherche. Après avoir résumé le contenu de l'image, le système recherche les images les plus similaires, du point de vue du résumé (descripteurs de l'information contenue par les pixels).

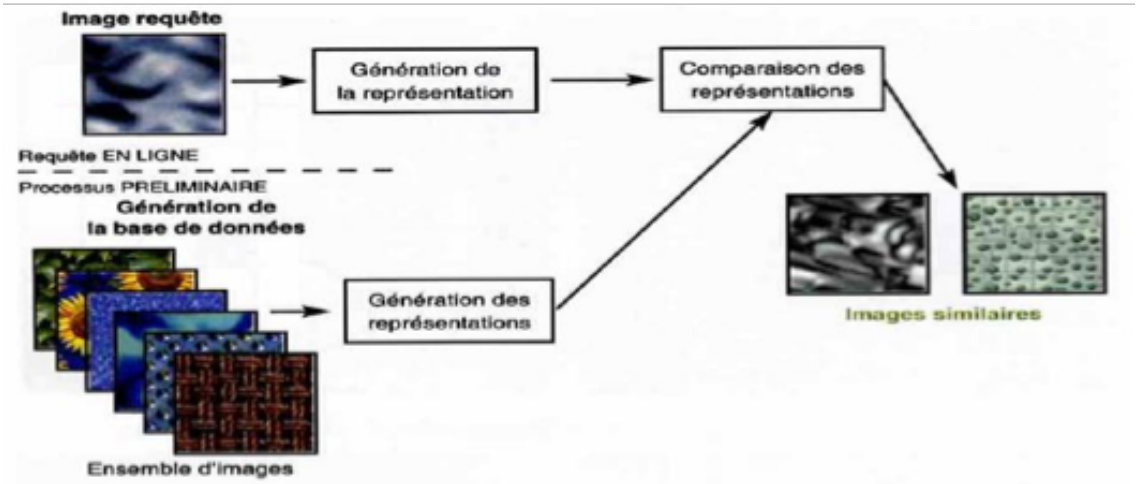


FIGURE 1.3.1: Schéma d'indexation spatiale

- **Recherche par esquisse** : approche similaire à une recherche par l'exemple mais dans ce cas précis, l'utilisateur ne propose pas une image mais dessine une esquisse de ce qu'il recherche. L'esquisse peut-être basée sur une forme particulière ou une couleur comme le propose RetrievR [1], service du célèbre site FlickR [2].
- **Spécification de caractéristiques visuelles** : l'utilisateur précise implicitement des caractéristiques qui résument l'image d'un point de vue bas-niveau. C'est-à-dire que les caractéristiques précisées n'ont pas de sens direct avec le contenu de l'image. Par exemple, l'utilisateur précise la taille, la couleur moyenne ou l'aspect des textures de l'image.

Dans la suite de ce chapitre, nous présentons les techniques d'analyse et d'indexation d'images par le contenu. L'objectif dans cette étape est d'extraire et de caractériser la structure d'images graphiques pour la navigation dans des masses de données. Cela pose plusieurs problèmes théoriques difficiles qui sont synthétisés dans la Figure 1.3.1

Les différents problèmes posés sont :

- **Génération de la représentation** : Il s'agit d'extraire de l'image des signatures caractéristiques qui la résument. Il existe différentes stratégies qui conditionnent la qualité de la navigation dans ces masses de données et le temps de traitement. Ces différentes caractéristiques peuvent être de trois types :
  - Globales [Adam 00b, Zhang 02, Pareti 06c, Tabbone 06, Rusiñol 10]
  - Spatiales/topologiques [Renz 02, Rusiñol 06, Rusiñol 09a, Coustaty 11]
  - Locales/semi-locales [Lowe 99, Bay 06, Kisku 10] ;

- **Comparaison des représentations** : une fois une représentation extraite des

1. Retrievr : http ://labs.systemone.at/retrievr/
2. FlickR : http ://www.flickr.com/

images, il s'agit de **mesurer la similarité entre deux images** à partir de leur représentation. Suivant la nature de la représentation, on se trouve confronté à un **dilemme qualité/rapidité**. Ces méthodes de comparaison sont étroitement liées à la représentation choisie (statistique [Cover 67, Chowdhury 10, McQueen 67]ou structurelle [Jouili 11, Raveaux 10, Raveaux 11, Bunke 11]) ;

– **Problématique de la masse de données** : les enjeux de cette thèse portent sur la navigation dans de **grandes masses de documents**. En plus de la problématique de la représentation et de leur comparaison, il s'agit de mettre en place une structuration optimale de l'information pour éviter la comparaison exhaustive.

Nous commençons tout d'abord par un état des lieux des approches utilisées pour générer la représentation des images.

1.4 Attributs pour décrire un contenu

La recherche d'objets dans des images est très largement utilisée par les particuliers et les industriels. Les premiers systèmes mis au point reposaient exclusivement sur la recherche de motifs prédéfinis à partir de corrélations [Derin 87, Rui 99, Smeulders 00]. Ces approches présentaient l'inconvénient de ne pas être invariantes aux changements d'illuminations, d'échelles ou aux rotations.

Afin d'apporter une réponse à ce problème, des travaux de recherche [Rui 99, Datta 08, Yang 08] ont proposé d'extraire des caractéristiques globalement sur l'image. Ces descripteurs résument le contenu des images par des statistiques calculées sur les valeurs des pixels, et présentent l'avantage d'être invariants en rotation et changements d'échelles quand ils sont normalisés. La recherche consiste alors à une comparaison de ces descripteurs. De nombreuses caractéristiques ont été proposées dans la littérature mais celles-ci ne permettaient pas de décrire l'organisation spatiale des objets contenus dans les images.

Une alternative aux recherches d'images décrites globalement consiste à extraire des éléments structurants des images et à décrire leur agencement [Tangelder 04, Yang 08]. De nombreuses approches ont été proposées en utilisant des segments, des contours, ou des régions comme éléments structurants. La description revient alors a indiquer les relations spatiales ou topologiques qui unissent ces éléments. Les caractéristiques spatiales fonctionnent bien pour certaines catégories d'images (telles que les images de symboles, les images composées de régions identiques). Dans le cas général d'images de scènes naturelles, ou composées de plusieurs éléments différents, il devient difficile de pouvoir désigner les primitives structurelles, de les segmenter facilement et de décrire leur agencement.

Pour répondre à ce problème, de nombreux travaux, basés sur des des approches semi-locales, ont vu le jour depuis la fin du 20ème siècle. Une approche semi-locale va tout d'abord extraire des zones d'intérêts et les décrire (en utilisant des descripteurs semblables à ceux des approches globales). C'est la concaténation de ces descriptions qui

permet de décrire une image. Dans certains cas particuliers, la description de l'agencement des zones d'intérêts est conjuguée à la description des zones elles-mêmes. Des exemples pourront être trouvés dans 1.4.3

Nous présentons un panorama des trois grandes catégories d'approches qui ont pu être présentées au cours des 50 dernières années dans cette section.

1.4.1 Approches par description globale

Très largement utilisées dans la littérature, ce type de méthode vise à calculer des attributs sur l'ensemble des pixels de l'image pour les décrire. Elles décrivent les images en résumant son contenu à partir de la couleur [Yue 11, Wang 11], de corrélations [Shechtman 07], de traitements particuliers (redimensionnement d'images[Torralba 08], caractérisation des sous-éléments d'objets [Coustaty 08]), de caractéristiques de formes génériques [Tabbone 06, Zhang 02], ou d'approches basées sur la texture [Lowe 99, Choksuriwong 05, Journet 08b]. Cette description est généralement représentée sous forme d'un vecteur de caractéristiques qui tente de décrire de la manière la plus unique et la moins ambigüe possible chaque classe d'images. Ce vecteur est alors utilisé pour indexer l'image, ou pour comparer des images via des distances entre vecteurs. Elles ont longtemps été les plus performantes dans des processus de reconnaissance de formes hors contexte.

Nous présentons, dans cette section, les grandes catégories d'approches existantes et les méthodes les plus utilisées dans la littérature. Pour chaque catégorie, nous présentons les descripteurs existants, nous spécifions ceux adaptés au cas des documents anciens, et une synthèse de ces descripteurs.

1.4.1.1 Représentation des formes

Présentation des descripteurs

Caractériser le contenu d'une image à partir de descripteurs est une tâche difficile de par la grande variabilité de formes possibles. La communauté de chercheurs qui travaille sur la recherche d'images par le contenu s'est inspirée des techniques de représentation des formes issues de la reconnaissance des formes. Les premières approches consistaient à reconnaître des images composées d'une seule forme blanche par image (les formes blanches peuvent être vues comme la silhouette des objets à reconnaître). Les premiers travaux du début des années 1960, consistaient à comparer les images elles-mêmes entre elles à l'aide de distances (Hausdorff[Alhichri 02, Baudrier 07b], Vinet [Karaf Falah 93], Baddeley [Coquin 01]), ou à rechercher des motifs particuliers par « matching » entre des motifs et l'image requête. Ces deux types d'approches présentent l'inconvénient de ne pas être invariantes aux rotations, aux changements d'échelles, et aux translations. De plus, avec l'augmentation de la taille des images, et du nombre de couleurs qui les composent, ces traitements deviennent particulièrement lourds en temps de calcul.

Les méthodes de caractérisation d'images dans leur globalité ont été initialement développées pour des images binaires, elles ont depuis été adaptées aux images en niveaux de gris.

Une première catégorie de descripteurs, les **moments géométriques**, sont apparus. Introduits par Hu [Hu 62], ces descripteurs ont fait l'objet de larges investigations et ont été largement utilisés. Ces moments déterminent de manière unique les images et présentent l'avantage d'être invariants en rotation, translation et changement d'échelle. Les premières versions de ces moments reposaient sur une projection dans un espace à base non orthogonales (possibilité de redondance d'information). Afin d'améliorer ce point, des moments polynomiaux à bases orthogonales ont été proposés dans la littérature (moments de Zernike, de Legendre, Fourier-Mellin, ...[Zernike 38, Teague 80, Belkasim 91, Adam 00b]) et permettent d'obtenir de très bons résultats sur des images de symboles ou de formes [Guillas 07, Choksuriwong 07, Rusiñol 08]. Enfin une dernière approche, Angular Radial Transform [Kim 99, Ricard 05], projette l'image en coordonnées polaires sur des bases sinusoïdales complexes et orthogonales. A partir de ces bases, un ensemble de caractéristiques est extrait, de manière similaire aux moments polynomiaux. Ce dernier descripteur est utilisé dans la norme MPEG-7 pour ses propriétés d'invariance aux bruits, aux changements d'échelle et aux rotations.

Une alternative aux descripteurs géométriques, que nous dénommerons les **transformées unitaires**, propose de décrire les images en projetant les pixels dans un espace. Toutes les méthodes de cette catégorie consistent à transformer l'image pour passer du domaine spatial dans un domaine qui permette d'observer les images sous un autre angle, tout en mettant en avant certaines caractéristiques particulières. Ces transformations peuvent être exprimées de manière générale par la formule suivante :

$$T = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} I(x,y)\, nt_{u,v}(x,y) \tag{1.4.1}$$

où M et N correspondent aux dimensions de l'images, I à la matrice des pixels de l'image, T à la matrice des pixels de l'image dans le nouvel espace, nt au noyau de transformation, et u, v aux vecteurs propres de l'espace de projection. De manière analogue, en utilisant un noyau transformation inverse nti, l'image d'origine peut-être reconstruite à partir de la formule suivante :

$$I = \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} T(u,v)\, nti_{x,y}(u,v) \tag{1.4.2}$$

Les noyaux sont propres à chaque transformée, et regroupent un grand nombre de méthodes très connues (transformée de Hadamard, transformée en cosinus discret, transformée de Fourier discrète, ...). Si l'on prend le cas de la transformée en cosinus discrets (utilisée dans la compression JPEG par exemple), les noyaux correspondent aux différents cosinus utilisés pour décomposer le signal de l'image. Dans le cas de la transformée de Fourier-Mellin, proposée par Adam[Adam 00a], ce sont deux transformations unitaires qui sont enchaînées (une transformée de Mellin appliquée sur la transformée de Fourier

d'une image après avoir passé l'image en coordonnées polaires). Les deux transformées découpent l'espace de l'image pour identifier premièrement les fréquences qui composent l'image, et dans un deuxième temps l'amplitude et la position de ces fréquences. Cette signature a été utilisée avec des symboles de schéma électrique [Guillas 07] ou encore avec des caractères d'imprimerie pour faire de la reconnaissance de caractères [Adam 00a].

Ces transformations ont été étudiées dès la fin des années 1960 [Andrews 71, Gonzalez 09] et des informations détaillées peuvent être obtenues dans [Gonzalez 09]. Toutes ces transformations projettent les images dans des espaces en utilisant un découpage de l'espace « temps-fréquences » régulier. Il devient donc difficile d'étudier un signal avec une bonne résolution spatiale et fréquentielle (étude des basses et des hautes fréquences dans une image).

Pour répondre à ce problème, les théories à base d'ondelettes ont été développées à la fin des années 1990, afin d'obtenir une bonne résolution fréquentielle et spatiale : les ondelettes [Mallat 99] (voir figure 1.4.1 pour une représentation du pavage temps fréquence avec les ondelettes).. Les ondelettes représentent l'image non plus sur des blocs de taille régulière où l'information est analysée identiquement, mais par une décomposition pyramidale où chaque « étage » correspond à la décomposition de l'image à une échelle de détail différente. Cette approche par échelle variable (multi-échelle) offre une approche plus flexible pour le traitement d'images que celle par taille fixe. Une fois les images projetées dans ces espaces, des invariants sont extraits de ces ondelettes pour décrire les images [Bui 99, Shen 99].

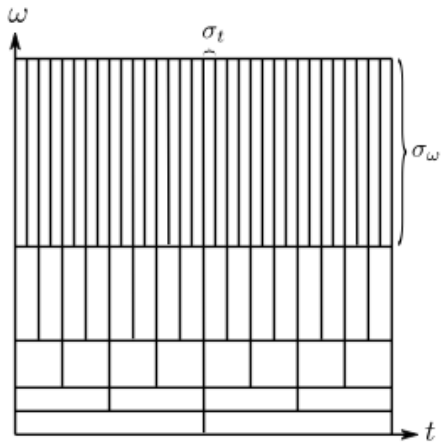


FIGURE 1.4.1: Pavage temps-fréquence obtenu lors d'une transformée en ondelettes. La résolution temps-fréquence dépend de l'échelle d'analyse.

Cas des images en noir et blanc, en niveaux de gris, et des documents anciens

Les méthodes présentées ci-dessus présentent l'avantage d'être applicables sur des images en niveaux de gris ou en noir et blanc. Dans le cas des images de documents anciens, qui étaient généralement créés en noir et blanc et composés de zones de traits, ces méthodes ne sont plus directement applicables. Les dégradations du temps, et les gradients forts propres aux images de documents, ont alors été traités par des méthodes développées pour répondre à ces problèmes spécifiques.

Tout d'abord, [Glauberman 56, Kasturi 90, ÿivind Due Trier 96]ont proposé et utilisé des *histogrammes de projection* pour caractériser le type d'information et sa localisation. L'idée de ce descripteur est de projeter tous les pixels d'une image sur un axe (horizontal ou vertical), et d'en faire la somme des niveaux de gris correspondant à une ligne ou une colonne (voir figure 1.4.2 pour un exemple sur du texte). Ce type d'histogramme permet d'obtenir la courbe moyenne des niveaux de gris d'une image selon une direction privilégiée, et ainsi de discriminer le type d'information (courbe oscillante dans le cas de texte ou de texture ; courbe rectiligne dans le cas de zones uniformes). L'intérêt de ces histogrammes repose également sur le fait qu'ils permettent de localiser l'information dans l'image quand ils sont utilisés conjointement.

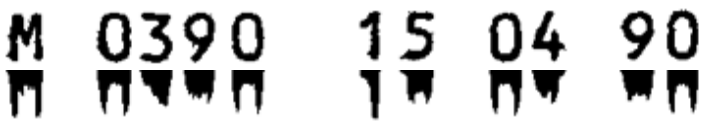


FIGURE 1.4.2: Exemple d'histogramme de projection vertical obtenu sur une image contenant des caractères pour les segmenter

De manière similaire, [ÿivind Due Trier 96, Heutte 98] ont développé des caractéristiques basées sur des profils. Inspirés des travaux de [Shridhar 84], un profil correspond au parcours de l'image à partir d'une zone de départ (à partir du bas pour le profil bas, à partir de la gauche pour le profil gauche, ...). Le profil stocke alors la position du premier pixel noir rencontré tout au long du parcours (la figure 1.4.3 présente des exemples de profils obtenus sur un caractère). Une version dérivée du profil est souvent utilisée pour caractériser la régularité d'une forme ou d'un caractère.

On peut remarquer que les deux premières approches présentées ci-dessus ont cherché à projeter les images sur les bords, sans permettre d'invariance à la rotation. Pour répondre à ce problème, [Ghorbel 94, Tabbone 03, Tabbone 06] ont proposé la R-Signature, qui consiste à utiliser une transformée de Radon pour caractériser les formes. La transformée de Radon consiste à projeter la forme de l'image sur des axes à différentes orientations. La signature correspond alors à la somme des pixels qui interviennent pour une orientation donnée (étalement/compacité de la forme pour un angle donné), le pas d'orientation étant un paramètre de la signature.

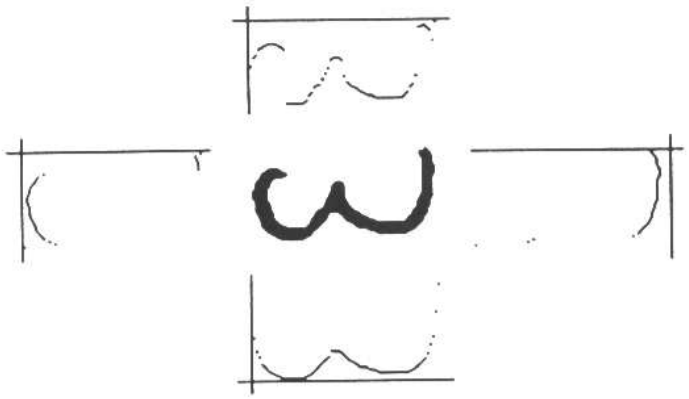


FIGURE 1.4.3: Profils haut, bas, gauche et droite du caractère 3

Bien qu'invariante en rotation, la R-signature ne permet pas de prendre en compte les les occlusions et les trous qui composent les formes. Pour répondre à ce problème, deux méthodes ont été développées dans la littérature, le nombre d'Euler et la F-signature. Tout d'abord, le nombre d'Euler [Pratt 07] permet de dénombrer le nombre de zones noires par rapport au nombre de zones blanches (soustraction réalisée suite au dénombrement des composantes connexes). Cette information statistique ne permet pas de connaître l'influence des différentes parties d'une forme sur les autres, et la dimension des trous. Une signature particulière développée par [Matsakis 99, Wendling 02], la F-Signature, permet de prendre en compte la topologie des trous (voir figure 1.4.4 pour un exemple). Cette méthode calcule un histogramme des forces entre les formes qui composent un objet ou une image, selon une direction particulière. C'est l'ensemble des valeurs obtenues à différentes orientations qui forme la signature d'une image. Cette signature permet de discriminer des formes et de différencier leur compacité et leur densité. Elle présente également un faible coût de calcul et est invariante aux déformations géométriques fondamentales. Enfin, cette signature a été utilisée avec des formes simples , ou des formes complexes (comme des logos ou du texte).

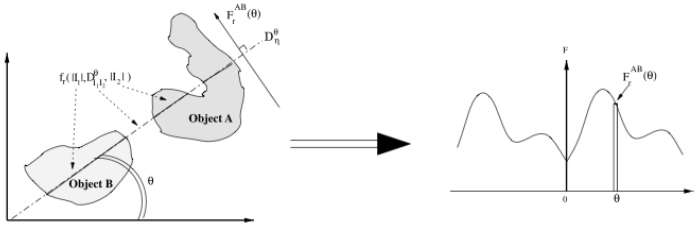


FIGURE 1.4.4: Exemple de calcul de l'histogramme de forces utilisé dans la F-Signature [Wendling 02]

Synthèse des descripteurs de formes

Nous venons de présenter une liste de descripteurs de formes applicables soit aux images en général, soit plus spécifiquement aux images de niveaux de gris ou en noir et blanc. Afin de proposer un comparatif le plus objectif possible de ces méthodes, nous proposons un tableau qui présente l'adéquation (ou non) de la méthode avec différents critères et types d'images. Tout d'abord, nous avons cherché à évaluer la robustesse face aux rotations et changements d'échelles, ainsi que le temps de calcul et la complexité de chaque méthode. Puis, nous présentons également l'adéquation de chaque méthode pour trois types de documents particuliers : les images de scènes naturelles, les images binaires composées de symboles et/ou textes, et enfin les images issues de documents anciens.

On peut constater dans un premier temps que les premières approches présentées, qui cherchaient à parcourir l'image pour y rechercher des motifs identiques, étaient très sensibles aux modifications qui pouvaient affecter une image. A contrario, toutes les méthodes développées par la suite sont invariantes aux altérations classiques. Un petit bémol doit être associé aux histogrammes de projection, ou aux profils qui ont été initialement conçus pour des images composées de texte ou de symboles numérisées dans le même sens et à la même échelle (l'invariance n'était pas forcément recherchée à l'origine).

En ce qui concerne les temps de calcul, toutes ces méthodes sont relativement rapides (à l'exception toujours des deux premières approches). Les deux approches les plus rapides le sont puisqu'elles sont généralement calculées sur un symbole ou un caractère isolé.

Enfin, l'adéquation des méthodes aux images naturelles ou aux images de documents anciens fait apparaître clairement deux catégories de descripteurs. Ces catégories peuvent s'expliquer en partie par le paradigme de segmentation / reconnaissance, puisque pour pouvoir reconnaître efficacement, il faut pouvoir segmenter une information pertinente. Sachant que les descripteurs présentés dans la première partie ont été initialement définis pour décrire des images au sens large, ceux-ci ont été orientés pour segmenter les caractéristiques des images de scènes naturelles. Ainsi, ces descripteurs visent à estimer les variations de niveaux de gris, ou à calculer des statistiques sur les formes présentes dans les images. Les images de scènes naturelles, généralement composées de transitions douces et de formes bien visibles, sont très bien décrites par ces approches. Dans le cas d'images de documents anciens, généralement binaires et composées de forts gradient (trace du crayon ou du tampon utilisée pour l'impression), ces descripteurs deviennent instables puisque les formes se limitent généralement à des contours qui délimitent le fond de la forme, et l'absence de nuances pose des soucis pour l'estimation des variations de niveaux de gris.

Dans cette seconde catégorie, les descripteurs présentés ont plus été orienté pour répondre aux caractéristiques des documents, c'est à dire des images créées par l'homme,

pour l'homme, composées d'éléments présentant une structuration forte. Cette structuration est généralement représentée sous forme de forts gradients, ou par la présence de structures linéaires, et apparaît plus rarement dans les images de scènes naturelles. Ainsi, on peut se rendre compte que ces descripteurs visent plus à étudier la topologie des formes que les variations de gris qui les composent. Ces approches donnent donc de bons résultats sur les documents anciens, puisque les formes correspondent aux éléments en noir (couleur de l'encre), généralement composées de segments (cas des symboles, des traits parallèles qui composent le fond des documents anciens, ...). Dans le cas des images de scènes naturelles, ces outils nécessitent généralement de binariser l'image, et les structures linéaires identifiables par ces caractéristiques sont beaucoup plus rares (surtout après binarisation de l'image). Tous ces éléments sont repris de manière exhaustive dans le tableau 1.1.

1.4.1.2 Approches contours

Toutes les méthodes présentées travaillent directement sur des formes binaires. Afin de simplifier ces formes, et d'améliorer les temps de calcul, des approches ont proposé de travailler sur les contours et le squelette des formes à partir de la fin des années 1970 [Pfaltz 67, Montanari 68, Montanari 69, Fraenkel 69].

La littérature peut être divisée en deux grandes catégories, les méthodes qui s'appliquent directement sur le contour de la forme (ou le squelette) et qui ont été appliquées sur tout type d'image. La seconde concerne les approches qui estiment la forme pour la caractériser, et qui ont été plus spécifiquement appliquées sur des images de documents.

Cas des contours issus des images en général

La première catégorie part du principe qu'un contour est une série de segments qui relient les points d'une forme. Partant de ce contour, des travaux ont cherché à décrire l'orientation des segments qui le composent. Ainsi, Freeman [Freeman 61] a proposé en 1961 de coder l'orientation à suivre pour atteindre le prochain pixel du contour. L'orientation est codée sur 4 ou 8 valeurs, et c'est cette chaîne de valeurs (l'enchainement des orientations discrétisées) qui décrit l'image. Une version relative du codage a également été proposée pour permettre une invariance aux rotations de 45 degrés. De manière similaire, [Takahashi 91, Kimura 91, Cao 95, Park 00] ont proposé d'étudier la fréquence d'apparition de certaines orientations dans les contours qui forment l'image. Cet histogramme est calculé soit sur l'ensemble de l'image, soit sur un découpage en zones de l'image. Enfin, un dernier descripteur plus récent, le Shape context [Belongie 02, Rusiñol 10, Escalera 11] calcule le nombre de points de contours qui apparaissent dans chaque case du quadrillage. La figure 1.4.5 présente un exemple de calcul de shape context à l'aide d'un quadrillage décomposé en 5 cases en profondeur et 12 en rotation. Les valeurs du shape context sont présentées sous forme matricielle dans

Méthode	Invariance en rotation	Invariance aux changements d'échelle	Temps de calcul / Complexité	Adaptée aux images naturelles	Adaptée aux images de binaires symboliques	Adaptée aux images de documents anciens
Images						
Image elle-même	- -	- -	- -	+	- -	+
Template Matching	- -	- -	- -	+	+ +	-
Moments géométriques et polynomiaux	+ +	+ +	+	+ +	+	-
Transformées unitaires (Fourier, Fourier-Mellin, ...)	+ +	+ +	+	+ +	+ +	+
Invariants issus d'ondelettes	+ +	+ +	+	+ +	- -	-
Images en Noir et Blanc ou en niveaux de gris uniquement						
Histogramme de projection	+ +	- -	+ +	- -	+ +	+ +
Profils	- -	+	+ +	- -	+ +	+ +
R-Signature	+ +	+ +	-	- -	+	+ +
Nombre d'Euler	+ +	+ +	+	- -	- / +	+
F-Signature	+ +	+ +	-	-	-	+

TABLE 1.1: Comparatif de différentes méthodes de caractérisation de formes

l'exemple, plus une case est noire, plus sa valeur est importante (beaucoup de points dans cette région). On peut remarquer que les deux premiers shape context sont très similaires, tandis que le dernier est très différent. Dans ce cas précis, le quadrillage est obtenu après avoir projeté l'image dans un repère polaire. Ceci rend la méthode invariante en rotation.

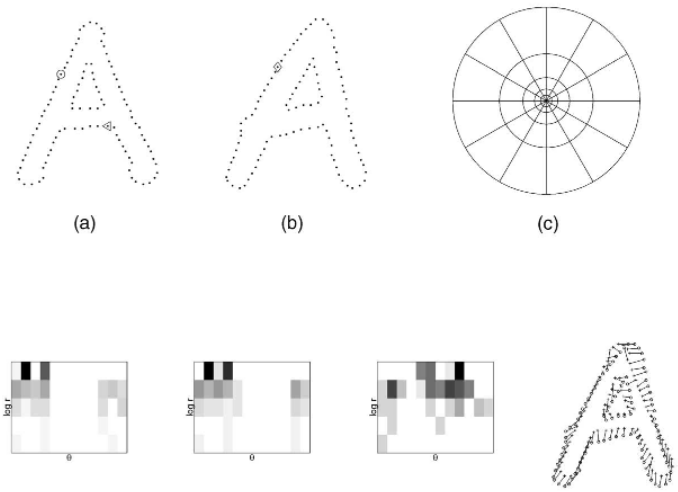


FIGURE 1.4.5: Calcul et recherche de correspondance du Shape context. a) et b) Exemples de contours discrétisés de deux formes. c) Quadrillage log-polaire utilisé pour dénombrer les points appartenant à chaque cadran. d), e), and f) Exemples de shape context, sous forme matricielle, obtenus avec les images a) et b) en partant des points de références indiqués sur les images. Enfin, e) présente la mise en correspondance des deux images obtenues à partir de l'approche présentée dans [Belongie 02]

Une seconde série de travaux a cherché à estimer l'allure générale du contour en le projetant dans des espaces particuliers. Ainsi, on peut citer les travaux de [Shridhar 84, Taxt 90] qui calculent les coefficients de la décomposition en série de Fourier du contour, qui est vu comme un signal en une dimension. Une version généralisée, et invariante en rotation, le Generic Fourier Descriptor (GFD), a été proposé par Zhang [Zhang 02]. Dans ce cas, l'image est d'abord transposée en coordonnés polaires avant le calcul de la transformée de Fourier. Enfin, une dernière approche, *Curvature Scale Space* (CSS), développée au début des années 90 [Mokhtarian 92], propose de créer une image des courbures à différentes échelles (voir figure 1.4.6 pour des exemples). Cette image permet de renseigner sur les courbures rencontrées et la longueur des plages entre deux points d'inflexion. La description d'une image correspondra à la recherche des maxima dans l'espace des courbures à différentes échelles.

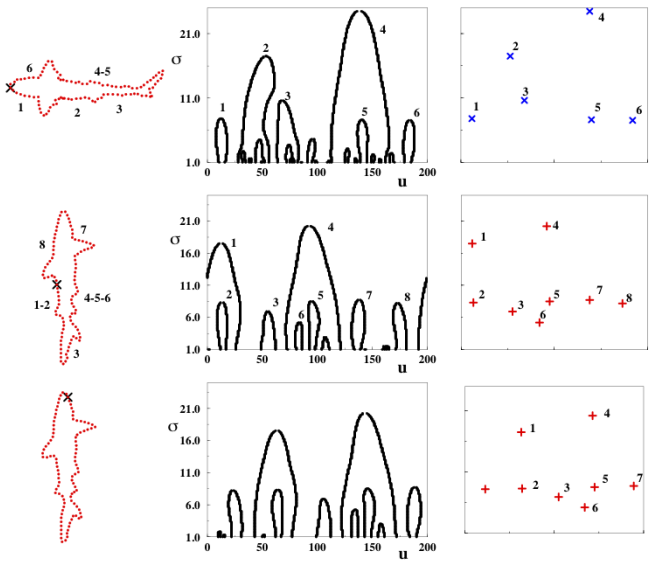


FIGURE 1.4.6: Exemples d'images CSS obtenues avec leur maxima sur trois formes similaires en utilisant la méthode présentée dans [Mokhtarian 96]. A gauche, contour de l'image d'origine avec l'origine de l'algorithme indiqué par une croix noire. Au milieu, courbe CSS obtenue, et à droite graphique des maxima des courbes. Ce sont ces maxima qui sont utilisés pour mettre deux images en correspondance

Cas des contours issus des images de documents

Les méthodes présentées dans cette première catégorie permettent de décrire statistiquement le contour d'une forme, en estimant l'allure générale de celle-ci. Elles sont donc applicables de manière générale à tous les types d'images. Cependant, bien que ces méthodes donnent des résultats intéressants, elles ne permettent pas d'analyser intuitivement les résultats (interpréter des rayons de courbures, des orientations, ou des coefficients de Fourier n'est pas toujours aisé). Une seconde catégorie d'approches a donc été créée dans la littérature pour proposer une approximation des contours non plus à l'aide de points mais d'éléments structurants de plus haut niveau (droites, segments, polygones, ...). Cette seconde catégorie a été plus largement utilisée sur des images de documents (documents techniques comportant des symboles électriques par exemple, ou sur des images graphiques de documents anciens). L'utilisation d'éléments structurants plus complexes permet l'utilisation de signatures structurelles qui renseignent sur la position relative des éléments. Cependant, l'utilisation d'éléments structurants de plus haut niveau impose de définir des hypothèses sur les formes à décrire (par exemple, dans [Coustaty 11], les auteurs supposent que les symboles sont approximables par un ensemble de segments), et ne garantit en rien la généricité de l'approche sur des bases hétérogènes.

L'utilisation d'éléments structurants de plus haut niveau impose tout d'abord de les extraire. On peut ainsi citer les approches qui cherchent à extraire des droites ou des segments avec une régression linéaire, ou par des méthodes plus complexes comme la transformée de Hough [Duda 72] (initialement conçue pour extraire des droites, puis modifiée pour l'extraction de segments [Coustaty 11] - un exemple peut être observé dans la figure 1.4.7).

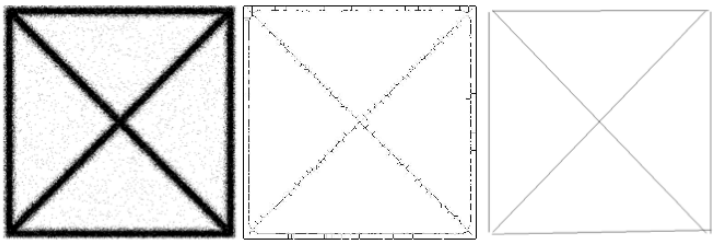


FIGURE 1.4.7: Exemple de symbole utilisé dans [Coustaty 11]. De gauche à droite sont présentés, d'abord le symbole original, puis au milieu l'image obtenue à l'aide d'une squelettisation, et enfin à droite le résultat obtenu à l'aide d'une transformée de Hough adaptée

De nombreux travaux ont également proposé d'extraire une version grossière de la forme, en la délimitant par une boîte englobante. Cette boîte peut-être de forme rectangulaire, elliptique, ou estimée par des polygones [3]. Un cas particulier de forme englobante des contours, le polygone isothétique [Pal 10] (voir figure 1.4.8 pour un exemple), permet de classer des types d'objets graphiques dans des documents. Un polygone isothétique correspond à un polygone construit uniquement à partir de rectangles horizontaux ou verticaux (impossible de rejoindre deux pixels en diagonale). Estimer l'image à l'aide de polygones de ce type donne une impression de crénelage, ce qui permet de simplifier la forme. En utilisant des caractéristiques basées sur la surface, ou à partir du périmètre total de la forme divisé par le nombre de polygones utilisés, Pal *et al* [Pal 10] arrivent à distinguer plusieurs catégories d'images graphiques.

Une fois les éléments structurants extraits, la signature structurelle revient à décrire l'agencement entre ces éléments, ou à mesurer la complexité et les occlusions de la forme et de son contour (plus une forme sera complexe et trouée, plus le nombre de polygones sera important, et les ratios faibles). Bien que l'extraction de caractéristiques soit facilitée par cette approche, elle est difficilement applicable aux images du début de l'imprimerie qui sont composées de nombreux traits. De plus, cette approche permet de distinguer du texte des objets graphiques mais ne permet pas directement la comparaison d'objets graphiques.

3. Dans le cas particulier de l'ellipse englobante, les axes de l'ellipse sont obtenus à partir des moments géométriques d'ordre 2.

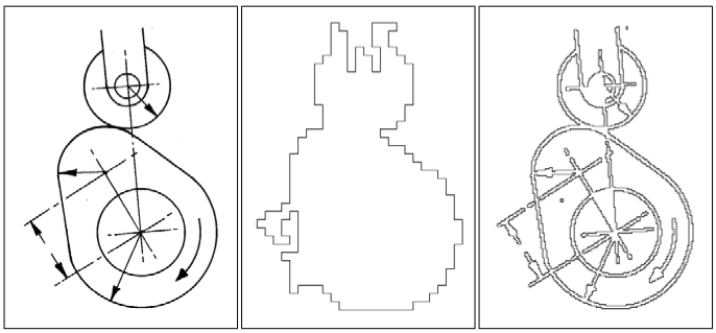


FIGURE 1.4.8: Exemple de polygonisation obtenue à l'aide de l'algorithme GOAL. L'image d'origine (à gauche) est simplifiée à l'aide de polygones isothétiques (1ère étape de simplification au milieu, résultat final à droite)

Synthèse des approches contours

Afin de comparer les méthodes présentées pour extraire et caractériser des contours, nous avons cherché à les comparer sur la base de cinq critères que nous avons voulu les plus objectifs possibles. Ces critères cherchent à évaluer la robustesse de ces approches face aux changements d'orientation ou d'échelle. Nous avons également pris en compte leur temps de calcul et leur complexité. Enfin, les deux derniers critères retenus prennent en compte le nombre de paramètres et la complexité d'adaptation de ces paramètres. Nous avons cherché à savoir, avec le critère de complexité d'adaptation des paramètres, si une approche peut être transposée à différents problèmes et types d'images sans avoir à modifier tous les paramètres.

Tout d'abord, on peut constater que les approches basées sur les contours qui ont été présentées sont toutes robustes face aux rotations et changements d'échelle (à l'exception du codage de Freeman qui n'est invariant en changement d'échelle qu'à la condition de connaître la différence d'échelle entre les deux formes). De même, on peut constater que toutes ces approches sont plutôt rapides à calculer de par la simplification des images en contours ou squelettes (le nombre de points à traiter est grandement réduit). Il faut cependant tempérer cet argument par la nécessité d'extraire le contour et le squelette de l'image avant de pouvoir la traiter. Enfin, si l'on observe les critères de complexité des méthodes (nombre de paramètres et facilité d'adaptation aux cas d'utilisation), on peut remarquer deux tendances. Tout d'abord, les approches qui cherchent à parcourir toute la chaîne de pixels nécessitent généralement plus de paramètres (nombre d'orientations à observer, nombre d'échelles, ...) et il est préférable d'adapter ces paramètres à chaque type d'image pour obtenir de meilleurs résultats. A contrario, les approches qui décomposent et estiment les contours pour en proposer une approximation sont généralement plus faciles à paramétrer (peu de paramètres ou fixés par défaut lors de la création de la méthode par l'auteur ; les paramètres par défaut conviennent à la grande majorité des images). Le tableau 1.2 présente les méthodes et les caractéristiques de manière

synthétique.

1.4.1.3 Représentation des textures

Sur Wikipedia, une texture est définie comme : « une région dans une image numérique qui a des caractéristiques homogènes répétitives ». Ces caractéristiques homogènes peuvent représenter un motif qui se répète, spatialement ou fréquentiellement. Plusieurs méthodes de la littérature cherchent à caractériser les textures, et certaines ont été appliquées aux images en général, tandis que d'autres ont été appliquées plus spécifiquement aux images de documents.

Cas des images en général

La première catégorie de méthode développée pour les images rassemble les méthodes à base de noyaux. L'idée générale de ces approches est qu'une texture peut-être estimée par un modèle stochastique. Ces modèles émettent l'hypothèse que la valeur de niveau de gris associée à chaque pixel ne dépend que de son voisinage et cherchent à estimer cette valeur en fonction de son voisinage. Les paramètres utilisés pour adapter le modèle à l'image sont caractéristiques de chaque texture, et sont utilisés pour identifier chaque image. Des exemples d'applications peuvent être trouvés dans [Besag 74, Cross 83, Derin 87, Komodakis 11].

Une seconde catégorie de méthodes, appartenant aux techniques à base de descripteurs, qui existent en grand nombre (attributs stochastiques, attributs issus de la matrice de co-occurrences, intégrales curvilignes, matrices des longueurs de plage, ...). L'idée de toutes ces méthodes est de résumer le contenu de l'image à l'aide de statistiques sur les pixels qui la composent. La plus connue d'entre elles est la matrice de co-occurences d'Haralick et de Laws [Haralick 73] (Également connue sous le nom de *GLCM* pour *G*ray *L*evel *C*o-occurrence *M*atrix). Très largement utilisée dans l'analyse de textures pour sa facilité de mise en oeuvre et ses bons résultats, cette matrice correspond au nombre d'occurrences d'une transition de niveau de gris entre pixels (transition sur une distance et une orientation données en paramètres). La réussite de la méthode repose sur le choix du vecteur de déplacement et plusieurs études [Karathanassi 00, Iften 04] ont montré qu'en pratique une distance courte permet d'obtenir des résultats optimaux. Comme précisé par Journet [Journet 06], les matrices de co-occurrences sont généralement très grandes et ne permettent pas une caractérisation aisée des textures. Elles sont donc généralement résumées à l'aide de quatorze indices définis par [Haralick 73]. Ces indices ont l'avantage de permettre une meilleure discrimination des images. Ils peuvent être rassemblés en trois familles de mesures : contraste, énergie et statistique. L'énergie permet de connaître la régularité de la texture et les mesures statistiques renseignent sur la répétitivité de la texture.

Méthode	Invariance en rotation	Invariance aux changements d'échelle	Temps de calcul / Complexité	Nombre de paramètres	Généricité
Codage de l'orientation des éléments du contour					
Freeman	+	-	+ +	-	+
Histogramme des orientations	+ +	+	+	-	+
Shape context	+ +	+ +	+	- -	-
Décomposition et estimation du contour					
Basées sur une transformée de Fourier (GFD, FM)	+ +	+ +	-	+	+
CSS	+ +	+	+	+ +	+ +
GOAL	+ +	+ +	N/A	+	+

TABLE 1.2: Comparatif de différentes méthodes de la littérature pour l'analyse et la description des contours

Une catégorie de méthodes, utilisables sur les images en général et spécifiquement sur les images de documents, rassemble les méthodes qui reposent sur des filtres particuliers. Ces filtres sont utilisés pour ne sélectionner qu'un certain type de textures, et les coefficients de ces filtres correspondent à des motifs fréquentiels particuliers que l'utilisateur souhaite faire ressortir dans les images. On peut ainsi citer les filtres de Fourier, ou de Gabor [Hsu 93, Pham 07] qui permettent de ne faire ressortir que certaines textures (un exemple de filtre de Gabor peut-être observé dans la figure 1.4.9). L'utilisation de ces filtres est réalisée au travers de banc de filtres, qui permettent une sélection large de l'information à retenir (il est rare qu'une image soit composée d'une texture unique). Cependant, bien que l'information extraite soit pertinente, elle impose à l'utilisateur d'avoir une connaissance sur les textures recherchées pour adapter les paramètres des différents filtres.



FIGURE 1.4.9: Exemple de construction d'un filtre de Gabor, avec la représentation imagée d'une gaussienne à gauche, d'un sinus au centre, et enfin du filtre obtenu à droite

Cas des images de documents anciens

Les approches textures ont beaucoup été utilisées, dans le monde des documents, pour identifier le type d'information qui compose une zone. En effet, le texte est généralement considéré comme une zone de texture, tandis que les autres objets représentés sont vus comme des formes. Une première catégorie rassemble les méthodes de corrélation et d'auto-corrélation. Couramment utilisées dans l'étude des textures [Lin 97, C. Rosenberger 01, Uttama 05, Journet 08c], elles permettent de connaître la similitude entre deux textures en réalisant une convolution entre les deux images. Cette convolution est généralement représentée sous la forme d'une courbe ou d'une fonction associée à la fonction d'auto-corrélation. Dans l'exemple présenté dans [C. Rosenberger 01], la finesse de la texture est représentée par une pente (plus la texture sera fine, plus la pente de la fonction sera importante). Cependant, cette convolution impose une connaissance a priori sur la (ou les) texture(s) recherchée(s), puisque le déplacement élémentaire du filtre d'autocorrélation est défini par la taille de la primitive texturelle. Pour être réellement efficace, il faut donc apprendre la texture à rechercher, et cela nécessite un pré-traitement pour être utilisable sur des documents composés de textures différentes (du point de vue fréquentiel). Enfin, comparer deux images consiste à convoluer la première image par la seconde. Dans le cas d'images de grande taille (supérieures à quelques centaines de

pixels), les temps de calculs deviennent long et cette méthode n'est alors plus envisageable.
Afin de pallier ce problème, et de permettre une caractérisation d'ensemble d'un document, Journet *et al.* [Journet 08c]ont proposé une *rose des vents* (également connue sous le nom de rose des directions), basée sur l'autocorrélation. Cette méthode permet de connaître l'orientation principale d'un document ou d'un symbole, en observant l'orientation principale de la rose (voir figure 1.4.10 pour des exemples de roses), mais également de discriminer le type d'information contenue dans un document (texte, zone graphique, image, ...) à l'aide de trois indices (angle de l'orientation principale, isotropie de l'image et orientations caractéristiques de la forme de la rose). Les auteurs ont également ajouté deux indices textures liés aux fréquences présentes dans l'image pour améliorer la caractérisation du contenu.

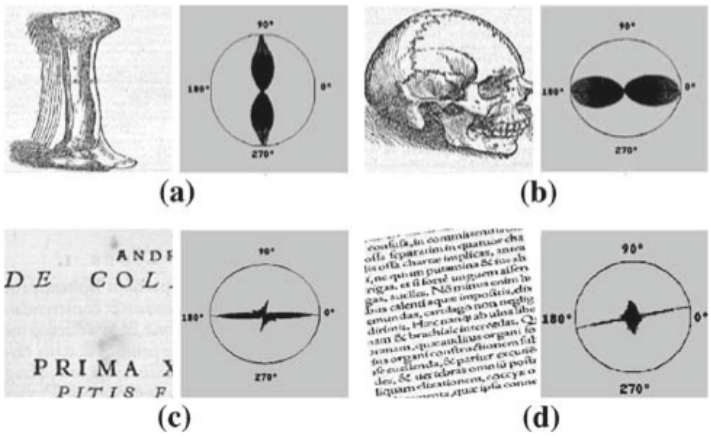


FIGURE 1.4.10: Exemple de roses des vents calculées sur différentes images. a) objet graphique vertical, b) partie d'un squelette, c) quelques lignes de textes (titre), d) texte écrit de manière penchée

Enfin, une dernière catégorie d'approche rassemble toutes celles qui consistent à segmenter l'image en zones homogènes (du point de vue de la texture). Parmi les méthodes les plus célèbres dans cette catégorie, on peut citer les algorithmes RLSA [Wong 82] et XY-CUT [Journet 06]. On peut également citer le diagramme de Voronoï (également appelé « partition de Voronoï » ou « pavage de Voronoï ») qui a été largement étudié dans le domaine de la géométrie et appliqué dans différentes disciplines [Aurenhammer 91]. Le diagramme de Voronoï [Fortune 86] consiste, en partant de germes [4] dans l'image, à générer un partitionnement en polygones, de telle sorte que les germes de deux sites voisins soient équidistants de la frontière (c'est à dire l'arête commune). Chaque polygone généré s'appelle un site du diagramme. Un exemple de diagramme de Voronoï peut être observé dans la figure 1.4.11, et un exemple interactif de construction du diagramme

4. Les germes sont généralement issus de tirages aléatoires

est visible sur [for 11]. De nombreuses méthodes de génération du pavage de Voronoi existent dans la littérature [Tuceryan 90] mais toutes nécessitent de définir les critères d'évaluation et la métrique entre des zones. A partir de ce pavage, des caractéristiques sont calculées pour décrire le contenu de chaque groupe. L'agencement et les propriétés des clusters permettent la reconnaissance et l'indexation d'images.

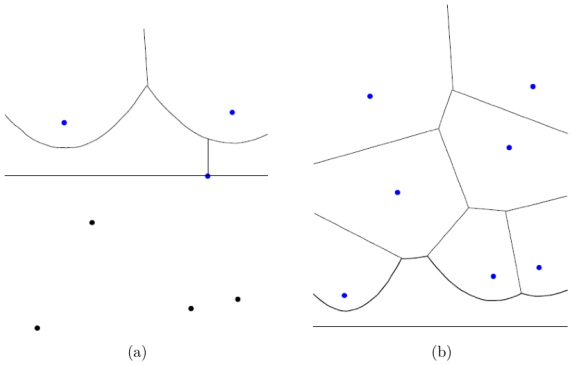


FIGURE 1.4.11: Exemple de construction du pavage de Voronoï à l'aide d'un algorithme par balayage : l'algorithme de Fortune [Fortune 86]

Synthèse des approches textures

Le tableau 1.3 présente sous forme synthétique les points forts et points faibles des méthodes présentées ci-dessus. Les méthodes orientées textures sont comparées sur la base de six critères. Tout d'abord, nous avons pris en compte, comme pour les méthodes orientées formes et contours, le temps de calcul et la complexité de ces méthodes. Le second paramètre donne une indication sur le nombre de paramètres utilisés pour chaque méthode. Le troisième critère est associé au second, puisqu'il indique la généricité de l'approche. C'est à dire s'il est aisé ou non de modifier les paramètres de chaque approche pour l'adapter au cas d'application. Enfin, les deux derniers critères permettent de connaître l'adéquation de la méthode avec les images de scènes naturelles, et les images de documents.

En ce qui concerne le temps de calcul et la complexité des approches présentées, on peut remarquer qu'aucune de ces approches n'apporte de solution réellement rapide au problème de segmentation et de description des images. Toutes ces méthodes imposent de parcourir l'image et de calculer soit pour chaque pixel, soit pour chaque zone, une valeur dépendante du voisinage.

Du point de vue du nombre de paramètres et de la généricité, la plupart des approches imposent une connaissance sur le type d'images que l'on souhaite analyser afin de régler les paramètres manuellement. Il faut noter le cas particulier des méthodes basées sur des noyaux stochastiques pour lesquelles aucun paramètre ne doit être fixé par l'utilisateur. Les paramètres sont estimés à partir du voisinage de chaque pixel. Par contre, de par la

nature discontinue des images de documents anciens (images binaires avec de grandes variations de niveaux de gris), ces méthodes ne donnent pas de résultats intéressants. En effet, la plupart des approches stochastiques sont applicables avec une hypothèse de continuité de la fonction étudiée. Dans le cas des images de documents anciens, la fonction est une matrice discrète, composée de grandes variations de niveaux de gris fréquentes (puisque l'information est généralement noire sur du papier blanc). De plus, bien qu'il soit possible de pré-traiter l'image, il devient difficile d'estimer la valeur d'un pixel lorsque son voisinage a subi les effets du temps (jaunissement non uniforme du papier par exemple).

1.4.1.4 Représentation des couleurs

Présentation des descripteurs

De par la conception du système visuel humain, la couleur joue un rôle important dans la reconnaissance des objets et des formes qui nous entourent. En effet, les informations perçues par la rétine sont décodées par notre cerveau pour nous permettre de percevoir les objets qui nous entourent. La vision humaine repose sur deux types principaux de capteurs, des bâtonnets et des cônes. Les premiers nous permettent de capter la luminosité qui nous entoure, tandis que les seconds nous permettent d'avoir une vision fine et en couleur. Seuls les objets qui se trouvent dans la zone centrale de la rétine, où la vision des détails est la plus précise (appelée fovéa - voir figure 1.4.12), peuvent être reconnus par notre cerveau. Or, cette zone est uniquement peuplée de cônes [5]. Ces cônes permettent une vision précise et en couleur. Afin de mimer la vision humaine, plusieurs approches ont été proposées pour représenter la couleur et décrire les images. Tout d'abord, plusieurs espaces colorimétriques ont été définis pour représenter les images et la couleur dans les ordinateurs. Ainsi, l'espace RGB est classiquement utilisé pour représenter la couleur de chaque pixel. Cependant, les distances entre deux points voisins représentent mal les différences de perception par le cerveau humain. Afin d'apporter des réponses à ce problème, plusieurs espaces colorimétriques (La*b*, YUV, ...) ont été proposé pour évaluer des distances entre points similaires aux phénomènes perceptifs.

Afin de représenter les couleurs qui composent les objets d'une image, plusieurs méthodes ont été utilisées dans la littérature, voici les plus connues :

– **Les histogrammes couleurs** [Swain 90, Ferman 02, Han 02] : Méthode la plus classique, elle consiste à considérer les trois canaux de couleur (R, G, B) indépendants et à représenter leur distribution. Cette distribution est obtenue par discrétisation et par comptage du nombre de pixels appartenant à chaque niveau de couleur. Les histogrammes présentent l'avantage d'être invariant aux translations et aux rotations. De plus, ils présentent une certaine robustesse face aux changements d'angle de vue, d'échelle et d'occlusion. Leur principal inconvénient repose

5. Pour plus de détails sur les bâtonnets (vision en niveaux de gris), les cônes (vision en couleur), et les différentes zones de vision de l'oeil, voir [Purves 05], chapitre 10

Méthode	Temps de calcul / Compléxité	Nombre de paramètres	Généricité	Adaptation aux images naturelles	Adaptation aux images de documents
Méthode à base de noyaux	- -	+ +	-	+ +	- -
Filtres particuliers	-	- -	- -	+ +	+
Méthodes basées sur les descripteurs	-	- -	+	+ +	+
Corrélation Auto-corrélation	- -	-	- -	+	+
Découpage de l'image	-	-	-	+	+ +

TABLE 1.3: Comparatif entre les 5 grandes catégories de méthodes d'extraction et de description de textures

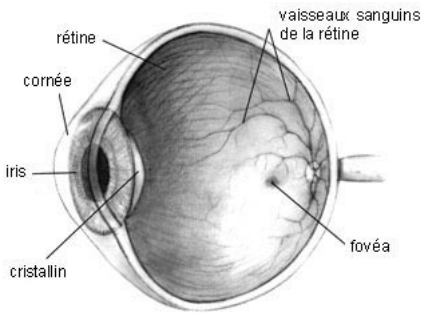


FIGURE 1.4.12: Système visuel humain - Image extraite de Wikipedia

sur l'absence d'information spatiale (deux images différentes peuvent être décrites par les mêmes histogrammes) ;

- **Les noms de couleurs par quantification** [Heckbert 82, Puzicha 00, Yang 07] : Tout comme nous le faisons en langage naturel, ces approches associent des noms aux couleurs à l'aide d'algorithmes de quantification couleur. Les noms des couleurs sont issus d'un dictionnaire, et un degré de confiance est associé à chaque couleur. La similarité entre les images repose alors sur la recherche d'images composées des mêmes noms de couleurs et degré de confiance. Tout comme pour les histogrammes couleurs, ces méthodes ne permettent pas de représenter l'information spatiale des régions couleurs ;
- **Les moments couleurs** [Stricker 95, Jau-Ling 02] : Initialement conçus par Stricker et Orengo, cette représentation couleur consiste à calculer les histogrammes des trois canaux couleurs. Pour chacun de ces histogrammes, les trois premiers moments géométriques centrés sont calculés. Le premier moment permet de définir la couleur moyenne de chaque canal, alors que le deuxième et le troisième moments permettent de représenter la variance et la dissymétrie de chacun des canaux. Une fois ces descripteurs obtenus, une fonction de similarité calcule une distance euclidienne pondérée pour identifier la similarité de deux images.

Toutes les méthodes citées précédemment souffrent d'un problème commun : bien qu'elles résument globalement le contenu des images, elles ne permettent pas d'encoder l'information spatiale et de répondre aux questions : où et comment sont reliés les pixels ? C'est ce que nous allons voir dans la section suivante.

1.4.2 Approches par description spatiale

Bien qu'elles permettent de résumer globalement les images, les approches globales ne permettent pas d'identifier la position de l'information (perte des repères spatiaux). Afin de pallier ce problème, de nouvelles signatures, souvent connues sous le nom de signatures structurelles ou topologiques, proposent des réponses à ce problème. Le principe de toutes ces méthodes repose sur trois étapes fondamentales :

Extraction de primitives

Ces primitives peuvent être des pixels, des segments, des zones homogènes, des régions, ... ([Eglin 99, Uttama 05, Journet 08b]). L'extraction de primitives a été appliquée sur des images naturelles (segmentations en régions homogènes), mais également sur des images de documents anciens. Dans ce cas, les objets présentent une structure forte, structure inhérente à leur construction. Par exemple, dans [Coustaty 08] et [Rusinol 07], le but est de comparer des symboles qui sont naturellement très structurés (composés de segments). Les deux approches proposent soit d'extraire les segments avec une transformée de Hough adaptée, soit d'extraire les polygones formés par les segments.
Cependant, d'autres approches ont proposé de décrire l'organisation spatiale de zones particulières d'une image (position du ciel par rapport à la mer, position des yeux dans un visage, ...). Les primitives sont alors extraites soit suite à une division de l'image en sous-blocs de taille fixe (quadrillage appliqué sur l'image), soit selon des critères d'homogénéité (principe identique à celui utilisé dans les algorithmes de sélection de régions uniformes ou texturées vu précédemment).

Codage de l'agencement entre ces primitives

L'idée générale des signatures structurelles et spatiales est de décrire l'agencement et la position des formes qui composent un objet et/ou une image au sens large. Une fois les zones d'intérêt extraites, plusieurs solutions ont été proposées pour coder les relations qui existent entre ces différentes primitives élémentaires. Le codage des relations spatiales a fait l'objet de nombreuses études, en particulier dans le cadre des informations géo-référencées, mais également en indexation et reconnaissance des formes. Ainsi, des approches proposées dans [Allen 83, Randell 92] cherchent à caractériser la position spatiale d'une région avec une autre. Ainsi, elles permettent de connaître la position relative d'une région par rapport à une autre (au-dessus, à gauche, ...), ou si les régions possèdent des zones en commun (superposition, inclusion, ...). Des approches se sont servi de la topologie implicite des graphes pour représenter les formes qui composent une image. Ainsi, les graphes d'adjacence des régions [Felzenszwalb 04, Bodic 09] décrivent l'agencement des régions qui composent une image (les noeuds correspondent aux régions et les arcs aux liens entre les régions). Enfin, des approches plus adaptées aux documents [Etemadi 91, Dosch 04, Rusiñol 06, Mas 06, Coustaty 08, Mas 10] ont proposé d'analyser la structure de primitives qui composent un symbole ou un objet. Cette structure est représentée à l'aide de grammaires 2D, ou d'identifiants propres à la relation qui unit deux éléments structurants (identification du lien existant entre deux polygones ou segments par des relations).

Représentation du codage

Enfin, une fois la liste des relations calculées sur l'ensemble des primitives élémentaires, toutes ces relations sont représentées sous forme d'arbres, de grammaires [Rusiñol 06,

Mas 06] ou de graphes [**?**]. L'utilisation de structures permet généralement de retrouver intuitivement la forme de départ, et facilite la comparaison entre deux images. Ainsi, la reconnaissance ou l'indexation se fait par comparaison de ces structures (similarités de structure, isomorphismes entre graphes, ou dérivations de grammaire).

1.4.3 Approches par description locale

Décrire une image globalement présente l'avantage de conserver un maximum d'information sur son contexte mais ne permet de description en détails. Pour pallier ce problème, plusieurs approches ont été développées pour identifier des points d'intérêt dans une image (zones visuellement importantes) et décrire leurs alentours. C'est le but des approches locales (points d'intérêt) ou semi-locales (zones d'intérêt). L'idée revient à utiliser les approches globales de manière locale sur une sous-partie de l'image.

Leur bon fonctionnement repose bien évidemment sur l'extraction et la sélection de ces sous-parties pour qu'elles soient les plus pertinentes possible dans la description des images. Ces zones peuvent, en général, être caractérisées par :

- une définition claire, avec des fondements mathématiques de préférence ;
- une position bien déterminée dans l'espace image ;
- la structure locale de l'image autour du point d'intérêt et riche en terme d'information sur son contenu, de telle sorte que l'utilisation de ce point réduit la complexité des traitements suivants dans le système de vision ;
- une stabilité aux perturbations locales et globales de l'image, incluant les déformations topologiques et de perspectives (transformations affines, changements d'échelles, rotations, translations), comme les changements d'illuminations et de contrastes, de telle sorte que les points d'intérêt possèdent une forte reproductibilité ;
- une intégration optionnelle d'une information sur l'échelle, pour pouvoir extraire des points d'intérêt d'images à différentes échelles et résolutions.

Les détecteurs de points d'intérêt appartiennent à trois catégories principales :

1. détecteurs de contours : parmi lesquels on peut citer le détecteur de contour de Canny [Canny 86] et Canny-Deriche [Deriche 87], les filtres différentiels, les opérateurs de Sobel, Prewitt, Roberts et Cross ;
2. détecteurs de coins : les détecteurs de Moravec [Moravec 79] et de Harris [Harris 88a] (et leurs différentes améliorations pour les rendre robustes aux changements d'échelles [Lindeberg 98]), aux transformations affines [Shi 94], invariant aux rotations [Harris 88b] qui extraient les coins à partir des dérivées des images. On peut également citer deux approches rapides [Smith 97, Rosten 10] qui déterminent les coins à partir des variations de niveaux de gris de l'image autour d'un noyau ;
3. détecteurs de régions d'intérêt : les régions d'intérêt permettent d'extraire des régions, et leur agencement, contrairement aux détecteurs de coins. Cependant, de nombreux détecteurs de régions d'intérêt associent un point maximal aux régions (généralement le centre de gravité) et peuvent être utilisés comme détecteurs de

coins. Les approches les plus répandues reposent sur des différences de gaussiennes à différentes échelles [Marr 80], le Laplacian de Gaussiennes [Lindeberg 94], le déterminant de la matrice Hessienne (utilisée dans l'opérateur SURF [Bay 06]) ou sur l'extraction des régions maximales stables [Matas 04].

D'autres détecteurs spécifiques ont également été mis en place pour extraire des éléments particuliers. C'est le cas des travaux présentés par Rusinol dans [Rusinol 07] dans lesquels l'auteur extrait les différents symboles de plans techniques pour les reconnaître. Il applique alors un traitement semi-local puisqu'il calcule les moments de Hu sur chaque polygone. Il utilise l'ambiguïté introduite par les moments de Hu, associée à un système de vote pour déterminer les images à retenir.

Une autre approche sur les méthodes semi-locales est abordée dans [Kauniskangas 99]. Les moments de Zernike et de Fourier-Mellin sont utilisés sur un voisinage des points d'intérêt détectés avec Harris. Il vient également comparer ces deux techniques aux descripteurs SIFT (Scale-Invariant Features Transform). Ces descripteurs sont largement utilisés dans la littérature car robuste et permet d'obtenir les meilleurs taux de reconnaissance dans de nombreux cas d'utilisation (voir [Lowe 99, Choksuriwong 07]). Les approches semi-locales offrent de bien meilleurs résultats par rapport aux approches globales puisqu'elles permettent de décrire ce qui est important dans l'image de manière plus fine. On ne cherche pas à décrire toute l'image mais seulement ce qui caractérise l'image et la discrimine des autres.

1.5 Les projets déjà existants

De nombreux projets ont été développés dans la littérature et utilisés dans des cadres très différents au cours des 15 dernières années. Tous proposent de rechercher des images par leur contenu, et d'appliquer cette technique à différents cas d'application. Nous proposons ci-après un tableau récapitulatif des différents projets par domaine d'application. Nous présentons également un bref résumé de leur fonctionnement [6].

6. Le fonctionnement de ces moteurs est bien plus complexe que ce qui est présenté dans ce tableau, mais nous avons volontairement raccourci leurs descriptions pour ne pas noyer le lecteur dans des explications trop complexes

Méthode	Domaine d'application	Fonctionnement
	Recherche dans des vidéos	
Jacob	Vidéos de paysages et de scènes naturelles	Des caractéristiques couleur (histogramme moyen des trois canaux RGB) et texture (descripteurs issus de la matrice de co-occurrence) sont utilisées pour décrire les images et les comparer entre-elles
[Dubois 10]	Recherche de textures dynamiques similaires dans la base DynTex	Les images sont décrites à partir de caractéristiques textures. Celles-ci correspondent aux coefficients d'ondelettes obtenus à l'aide de différentes approches (curvelet, MCA, ...)
	Recherche d'images de scènes naturelles	
BDLP	Images de plantes, animaux, personnes et paysages de scènes naturelles	Les images sont tout d'abord réduites à 13 couleurs par quantification. Puis, chaque couleur est décrite par un vecteur de six caractéristiques (surface de l'image contenant cette couleur, et nombre de fois que l'on trouve des formes 'très petites', 'petites', 'moyennes', 'grandes', 'très grandes' dans l'image)
Blobworld	Images de scènes naturelles en couleurs	Ce moteur repose sur la combinaison de caractéristiques de couleur (histogramme de 218 couleurs), de texture (contraste moyen et anisotropie), et de forme (aire, excentricité, et orientation)

Méthode	Domaine d'application	Fonctionnement
C-bird	Images de scènes naturelles	Pour chaque image, ce moteur décrit l'image à l'aide de quatre vecteurs (un vecteur couleur, un vecteur des couleurs les plus fréquentes, un vecteur des orientations les plus fréquentes, et un vecteur chromatique)
Chabot	Images de scènes naturelles	Combine de la recherche par mots-clés associés aux images (manuellement) et d'un histogramme de 20 couleurs
FOCUS	Publicités et images naturelles couleurs	Chaque image est décrite à l'aide d'un histogramme couleur. Cet histogramme est obtenu en découpant l'image sous la forme d'une matrice 100x100, et quantifiant les couleurs dans chaque case
NETRA	Base COREL	Les images de la base sont segmentées en régions de couleur homogène. Pour chaque région, un vecteur de caractéristiques couleur, forme, texture, et localisation spatiale est extrait
SurfImage	Images de visages, de paysages, de photos de rues ou composées d'une seule texture	Ce système offre la possibilité de combiner des caractéristiques bas-niveau (histogramme couleur, histogramme des orientations des contours, ...) et des caractéristiques élaborées comme les images propres (eigenimages en anglais)

Méthode	Domaine d'application	Fonctionnement
CBVQ, VisualSEEK, MetaSEEK	Appliqué aux images de scènes naturelles	Lorsqu'une image est ajoutée à la base, elle est automatiquement segmentée par rapport aux couleurs dominantes qui la composent. Puis les régions sont décrites par des caractéristiques de couleur, de texture et d'organisation spatiale
PicHunter	Images naturelles issues de la base COREL	Utilisation d'histogrammes couleurs, de distribution spatiales des couleurs, et d'annotation textuelle cachée
MARS [Nakazato 03]	Utilisé sur des images d'objets africains manufacturés anciens issus du musée d'histoire culturel Fowler (UCLA)	Basée sur le moteur ImageGrouper, cette approche repose principalement sur un algorithme de bouclage de pertinence pour la recherche d'images par le contenu. Ce système permet à l'utilisateur de valider ou d'invalider les images retournées par le système

Recherche d'images sur le web

ImageScape	Recherche par l'esquisse d'image du web	Pour chaque image de la base, les contours sont obtenus à partir d'un filtre de Sobel et d'un flou gaussien. Un histogramme des motifs 3x3 des contours présents dans l'image est calculé. D'autres caractéristiques sont également utilisées (couleur, laplacien, ...) pour comparer les formes
ImageRover	Recherche basée sur la couleur et la texture	Les caractéristiques utilisées pour décrire les images sont basées sur la couleur et l'orientation des textures

Méthode	Domaine d'application	Fonctionnement
imageRETRO	Testé sur une base de 10 000 images du web, le système réalise d'abord un clustering des images avant de proposer une recherche interactive	Cette méthode s'appuie sur un vecteur de caractéristiques couleurs (histogramme de 15 couleurs, nombre de couleurs, couleur moyenne, écart-type des couleurs, niveaux de gris moyen) et de caractéristiques sur les régions extraites (par croissance de régions couleur)
PicToSeek	Images issues du web et décrites par plusieurs caractéristiques couleur invariantes aux changements de points de vues et d'illuminations.	Les images sont transposées dans un espace couleur défini dans cette approche. Cet espace présente l'avantage d'être invariant du point de vue de prise de l'image, aux conditions d'illuminations, et déformations géométriques. Les contour des images dans cet espace sont décrits à l'aide d'un histogramme
Quicklook2 [Ciocca 99]	Recherche d'images de scènes naturelles à l'aide de mécanismes de bouclages de pertinences	Le moteur de recherche utilise à la fois des descripteurs textuels et des descripteurs de couleurs (histogramme de 64 valeurs, moyenne et écart-type de chaque couleur, ...)
Shoebox	Outil développé pour gérer ses photos personnelles à partir d'annotation vocales et d'extractions automatiques	L'extraction automatique repose sur la couleur moyenne et la variance de chaque canal couleur pour chacune des régions obtenues suites à une segmentation couleur
VIR Image Engine	Moteur de recherche d'images utilisé par *Alta Vista Photofinder* et *Illustra's Visual Intelligence system* (appliqué sur tout type d'image)	C'est un framework extensible qui permet de construire un système de recherche d'images par le contenu adapté. Ce framework permet d'intégrer des descripteurs de couleur (globaux ou locaux), de texture et de forme

Méthode	Domaine d'application	Fonctionnement
PIRIA[Joint 04]	Cet outil est utilisé dans le cadre de la base de donnée Xedix pour indexer les images (100 To de données)	Le moteur dispose de plusieurs descripteurs (>10) caractérisant le contenu des images en termes de couleur (11 espaces colorimétriques), de texture et de forme (silhouette) de manière globale ou locale
IKONA	Basé sur le contenu visuel, il intègre les travaux de recherches de l'équipe IMEDIA de l'INRIA depuis 10 ans	Il repose sur l'extraction automatique d'une grande variété de descripteurs visuels globaux et locaux à partir d'images, de vidéos, et de modèles 3D

Recherche d'images artistiques

AMORE	Images d'art	Les images sont segmentées en 8 régions (au plus) homogènes et ré-échantillonnées à 24x24 pixels. La comparaison des images est réalisée par matching entre ces régions
Picasso	Utilisé dans un catalogue électronique de peintures et de sculptures issues de galeries italiennes	Le système repose sur une segmentation couleur pyramidale, où chaque niveau à un niveau de segmentation (au plus bas niveau, chaque pixel est une région tandis qu'au plus haut niveau l'image ne forme qu'une seule région). Chaque région est alors décrite à l'aide de descripteurs de forme, de couleur, et de ses relations spatiales avec les autres
QBIC	Développé par IBM, ce système a été appliqué aux tableaux de musées américains	Le but est de retrouver des oeuvres à partir d'esquisses ou d'exemples (basé sur la description couleur des images)

Images spécifiques

Méthode	Domaine d'application	Fonctionnement
SMURF	Images marines	Bien que le système incorpore des descripteurs de couleur, texture et forme, il est pour le moment orienté vers la recherche de contours de poissons à partir de polylignes
SQUID	Recherche d'images de créatures marines sur fond uniforme	Description des contours des formes à l'aide de leurs excentricités, leurs circularités, et d'un descripteur obtenu à partir des CSS (maxima de l'image CSS)
FOCUS	Publicités et images naturelles couleurs	voir ci-dessus
ADL	Cartes satellites (recherche de cartes géo-référencées)	Recherche à partir de mots-clés, et de caractéristiques texture
CANDID	Images médicales et images multispectrales sattelitaires	La signature utilisée pour décrire les images correspond à une somme pondérée de gaussiennes. Chaque gaussienne est associée à un type de caractéristique (couleur et texture)
ASSERT	Images de tomographie	Après avoir délimité une zone qui contient une pathologie, le système calcule des caractéristiques dédiées au problème de reconnaissance de lobes (vecteur de 52 caractéristiques)
MIR	Recherche de photos de personnes et de scènes	Moteur de recherche multimodal qui repose sur différentes techniques issues des traitements de textes et d'images
Photobook	Reconnaissance de visage utilisé par la police américaine	Photobook reconnait des visages à partir de détecteurs de visages (eigenfaces) et de descripteurs de texture obtenus après utilisation de la décomposition de Wold (détection de points avec des gradients forts)

Méthode	Domaine d'application	Fonctionnement
IDIR	Reconnaissance de pages de document à partir de leur mise en page et de caractéristiques de texture et de forme	Chaque image est divisée en zones de N x N pixels. Quatre caractéristiques sont alors utilisées sur chaque zone pour la classer comme texte ou comme image (rapport entre le blanc et le noir, niveaux de gris moyen, longueur de plages, et inter-correlation verticale avec les zones voisines)

TABLE 1.5: Tableau récapitulatif des différents projets de recherche d'images par le contenu en fonction de leur domaine d'application

Synthèse du chapitre

Ce premier chapitre présente une étude des méthodes existantes dans la littérature pour décrire une image par son contenu. Après une présentation générale des méthodes existantes pour décrire des images à l'aide de mots-clés, et une présentation rapide des caractéristiques des documents étudiés dans ce manuscrit, nous avons présenté les services de recherche d'images proposés aux utilisateurs. Un des problèmes majeur de la recherche par le contenu repose sur la description du contenu des images. Nous présentons donc les grandes catégories de descripteurs existants, à savoir les approches globales (qui décrivent les formes, les contours, les textures, ou les couleurs), les approches spatiales (qui décrivent l'organisation des éléments présents dans l'image), et les approches locales (qui extraient des zones d'intérêts pour les décrire). Nous présentons également un tableau comparatif des méthodes de chaque catégories, et nous terminons par une présentation succincte des projets existants dans la littérature.

De par la nature même des documents anciens (papier dégradé, images composées de traits, dégradations dues à la numérisation, ...), toutes les méthodes classiques de la littérature que nous venons de voir ne permettent pas d'indexer le contenu de ces bases, principalement du fait de l'inadéquation entre les caractéristiques des images de documents, majoritairement binaires et composées de traits, et les mesures effectuées par ces techniques faisant souvent des hypothèses de continuité radiométriques. Plusieurs études et approches ont proposé des solutions spécifiques à ces documents. Le chapitre suivant présente en détails les documents anciens, vus par les historiens, et un état de l'art des approches informatiques proposées.

Chapitre 2

État de l'art de l'analyse des documents anciens

L'imprimerie naît à Mayence au milieu du XVe siècle, dans l'atelier de Gutenberg. Très vite cette nouvelle technique de (re)production de livres s'exporte à travers toute l'Europe et au-delà [Jimenes 06]. Depuis cette époque, ce sont des centaines de milliers de livres qui ont été imprimés et reproduits.

De plus en plus de bibliothèques nationales [BNF , BNS , BLSB], de projets nationaux, européens ou mondiaux [NaviDoMass 10, Europeana , pas 96, Google b, OCA , IMPACT] cherchent à préserver leur patrimoine documentaire. Dans cette optique, de grandes campagnes de numérisation sont actuellement menées par ces différents projets pour sauvegarder en masse, au format image, des copies de ces documents. De par la masse de documents générés par ces campagnes de numérisation, il devient nécessaire de proposer des services de navigation pour permettre aux usagers de retrouver les documents. Ces services impliquent la nécessité d'organiser les bases d'images, en les indexant sur la base de leur contenu visuel et/ou d'informations contextuelles liées aux spécialistes métiers, et de leur positionnement dans les ouvrages. L'extraction d'informations à partir des documents doit pouvoir refléter leur contenu.

2.1 Analyse des documents anciens par les historiens

Cette section présente comment les historiens classent ces documents pour pouvoir retrouver l'information qu'ils contiennent facilement.

2.1.1 Annotation manuelle et thesaurus

Dans le cadre du projet NaviDoMass, nous avons échangé et travaillé sur des livres avec des historiens de la Renaissance. Ces historiens se consacrent particulièrement à

la diffusion de fonds patrimoniaux grâce à des programmes de recherche associant des compétences en sciences humaines et en informatique.

Les informations contenues dans les documents représentent la mémoire de nos sociétés. Afin d'aider les historiens dans leur travail, différents outils ont été mis au point pour les accompagner dans leurs recherches, notamment pour leur permettre de retrouver ces documents. Ces dispositifs vont du moins élaboré au plus sophistiqué en suivant trois alternatives : **numérisation au format image**, qui transforme le document papier en document numérique brut sans aucune information sur son contenu ; **annotation textuelle**, proposant à un utilisateur d'associer des mots-clés à chaque document ; **transcription diplomatique,** rapportant tous les événements d'un manuscrit (non envisageable dans le contexte de millions de documents).

L'annotation manuelle des documents, et en particulier des images, nécessite l'utilisation d'un vocabulaire commun et standardisé de représentation de leur contenu. Les vocabulaires utilisés par les historiens sont des thesaurus qui répertorient de manière ordonnée et structurée les mots-clés utilisés pour la description. La retranscription ainsi obtenue permet de préserver le contenu et de sauvegarder de manière standardisée tous les ouvrages.

Description de documents anciens à l'aide de Thesaurus

Un thesaurus est un langage contrôlé, utilisé en indexation et pour la recherche documentaire, rendant la recherche documentaire plus efficace. C'est une liste de termes organisés (des descripteurs), se rapportant à un ou plusieurs domaines de connaissance. La plupart des bases de données ont leur propre thesaurus dans lequel on choisit les descripteurs décrivant les documents cités dans la base. Un thesaurus indique les termes utilisés pour décrire sans ambiguïté le contenu des documents. Le vocabulaire sur un thème précis est donc regroupé, limité, classé et régi par des relations de hiérarchie, d'équivalence ou de parenté entre les termes (descripteur générique, descripteur spécifique, descripteur associé).

Plusieurs thesaurus [1] ont été mis au point et normalisés afin de répondre aux différents besoins associés aux documents anciens. Le mot *document* dans ce cas représente aussi bien des images graphiques que des zones textuelles. Ainsi, nous pouvons citer les thesaurus RAMEAU ou Dewey qui sont parmi les plus utilisés par les bibliothèques pour des documents textuels, ou encore des descripteurs iconographiques comme Iconclass qui sont spécialisés pour annoter des images graphiques. Nous présentons succinctement ces thesaurus ci-après.

1. voir[DMOZ]pour de nombreux exemples

Dewey [Dewey 76]

La Classification décimale Dewey (CDD) est un système de classification numérique encyclopédique, couvrant tous les champs de la connaissance. Traduit dans de nombreuses langues, c'est le système de classification le plus utilisé dans le monde. La Dewey est développée et mise à jour régulièrement par le *Decimal Classification Editorial Policy Committee, OCLC/Forest Press* afin de tenir compte de l'évolution des connaissances. Les indices utilisés pour l'indexation systématique donnent lieu à l'établissement de notices qui alimentent un Fichier d'autorité Dewey.

RAMEAU [BNF 80]

Le Répertoire d'Autorité Matière Encyclopédique et Alphabétique Unifié (abrégé en RAMEAU) est un outil créé par la Direction des bibliothèques, des musées et de l'information scientifique et technique (DBMIST) [2]. RAMEAU est un langage d'indexation matière. Ce langage documentaire est utilisé, en France, par la Bibliothèque Nationale de France, les bibliothèques universitaires, de nombreuses bibliothèques de lecture publique ou de recherche ainsi que plusieurs organismes privés. Les mots-clés du langage sont structurés et sélectionnés pour assurer la concision, l'objectivité, la spécificité et la cohérence de description des documents. L'ensemble des mots-clés (ou notices) constitue la *Liste d'autorité nationale* qui est gérée par le Centre national RAMEAU. Cette indexation concerne tous les types de documents (imprimés, documents audiovisuels, documents iconographiques, ...) et permet une recherche par sujet dans les catalogues.

Dans le cas particulier d'annotations d'images dans les documents, des thesaurus spécifiques ont été développés et utilisés en France.

Mandragore

La base de données Mandragore, base iconographique du département des Manuscrits qui réunit les collections occidentales et orientales de miniatures du département, permet une recherche à partir d'un thesaurus spécifique à la BnF.

Le thesaurus employé dans Mandragore est utilisé dans le cadre d'un projet européen sur les accès multilingues, le projet STITCH.

La banque d'images

Les images de la banque d'images, base d'images fixes numérisées du département de la Reproduction alimentée par les images libres de droits des documents reproduits à la demande des clients internes et externes à la BnF, sont indexées à partir du Thesaurus iconographique établi par François Garnier. Ce thesaurus permet l'analyse documentaire

2. RAMEAU est inspiré de la liste d'autorité établie par la Bibliothèque de l'Université Laval au Québec elle-même dérivée des Library of Congress Subject Headings

de figurations anciennes et modernes. Il est utilisé, notamment, pour l'indexation des images de la base Joconde (Catalogue des collections des musées de France) et la base Liber Floridus (Manuscrits médiévaux enluminés des bibliothèques de l'enseignement supérieur).

Iconclass [Iconclass]

Dans le cadre des images traitées dans ce manuscrit, les historiens du Centre d'Etude Supérieur de la Renaissance [BVH] utilisent Iconclass qui fut initialement développé par Henri Van de Waal (professeur d'histoire de l'art à l'université de Leiden). Il permet d'indexer et de classer des bases d'images de documents anciens à l'aide de mots-clés. Ces mots-clés permettent de décrire le contenu sémantique des images à partir d'un dictionnaire structuré hiérarchiquement à partir du sens de ces mots. Ce dictionnaire est composé de 10 classes et 28 000 définitions.

De nombreuses institutions à travers le monde utilisent Iconclass pour décrire et classer leurs collections de manière standardisée, et déjà plus de 40 000 références à des livres ou des articles ont put être indexées. Iconclass s'adresse aussi bien à un public d'historiens de l'art qu'à des visiteurs d'un musée puisqu'il permet de chercher et de retrouver des ouvrages dans des collections à partir du sens de la scène ou d'éléments particuliers qui la composent. Iconclass est actuellement le système de classification par le contenu visuel le plus répandu à travers le monde.

Suite à l'acquisition d'Iconclass en 2006, l'institut d'histoire de l'art *Rijksbureau voor Kunsthistorische Documentatie* (ou RKD), de la Hague aux Pays-Bas, s'occupe de la maintenance quotidienne et des développements futurs du système. Cette acquisition a fait suite aux collaborations qui existaient entre RKD et Henri Van de Waal depuis les années 1950.

2.1.2 Problèmes de l'annotation manuelle d'images

Afin de résumer le contenu des documents ou des images, les historiens les annotent à l'aide de mots-clés issus de thesaurus. Les images sont annotées manuellement sur la base d'une interprétation visuelle des contenus. A partir de cette analyse, des mots-clés sont choisis pour représenter les images.

Annoter des images manuellement pose plusieurs problèmes majeurs. Tout d'abord, il faut préciser que c'est une tâche fastidieuse. De plus, dans un contexte d'analyse automatique de grandes masses de documents anciens (plusieurs millions de pages et d'images), il faudrait de nombreux spécialistes du domaine dédiés à cette tâche. En dehors du coût économique de ces spécialistes, le problème majeur repose sur la variabilité des domaines d'étude des différents chercheurs en histoire. On peut ainsi citer les chercheurs en histoire du livre qui s'intéressent aux techniques d'imprimerie, à l'esthétique des livres, ou à l'économie qui entourait les livres (contrefaçons, échanges de fontes, ..) ; les historiens de la langue et de l'orthographe qui observent l'évolution des pratiques orthographiques

et du matériel typographique ; ou encore les chercheurs en histoire culturelle et histoire des mentalités qui s'appuient sur l'évolution des caractères pour déduire des indices de l'évolution des cultures et mentalités.

Là où certains s'intéresseront plus particulièrement à l'aspect esthétique d'une image, d'autres s'intéresseront plus particulièrement à d'autres détails de l'image (personnages qui la composent, architecture représentée, etc). Cette variabilité apparaîtra alors au travers des descriptions et des mots-clés retenus, puisque chaque domaine d'étude spécialise donc sa description en fonction de ses attentes. Par exemple, un historien du livre s'intéressera d'avantage à l'esthétique d'une image, là où un historien culturel s'intéressera plus au côté social de la scène qui est représentée.

Cette variabilité de but et d'analyse pose le problème d'une description générique et adaptée aux différentes recherches. En effet, est-il possible de répondre de manière pertinente aux attentes des différents domaines de recherche de l'histoire quand une image a été annotée par un chercheur d'un domaine particulier ? Est-ce que la description via un thesaurus permet de représenter fidèlement le contenu d'une image ? Faut-il choisir un thesaurus utilisé par toutes les bibliothèques ou un thesaurus spécifique à chaque domaine ?

Les réponses à ces questions sont difficiles, et afin d'aider les historiens dans leurs recherches, des outils informatiques de détection d'éléments particuliers ont été développés. Ces outils ne prétendent pas annoter les images avec des mots-clés exacts, mais proposent un pré-traitement qui doit pouvoir faciliter la tâche de recherche des historiens (séparation des différents types d'éléments que l'on peut trouver dans une page, catégorisation de ces éléments à partir de considérations générales et validées par tous les domaines, ...). Nous présentons ces outils ci-dessous.

2.2 Outils informatiques d'analyse des documents anciens - État de l'art

L'annotation de documents ou d'images présente plusieurs problèmes difficiles (subjectivité de l'annotation, description non identique en fonction des scripteurs et du temps, ...). Pour pallier à ces problèmes, des outils développés à la base par les communautés scientifiques et repris par des organisations telles que la bibliothèque nationale française[Gallica], ou celui de la bibliothèque nationale anglaise [BLSB], commencent à apparaître, afin de proposer des services de recherche à partir du texte des documents. Toutefois, ces moteurs ne fonctionnent que sur les documents textuels et ne permettent pas de naviguer dans des contenus graphiques.

D'autre part, de nombreux projets de la littérature développés au cours des 20 dernières années, réalisent l'indexation d'images par leur contenu, offrant ainsi des services de navigation dans des documents riches, entre autres contenant des images. Toutes les techniques de recherches d'images ont conduit à des systèmes de recherche d'images que

l'on peut catégoriser :

1. recherche par **navigation libre** dans la base d'image : ce type de recherche consiste à proposer à l'utilisateur l'ensemble de la base d'images, qu'il parcourt selon ses envies. Cette technique n'est que peu utilisée puisqu'elle impose de naviguer longtemps au sein de la base sans garantir de résultat à l'utilisateur. Cependant, dans certains cas d'utilisation précis, ce type de recherche pourra être mis en place.
2. recherche par **mots-clés** : cette technique est à la base des moteurs de recherche utilisés tous les jours par des millions de personnes. Le principe général consiste à associer des mots-clés aux images (manuellement ou (semi-)automatiquement) et à rechercher des images à partir de ces mots-clés. Cette technique est détaillée dans la section **??**.
3. recherche par le **contenu** : plus communément appelée Content-Based Image Retrieval (CBIR), c'est un domaine de recherche très actif depuis les années 1990. La section 1.3 présente un panorama des approches de description des images pour une indexation par le contenu avec des outils génériques, dont nous avons montré en quoi ils ne sont malheureusement pas ré-utilisables pour les images de documents [3]. Le type de données qu'ils traitent, les méthodes sur lesquelles ils reposent et enfin les descripteurs qu'ils utilisent sont évoqués ci-après. Enfin, la section 2.2.2 présente les approches de la littérature dédiées à la description et à l'indexation des images de documents.

2.2.1 Pourquoi les méthodes usuelles ne fonctionnent pas sur les documents anciens ?

Les documents anciens sont différents des images « traditionnelles » du fait qu'ils ont été conçus pour une lecture humaine et véhiculent de ce fait des messages explicites ou implicites que l'on peut retrouver dans l'organisation spatiale des objets (illustrations, pages de journaux, ...). De ce fait, leurs propriétés radiométriques sont extrêmement différentes de celles des images classiquement utilisées dans les communautés d'analyse d'images. L'adéquation des méthodes classiques issues de l'analyse d'image aux documents doit donc être étudiée pour analyser la ré-utilisabilité de certains outils. Cette complexité découle de la présence d'éléments de nature différente (textes, images, etc), et de la manière dont ils ont été créés. Les méthodes usuelles, comme vu dans la partie 1.4, sont généralement décrites pour reconnaître des images couleurs, ou des structures particulières dans les images (détection de formes, de textures, ...). Extraire de l'information à partir des documents anciens est une tâche complexe puisqu'elle doit permettre de décrire leurs contenus (textuel, graphique, etc) tout en tenant compte de leur complexité (pas de structure apparente, images monochromes, images de traits, ...). Cette complexité impose généralement de traiter les documents selon la chaîne de traitement suivant :

3. La non adéquation de ces méthodes avec les documents anciens sera présentée par la suite

1. Pré-traiter l'image pour la restaurer : les images sont très souvent dégradées par le temps et leur utilisation (voir A). Cette restauration (débruitage, filtrage) transforme l'image pour améliorer les conditions de travail des traitements suivants ;
2. Décrire le contenu des images à l'aide des caractéristiques vues précédemment 1.4 (moments de Zernike, Descripteur de Fourier Généralisé, Shape-Context, ...). Cette description peut être soit globale, c'est à dire sur l'ensemble de l'image, soit locale et dans ce cas précis elle ne décrira que certaines zones précises des images (cas du détecteur et descripteur SIFT) ;
3. Définir des métriques pour mesurer la similarité entre les images : ces métriques permettent de comparer deux descriptions, et de dire si deux images sont semblables. Elles sont donc en lien direct avec les techniques de description utilisées pour décrire le contenu des images (isomorphisme exact ou inexact dans le cas de graphes, distance dans le cas de vecteurs ou d'histogrammes, ...) ;
4. Structurer l'espace des caractéristiques : pour proposer une réponse rapide à l'utilisateur et éviter une comparaison exhaustive pendant la navigation dans la base de données. Cette hiérarchisation de l'espace dépend également du type de structure utilisé pour décrire les images (graphes, vecteurs, ...), et repose généralement sur des approches de clustering ;
5. Penser des interfaces Homme-Machine adaptatives qui exploitent des informations sur l'utilisation de l'utilisateur (bouclage de pertinence - relevance feedback).

2.2.1.1 Spécificités des documents étudiés au cours de la thèse

Au cours de cette thèse, nous nous sommes plus particulièrement intéressés aux étapes 1, 2, 3 et 4 évoquées précédemment. Ces étapes ont été étudiées pour proposer un outil d'analyse et de recherche d'images graphiques de documents du XVème et XVIème siècles par le contenu. Les images nous ont été fournies sous forme papier et au format numérique par le *Centre d'Etude Supérieur de la Renaissance*. Ces documents datent de La Renaissance, période qui correspond au début de l'imprimerie. A l'époque, les lettrines étaient imprimées à l'aide de tampons en bois qui étaient pressés sur le papier. Ces tampons étaient taillés à la main dans le but d'illustrer des documents (soit à l'aide d'éléments décoratifs, soit à l'aide de représentation de scènes du passé). Ces tampons présentent donc deux types de spécificités, le support et la technique d'impression.

Le support

Le support utilisé pour imprimer les documents de l'époque est le papier. Celui-ci est composé d'éléments naturels et n'était pas, à l'époque, traité pour résister aux dégradations du temps (voir Annexe A). Les fibres organiques qui composent le papier se dégradent au cours du temps et font apparaître plusieurs problèmes (jaunissement, affinement, et fragilisation des pages qui peuvent se déchirer à chaque manipulation).



FIGURE 2.2.1: Exemples de lettrines composées de traits

FIGURE 2.2.2: Exemples de lettrines en gros plan pour mettre en évidence les spécificités des images imprimées avec des tampons en bois

Ces problèmes obligent donc à réfléchir à de nouveaux pré-traitements pour faire disparaître les effets du temps. Ces pré-traitements ne concernent que les traiteurs d'images puisque les historiens sont parfois intéressés par la conservation de ce fond bruité (traces représentatives de la vie du livre).

La technique d'impression

La seconde particularité des documents de La Renaissance repose sur les techniques d'impression utilisées. Cette époque correspond aux débuts de l'imprimerie et les documents étaient imprimés à l'aide de tampons en bois. Ces tampons étaient utilisés comme une signature artistique propre à chaque imprimeur. Les historiens s'en servent aujourd'hui pour identifier de manière unique la provenance et l'époque des documents.

De plus, de par leur conception, les tampons ne permettaient d'imprimer que des motifs en noir et blanc sans nuances de gris (le papier était recouvert ou non d'encre). Pour créer des nuances de gris, et ainsi créer des ombres et des effets de relief, les imprimeurs utilisaient des zones rayées, composées de traits parallèles. Des exemples peuvent être observés dans la figure 2.2.1, et on peut noter que les zones de traits font partie intégrante des membres des personnages.

2.2.1.2 Les lettrines

De nombreux types de documents graphiques issus de documents anciens existent (une liste détaillée peut-être obtenue dans [Delalandre 09]). Nous nous sommes particulièrement intéressés aux lettrines (voir figure 2.2.1) qui ont été très réutilisées à travers les ouvrages et le temps. Une lettrine est une image à caractère décoratif qui représente une lettre. Celle-ci marque généralement le début d'une page et/ou d'un chapitre. Même si l'objet d'application est la lettrine, les études menées dans ce manuscrit ont pour objectif de définir des outils génériques d'indexation d'images graphiques au sens large. Plusieurs travaux se sont penchés sur les documents anciens et plus particulièrement sur

les images graphiques (dont les lettrines).

Tous les types de documents anciens du début de l'imprimerie présentent des traces du temps qui passe (jaunissement et déchirement du papier, étalement de l'encre, apparition du verso sur le recto, ...), et des outils utilisés pour les créer (images binaires composées de traits, propres à chaque imprimeur, créées à l'aide de tampons en bois qui s'usent, ...). Ces documents présentent donc des particularités fréquentielles (forts gradients), des discontinuités radiométriques, et une complexité de l'information qu'elles renferment (images créées par l'homme, pour l'homme, composées de formes et de textures). Toutes ces caractéristiques rendent les méthodes usuelles inefficaces sur les documents anciens. Des travaux se sont alors orientés, pour apporter des réponses à ces problèmes.

2.2.2 Caractérisation de la structure et de l'écriture dans les documents anciens

Au cours des quinze dernières années, plusieurs travaux et projets ont cherché à caractériser et indexer les images de documents anciens par leur contenu. On peut ainsi citer deux projets français (les projets Madonne [4] et Navidomass [5]), qui ont permis de développer de nombreux travaux de recherche sur l'étude de documents issus du patrimoine. L'étude d'images de documents anciens peut-être résumée en trois catégories [Ogier 06] :

- L'analyse de la structure des documents
- La reconnaissance de l'écriture manuscrite ou typographique
- L'indexation et la reconnaissance des objets graphiques (images, logos, symboles)

Chacune de ces catégories a fait l'objet de travaux de recherche dont certains prennent en compte le bouclage de pertinence et les interactions avec l'utilisateur (comme le système AGORA pour l'analyse de la structure des documents anciens). Nous présentons succinctement les deux premières approches ci-dessous. Ce manuscrit s'inscrit dans la troisième catégorie (reconnaissance d'objets graphiques) et nous présentons un état de l'art des méthodes dédiées à cette catégorie dans la suite de cette section.

Analyse de la structure des documents

La structure d'un document est généralement liée à un modèle de présentation et/ou d'organisation qui vise à aider l'utilisateur dans sa compréhension du document. Cette structure permet donc d'identifier comment les éléments sont placés sur une page. Elle revient à séparer les différents blocs qui composent une page puis à les identifier, pour retrouver la structure physique du document. La reconnaissance des éléments segmentés (texte en gras, image couleur, dessins, petits caractères...) donne accès à ce qu'on appelle la structure fonctionnelle intermédiaire (voir figure 2.2.3.c), pour finalement permettre

4. http ://madonne.univ-lr.fr/
5. http ://navidomass.univ-lr.fr/

de retrouver la structure logique du document (figure 2.2.3.d) via l'interprétation de la structure fonctionnelle intermédiaire (l'image est composée d'un titre, d'un édito, d'une grande photo...).

Selon [Doermann 98], la structure d'un document se présente sur 3 niveaux :

- **la structure logique**, qui correspond au sens que l'on souhaite donner à une page, et se traduit par un choix de titres, de photos et de mise en page ;
- **la structure fonctionnelle intermédiaire**, qui consiste à affecter une caractérisation visuelle concrète, et à définir l'organisation spatiale des éléments à mettre en page (un titre sera une zone de texte de gros caractères en haut de la page, un édito est un texte en italique sur deux colonnes et toujours à gauche) ;
- **la structure physique** : c'est le résultat du processus de réalisation intégrant les contraintes logiques, fonctionnelles et enfin la taille de la page.

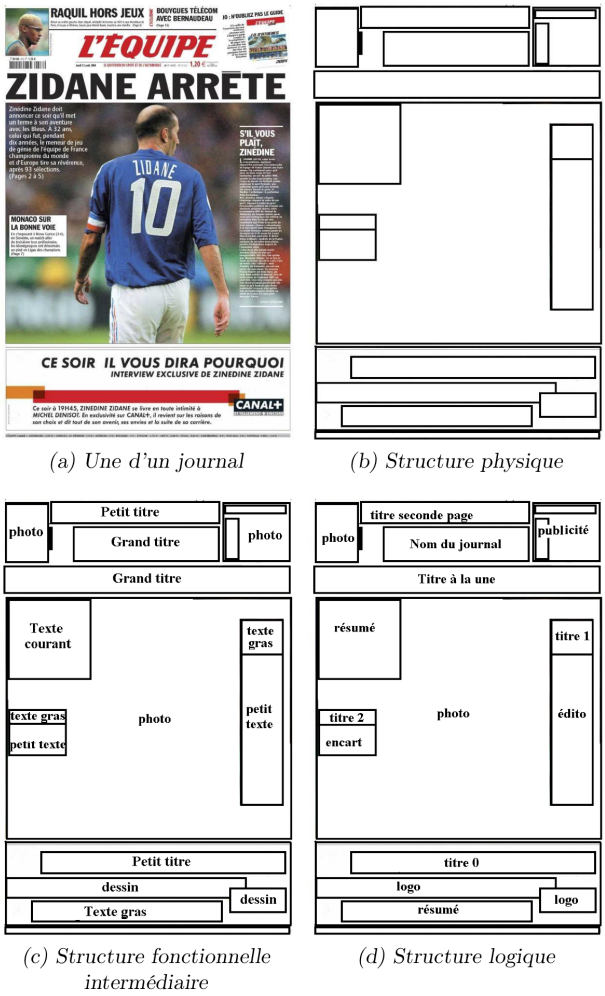
Dans le cadre des projets Madonne et Navidomass, plusieurs travaux ont proposé d'étudier la structure particulière des documents anciens. Deux époques particulières peuvent être citées en terme de structure. Ainsi, de par les contraintes des premières machines d'imprimerie, les documents ont une structure propre à chaque imprimeur. Chacun d'entre eux choisissait la manière dont il positionnait les éléments sur une page. Avec le temps, les nouvelles techniques d'impression, moins contraignantes, ont laissé place à l'imagination des rédacteurs et il n'est pas rare de trouver des pages sans réelle structure apparente dans les journaux que nous achetons tous les jours. Cependant, de plus en plus de normes sont apparues avec le temps pour uniformiser les structures actuelles des documents, et ainsi proposer des schémas types d'impression de documents.

Identifier la mise en page permet donc d'identifier la page en elle-même, mais oriente également les traitements à utiliser par la suite (appliquer un OCR sur les zones de textes, appliquer un algorithme de reconnaissance d'images sur les illustrations, ...). Nous pouvons citer les travaux de Bertrand Couasnon [Coüasnon 03] qui annote des registres militaires du 19ème siècle (voir figure 2.2.4) en utilisant des techniques de grammaire en 2 dimensions intégrées au système DMOS.

Une autre contribution centré sur l'utilisateur, le système AGORA, a été conçu pour analyser de manière interactive la structure de documents de la Renaissance [Ramel 06, Journet 08a]. En fonction de ses besoins (extraction de lettrines, d'annotations dans les marges, des titres, etc), l'utilisateur peut construire des scénarii permettant d'étiqueter, de fusionner ou de supprimer les blocs extraits automatiquement. Les scénarii peuvent être stockés et modifiés pour être réutilisés sur d'autres séries d'images, et ainsi proposer des traitements par lots.

Reconnaissance de l'écriture manuscrite ou typographique

La reconnaissance d'écriture, manuscrite ou typographique, est un domaine de recherche très actif depuis plusieurs décennies [Prum 10, Mohand 10]. Elle rassemble de nombreux chercheurs à travers le monde et de nombreuses compétitions prennent place pour évaluer et comparer les différentes approches proposées (Handwriting Competition,



(a) Une d'un journal (b) Structure physique

(c) Structure fonctionnelle intermédiaire (d) Structure logique

FIGURE 2.2.3: Les différentes structures d'un document, telle que définies dans [Journet 06]

Online Arabic Handwriting Recognition, Handwritten Farsi/Arabic Character Recognition Competition, Handwriting Segmentation Contest, Arabic Handwriting Competition - pour ne citer que celles qui ont eu lieu au cours de la conférence ICDAR 2009 [Grosicki 09]).

En marge des techniques classiques de reconnaissances de caractères (OCR), des travaux récents [Retornaz 07] ont également cherché à détecter et identifier les zones de textes dans différents types de documents. Le but n'est plus simplement de reconnaître le texte, mais également de le détecter à l'aide d'opérateurs de morphologie mathématique (ouverture ultime).

Nous allons maintenant nous intéresser plus particulièrement à l'analyse et la reconnaissance d'images graphiques de documents anciens.

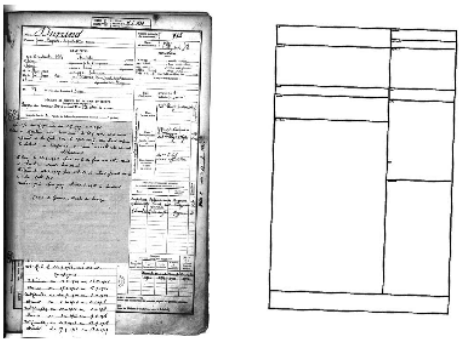


FIGURE 2.2.4: Exemples d'annotation automatique de formulaires militaires en utilisant l'approche DMOS [Coüasnon 03]. A gauche, un exemple de formulaire militaire sur lequel a été ajouté une feuille, et à droite la structure reconnue par l'approche

2.2.3 Analyse et reconnaissance d'images graphiques de documents anciens

2.2.3.1 Cas des documents anciens en général

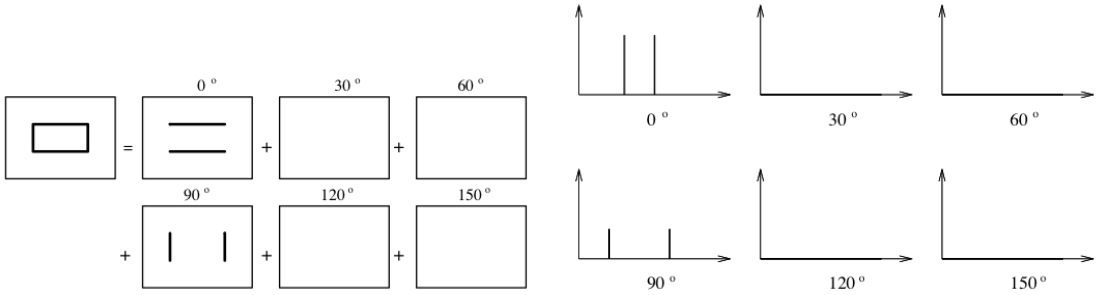
Durant cette dernière décennie, de nombreux travaux ont cherché à caractériser et indexer les images de documents anciens par leur contenu.

Une première approche [Nies 96], à la base du moteur de recherche d'images par le contenu *LCPD*, permet de rechercher des cartes de visites et des portraits du XIVème et XXème siècles dans une base de 1154 images (la base devrait être complétée pour atteindre 50 000 images). L'utilisateur peut choisir entre différents vecteurs de description des images, soit issus des valeurs des pixels (vecteur obtenu via l'histogramme de projection des niveaux de gris de l'image ou du gradient de l'image), soit obtenu à plusieurs niveaux d'échelle (*local binary pattern vector* définit par les auteurs). La recherche se fait par comparaison du vecteur entre l'image requête et les images de la base.

Delalandre et al. [Delalandre 09] présentent un état de l'art d'images de documents anciens, dans lequel deux méthodes sont appliquées sur des documents ornementaux anciens. Une première méthode, développée par Chen et Szabo [Chen 03], est utilisée sur des images du XVIIème siècle. Elle consiste à chercher des points d'intérêt dans les images et à calculer les moments de Zernike autour de ces points. L'extraction de points d'intérêt dans des images sans motif (au sens des lettrines) est possible puisque seuls les contours des objets seront détectés. Dans le cas des lettrines, images composées de traits, un grand nombre de points d'intérêt seront détectés et ne seront pas représentatifs du contenu des images.

L'autre méthode présentée [Bigun 96], est appliquée sur des ornements de la biblio-

thèque de Lausanne [6], dans le cadre du projet suisse Passe-Partout [pas 96], et utilisée sur la base Maguelone [Mag]. Elle repose sur l'extraction de radiogrammes d'orientations sur les images pour les indexer. Ces radiogrammes permettent d'identifier des structures linéaires parallèles dans différentes orientations. Six radiogrammes à différentes orientations sont obtenus et décrits par dix coefficients de Fourier. Une distance euclidienne est calculée entre les six paquets de dix coefficients pour renvoyer les images les plus proches. Cette méthode obtient de bons résultats sur les ornements qui ne sont composés que de motifs. Elle est également utilisée dans le moteur de recherche TODAI, pour décrire et retrouver des images ornementales suisses.



(a) Orientations choisies pour projeter l'image (b) Histogrammes associés aux projections effectuées

FIGURE 2.2.5: Schéma de décomposition d'une image selon des radiogrammes. Dans cet exemple, l'image d'origine (le rectangle) est décomposée selon 6 orientations (sous-figure de gauche) et des histogrammes sont obtenus à partir des projections (sous-figure de droite) selon un principe similaire à la transformée de Hough [Duda 72]

Enfin, notons l'approche appliquée sur des images ornementales colorées qui a été développée par Hurtut et al [**?**]. Elle propose une comparaison de l'organisation spatiale des couleurs entre les images en utilisant la distance de Earth Mover's. Cette distance mesure le coût minimal pour transformer une distribution de points A en une distribution de points B. Les distributions de points correspondent aux régions couleurs qui sont extraites puis sous-échantillonnées. Les résultats de cette distance sont utilisés en recherche d'image, en utilisant la couleur de chaque pixel d'images sous-échantillonnées. Cette méthode ne nécessite donc pas de segmenter les images, par contre elle n'est pas applicable aux images du début de l'imprimerie puisqu'elles sont en noir et blanc, et les sous-échantillonner ferait disparaître leurs détails.

Les méthodes que nous venons de présenter ont été utilisées sur des images ornementales du XVIIème et XVIIIème siècles. Dans le cas d'images de documents de la

6. [BNS]

Renaissance, plusieurs travaux ont également été proposés.

Une première méthode [Journet 08b], plus orientée sur les pages de documents, propose une segmentation et description de celles-ci à l'aide de descripteurs textures (méthode utilisée pour segmenter les blocs d'une page dans le logiciel Agora - voir ci-dessus). Ces descripteurs permettent d'identifier et de segmenter les différents éléments constitutifs d'une page. Ils s'appuient sur la fonction d'auto-corrélation bi-dimensionnelle pour obtenir une *Rose des directions* (voir figure 2.2.6 pour des exemples) accompagnée de trois indices (orientation principale, intensité de la rose des directions et description de la forme globale de la rose). Ces indices sont utilisés pour différencier les types de documents et segmenter les lettrines des documents originaux.

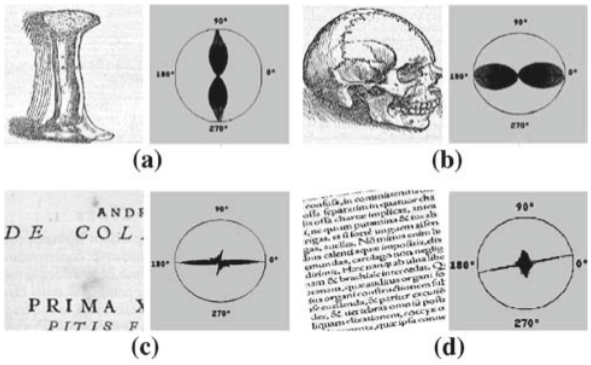


FIGURE 2.2.6: Exemples de roses des directions obtenues sur différentes images : a) dessin vertical, b) un crâne, c) quatre lignes de texte, d) texte penché

Dans un contexte un peu différent puisqu'il s'agissait de dater les images, une approche proposée par [Baudrier 07a, Baudrier 08], recherche les dissimilarités entre les images. Cette information est utile aux historiens puisqu'elle permet de détecter les différences minimes qui existent entre des images semblables. Étant donné que les images étaient toutes issues du même tampon en bois, le tampon se dégradait au cours du temps. Détecter les différences permet de connaître la chronologie d'impression de ces images. Cette approche utilise l'information contenue par tous les pixels pour comparer les images à l'aide d'une carte de dissimilarité locale (des exemples de cartes peuvent être observés dans la figure 2.2.7). La caractéristique principale de cette méthode repose sur une comparaison d'images sans extraction de descripteurs. Cependant, la méthode n'est pas utilisable dans un processus en ligne par rapport aux temps de calcul, n'est pas robuste aux changements d'échelle et compare les images sur une information globale sans prendre en compte l'information locale.

2.2.3.2 Cas des lettrines

Le cas des lettrines a été traité dans la littérature de ces cinq dernières années pour proposer de reconnaître soit la lettre, soit le style. Ainsi, Naegel *et al* [Naegel 09] proposent

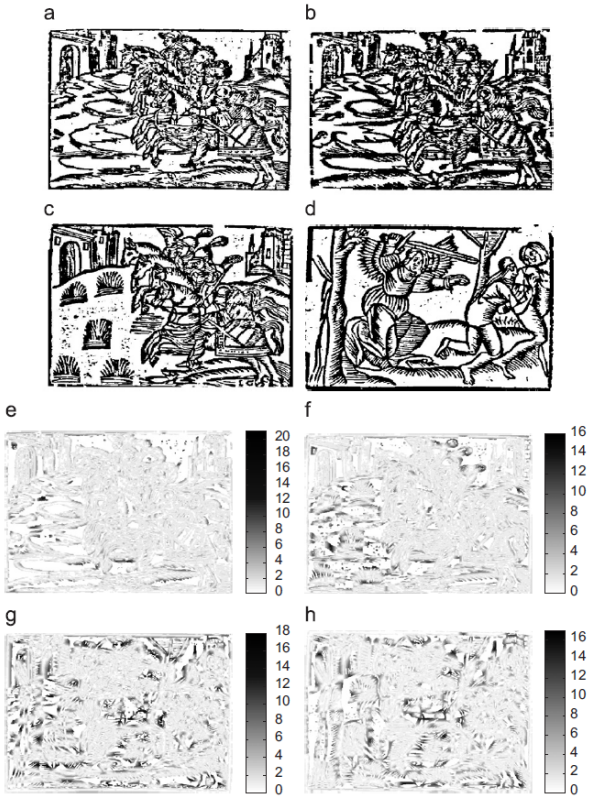


FIGURE 2.2.7: Impressions médiévales et leurs cartes de dissimilarités globales telles que définies dans [Baudrier 08]. Dans cet exemple, les cartes représentent les dissimilarités entre la première image (a) et les autres

de binariser les images de lettrines, puis de ne retenir que les composantes connexes qui présentent certaines propriétés. La sélection de ces composantes est réalisée à l'aide d'un arbre des composantes, où chaque noeud (*ie* chaque composante connexe retenue) est décrit à l'aide du descripteur de Fourier généralisé (GFD). L'extraction de la lettre donne de bons résultats mais ne propose pas de reconnaissance du style.

Enfin, deux méthodes, utilisant la loi de Zipf [Zipf 49], se sont intéressées plus particulièrement aux styles des lettrines. Cette loi, définie empiriquement par George Kingsley Zipf, est une *loi puissance* qui se base sur la fréquence et le rang d'apparition des mots dans un texte. Une première méthode [Pareti 08]applique cette loi sur l'image en prenant des sous-images comme motifs de l'image et en calculant leur fréquence et leur rang. A partir de ces calculs, une courbe représentant la répartition des différents motifs dans l'image est calculée. Des descripteurs sont extraits de cette courbe et les lettrines sont classées par style à partir de ces descripteurs. Les résultats obtenus sont intéressants mais ne permettent pas de décrire les images sur leur contenu de manière explicite. En effet, les styles obtenus permettent uniquement de différencier les images en fonction de leur motif et/ou fond.

Dans [Chouaib 09], les auteurs proposent une seconde méthode propre aux lettrines, reposant sur une approche connexe à celle citée ci-dessus. Ces travaux ont également pour but d'indexer les lettrines à partir du fond et du motif. Toujours sur la base d'une loi de Zipf, ils extraient les motifs des lettrines et calculent des descripteurs sur ces motifs. Les descripteurs sont basés sur le rang et la fréquence des motifs et un vecteur descriptif. Ce vecteur correspond à une pondération des fréquences d'apparition des différents motifs en utilisant la méthode TF-IDF [Jones 79]. Les résultats améliorent la méthode présentée par Pareti et al. dans [Pareti 06b] mais ne différencient les images qu'au niveau du style.

Enfin, pour clore cet état de l'art, le tableau 2.1 propose un récapitulatif des méthodes présentées dans cet état de l'art.

Référence	Type de données	Description
[Nies 96]	Portraits et Cartes de visites	Plusieurs vecteurs peuvent être obtenus pour décrire les images (vecteurs obtenus en utilisant des histogrammes de projection). Le moteur de recherche d'images par l'exemple LCPD se base sur ces vecteurs pour comparer les images.
[Chen 03]	Images du 15ème siècle	Recherche de points d'intérêt et description de ces points par le descripteur de Zernike, dans des images non composées de traits.
[Bigun 96]	Images d'ornements	Identification des structures linéaires parallèles à des orientations particulières pour identifier des images ornementales (issu du projet Passe-Partout)
[?]	Pages entières	Comparaison d'images couleurs issues de documents anciens. La comparaison repose sur l'utilisation de la distance de Earth Mover's entre les régions de couleurs extraites
[Journet 08b]	Pages entières	Séparation des éléments d'une page à l'aide de descripteurs textures
[Baudrier 07a]	Images graphiques	Comparaison d'images à partir de cartes de dissimilarités locales
[Pareti 06a]	Images graphiques	Identification du style des lettrines à partir d'une loi puissance
[Chouaib 09]	Lettrines	Identification du motif à partir d'une loi puissance

TABLE 2.1: Tableau récapitulatif des méthodes de description d'images de documents anciens

2.3 Discussions sur l'état de l'art

Toutes les méthodes que nous venons de présenter reposent sur une approche cartésienne du problème de caractérisation d'images, c'est à dire qu'elles simplifient les images pour en étudier un aspect particulier (hors de son contexte). Cependant, les images de documents anciens sont complexes et composées de différents types d'information. Afin de prendre en compte cette complexité, il faudrait être capable d'en extraire les différents types d'information, de les décrire, et de regrouper ces différents types d'information pour obtenir une description à la fois exhaustive, et représentative de cette complexité.

En effet, les images de documents anciens sont composées de zones uniformes et de zones de traits. Les zones uniformes représentent généralement les objets de l'image, pendant que les zones de traits sont utilisées pour créer du relief et un effet de perspective dans l'image. C'est la combinaison de ces deux types de zones qui permet de construire une image visuellement esthétique et compréhensible. Si l'on prend l'exemple d'une lettrine, la décrire d'un point de vue texture uniquement, sans utiliser les formes qui la compose, ou décrire le fond sans décrire la lettre ne paraît pas cohérent. Cela correspond à une vision de traiteur d'images qui consiste à ne s'appuyer que sur des descripteurs bas-niveau (de couleurs, de formes, de textures, ...). L'objectif global est d'essayer de revenir à des informations utilisées dans la production des lettrines pour obtenir une esthétique visuelle. Dans ce cadre, l'idée est de regrouper par "catégories" les différentes couches visuelles (dans notre cas : textures, formes, contours, ...), et d'essayer de les extraire.

Pour reprendre un extrait d'Edgar Morin [26], sociologue, philosophe et grand penseur français de la complexité, le principe de simplicité impose de disjoindre, le principe de complexité enjoint de relier, tout en distinguant. Toutes les approches scientifiques jusqu'à la moitié du XXème siècle cherchaient à simplifier les problèmes compliqués pour les résoudre. Il faut dissocier les problèmes complexes des problèmes compliqués. La complexité par définition signifie ce qui est tissé ensemble, ce qui est relié, ce qui n'impose en rien d'être composé de problèmes compliqués. Au contraire, là ou un problème compliqué va nécessiter une simplification sans chercher à respecter son environnement, un problème complexe nécessitera des allers-retours entre description certaine par simplification et description incertaine de son contexte. Il paraîtrait donc essentiel d'intégrer, dans un schéma d'ensemble, séparabilité et logique, avec une séparabilité de l'information et un respect de la logique globale.

Une méthode développée par Uttama et al. dans [Surapong 08], propose de décomposer une image pour identifier le contenu des lettrines d'une manière complexe. Le processus de segmentation repose sur une décomposition en différentes couches d'informations (formes, textures, contours, ...) inspiré des principes de perception visuelle. A partir de cette étape, chaque couche va contenir les zones propres à un type d'information. Deux méthodes ont été implémentées pour décrire les images à partir de ces couches, tout en conservant une description globale de l'image. Une première repose sur

le calcul de la longueur de l'arbre couvrant minimum (MST [Hero 99]) tandis que la seconde méthode calcule un histogramme des fréquences d'apparition des relations entre composantes connexes de l'image. Cette dernière approche extrait des zones d'intérêt, décrit le contenu des régions, et les relations spatiales entre ces zones. Par contre, les zones extraites n'ont pas de lien direct avec l'information recherchée par les historiens.

Synthèse du chapitre

Le dernier chapitre de cette première partie du manuscrit, décomposé en trois thèmes, présente un état de l'art de l'analyse des documents anciens. Le premier thème s'intéresse aux outils utilisés par les historiens pour annoter les documents, à savoir les thesaurus, et les problèmes liés à l'annotation manuelle (variabilité, tâche fastidieuse). Puis, dans un second thème, nous nous sommes intéressés aux outils informatiques dédiés aux documents anciens. Nous avons commencé par une démonstration de l'inadéquation des méthodes usuelles aux caractéristiques des documents du début de l'imprimerie. Dans un second temps, nous avons présenté les méthodes de la littérature qui se sont intéressées plus particulièrement à la structure et aux écritures de ces documents, puis nous avons continué par une étude des méthodes d'analyse des images graphiques de documents anciens. Pour cette partie, nous avons différencié les approches orientées documents anciens, des approches spécialisées pour les lettrines. Enfin, le dernier thème de ce chapitre proposait une discussion de l'état de l'art, et présente la complexité des images graphiques de documents anciens, et la nécessité d'étudier ces images par une approche complexe.

2.3.1 Démarche proposée : une approche complexe d'analyse d'images de documents anciens

Partant de l'idée d'Edgar Morin, et du constat que les méthodes de la littérature sont spécialisées dans l'analyse d'une seule catégorie d'information, il nous a paru pertinent de développer **une méthode complexe**. Une méthode complexe est une méthode qui doit permettre de lier la connaissance métier et la connaissance des traiteurs d'images.

Soucieux d'obtenir une description plus proche des attentes des utilisateurs, nous avons cherché à utiliser et à structurer leurs connaissances. En ce qui concerne les connaissances des traiteurs d'images, cette méthode repose sur un enchaînement de trois étapes :

1. Simplification de l'image pour isoler les zones formes des zones texturées ;
2. Appliquer des méthodes de description spécialisées sur chaque sous-image obtenue après simplification ;
3. Reconstruire une description complexe de l'image.

La simplification et la description spécialisée consistent à étudier chaque élément de l'image et à décrire son contenu en fonction de sa typologie de départ. La reconstruc-

tion quant à elle consiste à agréger les différentes descriptions tout en caractérisant les relations entre les différentes zones d'intérêts.

Enfin, une fois les connaissances des deux domaines exprimées, notre but est de lier ces domaines pour réduire le fossé sémantique et ainsi proposer une recherche qui associe la connaissance des différents experts (experts en traitements des images d'une part, et experts du domaine historien de l'autre). Nous présentons la démarche globale et les étapes dans la suite de ce manuscrit.

Deuxième partie

Approche proposée

Chapitre 3

Représentation des connaissances d'experts d'un domaine

3.1 Problématique liée à la représentation des connaissances d'experts d'un domaine

De plus en plus de documents anciens sont numérisés pour sauvegarder leur contenu. Cette numérisation génère une masse de documents et impose la nécessité de les gérer efficacement. Leur gestion repose sur des mécanismes d'indexation et de recherche qui tentent de répondre aux attentes des utilisateurs. Afin de répondre à la croissance rapide des contenus multimédia disponibles, plusieurs tentatives ont abordé le problème très connu du fossé sémantique [Smeulders 00] pour proposer une analyse et une interprétation sémantique des images.

L'annotation automatique d'images, qui a été un des défis de cette dernière décennie, a été introduite au début des années 2000, et les premiers travaux se sont orientés sur les approches d'apprentissage statistiques et probabilistes. Ces approches fournissent des outils puissants et efficaces pour créer des associations entre les caractéristiques visuelles et les concepts sémantiques [Barnard 03, Lavrenko 03].

Cependant, même si les approches statistiques décrivent bien le contenu visuel des images, l'interprétation sémantique qui en découle se limite souvent à la détection de manifestations particulières de la sémantique, sans permettre de modéliser le sens des images tel qu'il est perçu par les humains. Ces approches présentent également plusieurs limitations dans le cas où elles sont appliquées à de grandes bases d'images [Liu 07] puisque les performances varient en fonction du nombre de classes considérées et des images traitées. Ces variations de performances peuvent s'expliquer par la grande variabilité qui réside en intra-classes dans le cas de grandes bases d'images (il peut être difficile de représenter toutes les formes de bâtiments à l'aide de caractéristiques visuelles), et des similitudes qui peuvent apparaître entre des classes. Ces disparités intra-classes,

et ces similitudes inter-classes, peuvent conduire à des situations conflictuelles dans le processus d'annotation, et prouvent le manque de corrélation entre les caractéristiques sémantiques et bas-niveau.

Toujours à la recherche de modèles qui pourraient aider à mettre en corrélation les caractéristiques de bas niveau et la sémantique de haut niveau, des approches se sont basées sur les "connaissances contextualisées". En construisant des hiérarchies sémantiques, ou en intégrant des connaissances à priori, on peut améliorer l'analyse des images et leur interprétation. Ces approches partent du principe que les objets du monde réel sont toujours associés à leur contexte, et la représentation de ce contexte est essentiel pour l'analyse et la compréhension des images. Les connaissances contextuelles peuvent provenir de multiples sources d'information, et c'est la complexité de ces informations qui permet d'obtenir une description la plus fidèle possible. Dans un contexte particulier, l'introduction de ce type de connaissances aide à raisonner et à améliorer l'annotation d'images [Neumann 08, Hudelot 08, Bannour 11].

Lavrenko *et al.* [Lavrenko 03] ont été les premiers à utiliser l'information issue du contexte (contexte perceptif) dans l'annotation d'image. Ils ont proposé un modèle statistique génératif qui se penche sur la probabilité d'associer des mots avec les régions de l'image. Ce modèle calcule une probabilité conjointe qui associe des mots-clés aux régions des images. Ainsi, une image est décrite par l'ensemble de ses régions et de ses mots-clés. L'association de différentes régions fournit un contexte, tandis que l'association des mots avec les régions sur l'image donne un sens. Cependant, cette méthode ne capture pas vraiment la sémantique des images, mais utilise plutôt des statistiques sur le contexte pour améliorer la description de l'image.

Par conséquent, il est important d'utiliser des méthodes explicites et formelles pour représenter les connaissances. Cela permet de prendre en compte la connaissance associée au contexte général, au contexte spécifique, et d'améliorer l'interprétation de l'image, tout en s'affranchissant de l'implémentation utilisée. La connaissance issue du contexte général est considérée comme la connaissance du domaine et la connaissance issue du contexte spécifique est généralement vue comme la connaissance issue des images.

Sur la base de techniques en relation avec l'ingénierie des connaissances, une approche intéressante proposée par Newell [Newell 82] suggère, lors de la conception d'un système à base de connaissances, de faire une distinction entre le niveau "connaissance" et le niveau "symbole". « Le niveau connaissance est un niveau abstrait qui traite du contenu de la connaissance. A ce niveau, il s'agit de discuter du contenu de la connaissance qui doit être exprimée, et non de la manière dont la connaissance doit être codée dans un langage de représentation. ». « Le niveau symbole traite de la manipulation des symboles représentant la connaissance. Au niveau symbole, il s'agit de définir les structures de données informatiques capables de coder le contenu défini au niveau connaissance, connaissances qui sont exprimées dans l'ensemble des concepts définis au niveau symbole ». Le *niveau connaissance* de Newell peut être décliné en quatre catégories que sont : (i) les connaissances descriptives (relatives au domaine concerné), (ii) les connaissances opératoires (relatives aux opérateurs de traitement d'image), (iii) les connaissances stratégiques (re-

latives à la coordination des opérateurs) et enfin (iv) les connaissances système (relatives à l'implémentation "génie logiciel" d'une plate-forme d'interprétation).

Comme le souligne [Kompatsiaris 08], il existe 5 catégories de représentation des connaissances :

1. Les représentations basées sur les ontologies qui définissent des concepts de haut-niveau ;
2. Les représentations qui utilisent des méthodes d'apprentissage supervisées ou non supervisées pour associer des caractéristiques bas-niveau aux concepts haut-niveau ;
3. L'introduction de bouclage de pertinence dans la recherche pour de l'apprentissage incrémental des intentions des utilisateurs ;
4. Les modèles sémantiques (*semantic templates* en anglais) forment une interface entre les concepts de haut-niveau et les caractéristiques bas-niveau en proposant une image ou un icône ;
5. Les modèles utilisés dans les moteurs de recherche d'images du web, qui utilisent à la fois des informations textuelles et le contenu visuel des images.

De nombreux systèmes combinent plusieurs de ces techniques pour proposer des moteurs de recherche d'images à l'aide de mots-clés sémantiques.

Le problème de recherche d'images de documents anciens par leur contenu impose deux contraintes majeures. Premièrement, il faut concevoir un système qui puisse modéliser et structurer les connaissances issues du domaine des historiens (concepts sémantiques), et celles issues du domaine du traitement des images (caractéristiques bas-niveau). D'autre part, il faut que ce système puisse répondre au problème de fossé sémantique, et faire le pont entre les caractéristiques bas-niveau et les concepts sémantiques utilisés pour l'interrogation par les historiens.

Dans un autre contexte, on retrouve des problématiques équivalentes liées à l'avènement du numérique et au volume d'information multimédia ajouté et diffusé sur Internet qui connaît une croissance rapide (plus de 2 milliards de vidéos diffusées sur Youtube chaque jour sur la planète). Ce volume de données entraîne l'apparition de nouveaux challenges pour partager et retrouver l'information qu'il renferme, challenge que tente de relever le web sémantique[Uschold 96]. Le web sémantique repose sur l'utilisation de langages de description structurés (XML, RDF, ...), à la source des ontologies, qui permettent de modéliser les données, l'information qu'elles renferment, et l'information recherchée par les utilisateurs.

Parmi les représentations des connaissances du domaine existantes, les ontologies présentent de nombreux avantages. Elles fournissent un cadre formel et explicite d'une conceptualisation partagée [Gruber 95]. *"Formel"* puisque ces ontologies sont lisibles par la machine, et qu'il est possible de raisonner sur son contenu pour l'être humain et la machine. *"Explicite"* signifie que le type de concept utilisé, et les contraintes sur leur utilisation sont définies explicitement. L'idée de *"partagée"* renvoie aux connaissances qui appartiennent à une communauté. Enfin, *"conceptualisation"* fait référence aux modèles

obtenus par abstraction de certains phénomènes qui existent dans le monde réel, et par identification des concepts pertinents de ces phénomènes. Ainsi, les ontologies capturent la connaissance pertinente d'un domaine, fournissent une compréhension commune des connaissances de ce domaine, détermine le vocabulaire reconnu du domaine, donne la définition explicite de ce vocabulaire et des relations entre les termes du vocabulaire, et tout cela sous forme de modèles formels. Nous avons donc choisi d'utiliser ce formalisme pour notre approche. Nous présentons les détails de cette approche ci-dessous.

3.2 Approche proposée : Vue d'ensemble

Nous avons vu, depuis le début de ce manuscrit, que la recherche d'image est un domaine complexe. Celle-ci peut s'appuyer sur des mots-clés associés aux images, ou sur le contenu des images en lui-même (contenu déduit à partir des pixels qui composent l'image). Dans le cadre de cette thèse, nous nous sommes plus particulièrement intéressés aux images de documents issus du début de la Renaissance. Ces documents présentent des caractéristiques particulières qui rendent les outils classiques d'analyse d'image inefficaces.

Pour être précis, le but de cette thèse est de proposer une chaîne de traitements qui permettent l'analyse et l'indexation d'images de documents anciens. Cette chaîne de traitements adaptée aux documents anciens nécessite d'intégrer de multiples aspects liés aux connaissances des experts du domaine (les historiens dans le cadre de ce manuscrit), et aux connaissances des experts du traitement des images. Les connaissances des historiens permettent de décrire les images en fonction de leurs attentes (images composées de personnages, images purement décoratives, ...), tandis que les connaissances des traiteurs d'images permettent de séquencer la chaîne de traitements et surtout de la paramétrer de la manière la plus adaptée.

Afin de répondre à ce but, nous proposons un modèle en trois parties. Dans un premier temps, nous modélisons l'information recherchée par les historiens afin de proposer une approche en accord avec leurs connaissances et leurs attentes. Le but de ce modèle est de comprendre comment ils analysent ces images et ce qu'ils recherchent.

Dans une second temps, en partant des attentes des historiens et de l'étude des documents anciens (images complexes dégradées par le temps), nous proposons un modèle informatique d'analyse automatique des documents anciens. Celui-ci permet de prendre en compte les spécificités de ces documents. Nous présentons cette seconde approches dans la chapitre 5.

Enfin, ces deux modèles seront utilisés conjointement dans le dernier chapitre. L'utilisation conjointe de ces deux modèles permet de proposer des requêtes basées sur la connaissance de l'utilisateur tout en extrayant de la connaissance à partir des résultats de traitements d'images, et propose une solution pour réduire le fossé sémantique qui existe entre les traitements d'images et les attentes des historiens.

L'approche proposée est résumée dans la figure 3.2.1.

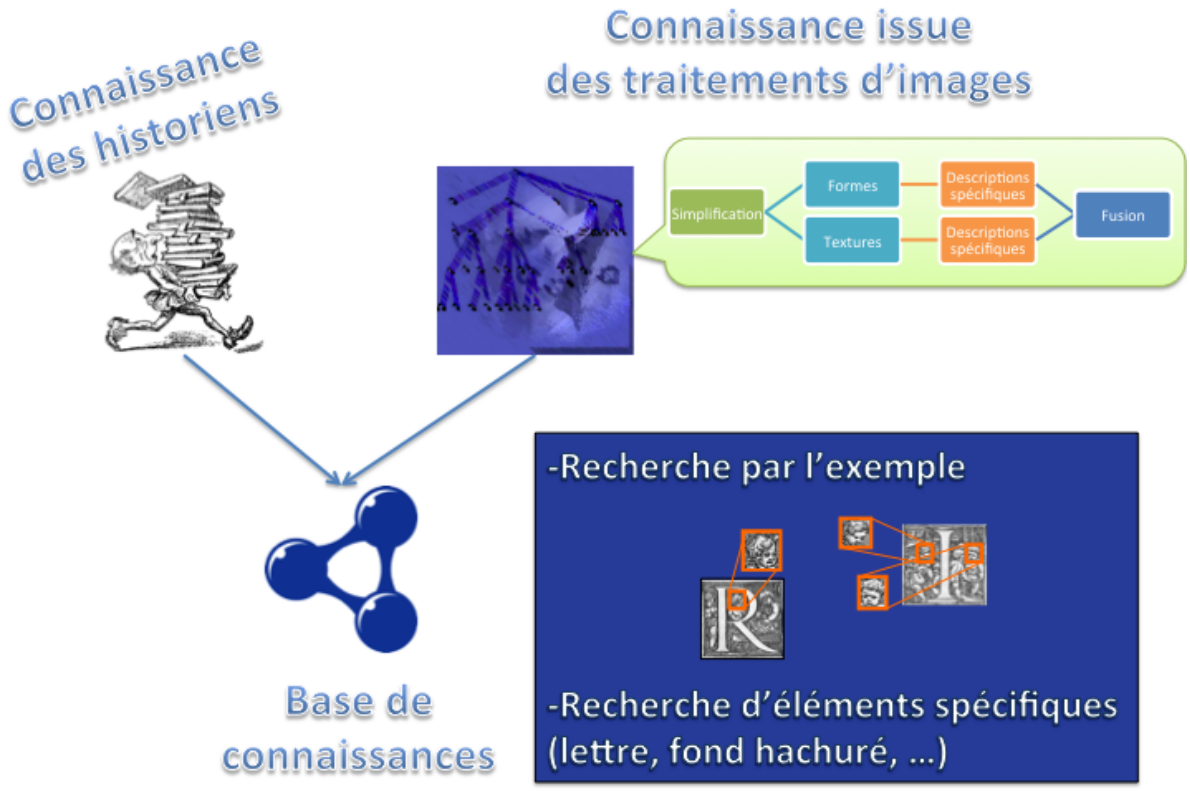


FIGURE 3.2.1: Boucle d'analyse automatique d'un document ancien

3.2.1 Modélisation Connaissance historiens

Dans le cadre du projet NaviDoMass, nous souhaitons proposer un moteur de recherche d'images par le contenu qui permet de répondre aux attentes spécifiques des historiens. Ceci impose donc d'avoir une connaissance sur leurs attentes. Afin de modéliser les connaissances des différents partenaires du Centre d'Études Supérieures de la Renaissance, nous avons échangé sur leur vision des documents anciens et avons défini un modèle basé sur une ontologie qui permet de recenser les éléments constituants des lettrines.

Ce modèle propose une structuration hiérarchique des concepts utilisés par les historiens pour décrire les images. Il permet également de rechercher dans cette base de connaissances pour retrouver les images qui partagent un certain nombre de concepts. Ce modèle est présenté en détails dans la section 4.

3.2.2 Cas des connaissances des informaticiens

Comme évoqué dans [Ogier 00], la conception d'un système d'interprétation d'images de documents anciens met en jeu des compétences très diverses et nombreuses, surtout si l'on se positionne du point de vue des connaissances. En effet, partant de l'image numérisée et des intentions de l'utilisateur, le problème est d'extraire de manière *intelligente* et *coordonnée* des données cohérentes au regard de l'expertise du domaine concerné, tout en conservant la propriété de généricité. Généralement, cet objectif est atteint en coordonnant des opérateurs de plus ou moins haut niveau (traitement des images, reconnaissance des formes, fusion de données, ...), guidés par une stratégie d'interprétation issue d'une requête initiale. Ce procédé fait également appel à des mécanismes de coopération entre l'utilisateur et le système.

Pour reprendre le modèle présenté par Newel [Newell 82], la connaissance issue des images est catégorisable en deux niveaux. Le niveau *"connaissance"* (connaissance métier), qui correspond aux connaissances des historiens dans le cadre de ce manuscrit, et le niveau "*symbolique*", associés aux connaissances issues des traitements d'images. Le niveau "*connaissance*"a été traité dans la section précédente, puisque la connaissance et les attentes des historiens ont été modélisées via une ontologie. Le second niveau, quant à lui, a pour but de trouver un formalisme informatique générique permettant d'intégrer toute la connaissance issue des traitements d'images.

Afin d'apporter une réponse à ce second niveau, nous proposons un modèle d'analyse de documents anciens qui permet d'extraire automatiquement de l'information à partir du contenu pixellaire. Notre modèle s'appuie sur le formalisme des ontologies et consiste à voir une image comme un ensemble de zones décrites par des caractéristiques qui leurs sont propres (dépendantes de leur contenu). Les informations liées à ces régions, réunies au sein d'une même base forment la connaissance liée aux traitements des images. Le formalisme retenu pour structurer l'information issue de ces traitements, a été choisi pour être inter-opérable avec les connaissances *"métiers"* (c'est à dire les connaissances des historiens). Ce modèle est présenté en détails dans la section 4.4.

Nous tenons à préciser qu'en aval de ce modèle, nous présenterons dans la partie III les traitements et algorithmes que nous avons définis et appliqués sur les images pour en extraire les régions et leurs descriptions (la figure 3.2.2 résume ces traitements). Ces traitements conditionnent, et sont conditionnés par le modèle ontologique. Nous avons fait le choix d'utiliser une chaîne de traitements adaptée à la recherche d'informations exprimées par les historiens.

Dans le cas des lettrines, images complexes composées d'au moins deux types d'informations, une première étape de simplification en couches d'informations homogène est nécessaire pour séparer les différentes catégories d'informations présentes. Puis, partant de chaque couche d'information, il convient d'extraire de l'information des images en proposant des outils adaptés aux spécificités de ces couches. Afin de conserver une

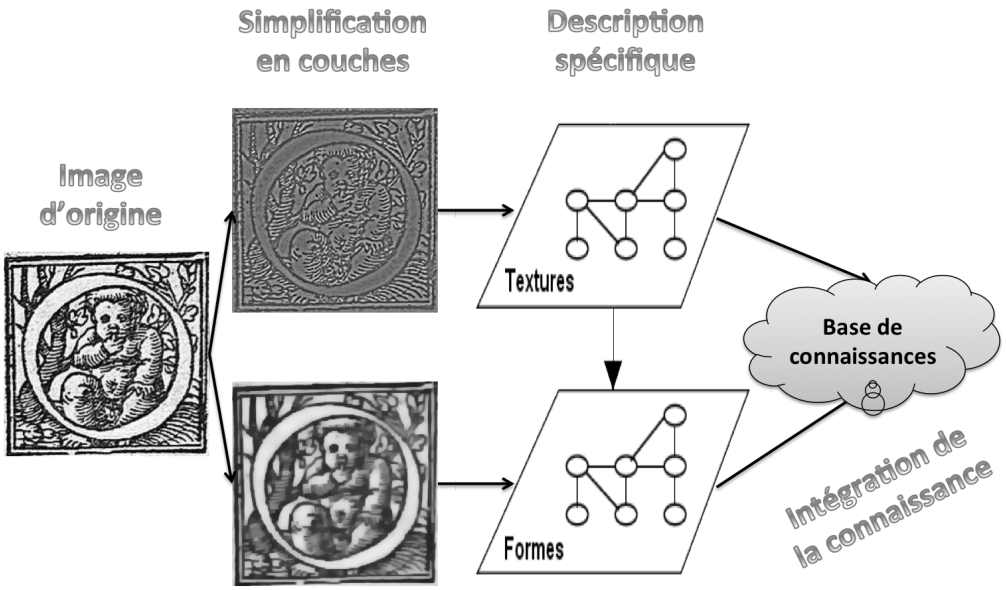


FIGURE 3.2.2: Schéma global de structuration de la connaissance des traiteurs d'images

certaine généricité dans la démarche nous permettant d'extraire de l'information, nous proposons l'enchaînement générique en trois étapes suivant :

1. Extraction de régions liées à un type d'information particulier : l'information qui nous intéresse dans les images peut être vue comme une recherche de zones d'intérêt. Comme l'enchaînement de ces étapes est réalisé sur des couches d'informations différentes, il devient nécessaire d'utiliser des approches spécialement adaptées à chaque couche retenue, et robustes aux dégradations des documents anciens ;
2. Caractérisation de la spécificité de ces régions : pour chaque couche, des zones d'intérêt sont extraites. Il convient de décrire celles-ci à l'aide d'un ensemble de caractéristiques adaptées. Le choix des caractéristiques repose principalement sur la volonté de choisir des caractéristiques de la littérature dans la mesure du possible, proches de la perception visuelle humaine pour conserver une cohérence avec l'information recherchée par les experts du domaine historien, et qui permettent de conserver une certaine généricité (pour pouvoir réutiliser celles-ci avec d'autres types de documents). Ces caractéristiques doivent également permettre de décrire le contenu des images, à la fois par rapport à leur contenu radiométrique, mais aussi par leurs propriétés spatiales. Le but final étant de pouvoir utiliser ces régions de l'image pour extraire des éléments sémantiques de l'image ;
3. Comparaison de ces régions : comme nous séparons l'image d'origine en plusieurs couches d'informations différentes, il est important de proposer une méthode de comparaison entre les régions. Celle-ci peut-être utilisée pour classer et/ou indexer les documents, et nous nous en servons également pour valider chaque étape de

l'approche globale.

3.2.3 Combinaison de nos modèles

Si l'on observe le problème de recherche d'image, on peut remarquer que les utilisateurs possèdent un vocabulaire différent de celui utilisé par les traiteurs d'images. Un algorithme de traitement d'image ne propose pas directement d'extraire un personnage ou une lettre, tandis qu'un historien ne recherche jamais des zones d'intérêt texturées ou uniformes. Ce problème de fossé sémantique se pose rapidement, et nécessite de créer un pont entre les pixels et les mots-clés des historiens.

L'intérêt majeur de modéliser les connaissances du domaine métier et du domaine symbolique via des ontologies, est de pouvoir les lier. Ce lien doit permettre de réduire le fossé sémantique, et ainsi permettre à l'utilisateur final de rechercher une image soit en s'appuyant uniquement sur des attributs de comparaison d'images (type moteurs CBIR classiques de la littérature), soit en ne s'appuyant que sur la connaissance métier, ou enfin de proposer une requête métier dont les résultats s'appuient sur la connaissance symbolique.

La combinaison de nos modèles permet de construire une base de connaissances qui contient à la fois les connaissances symboliques (traiteurs d'images) et les informations exprimées dans le niveau connaissances (dans notre cas celle des historiens). L'extraction de connaissances de cette base (réponse à une requête utilisateur) correspond alors à une règle d'inférence qui recherche l'information dans la source la plus adaptée.

Synthèse du chapitre

Cette deuxième partie du manuscrit est là pour présenter le formalisme de l'approche proposée. Ainsi, dans ce premier chapitre, nous commençons par une présentation de la problématique liée à la représentation des connaissances d'un domaine du point de vue des experts. Après un bref état des lieux des catégories de représentation des connaissances présentes dans la littérature, nous présentons le problème de représentation de connaissances issues des historiens et des traitements des images. Partant de cette description, nous justifions le choix retenu (à savoir, les ontologies), et nous présentons une vue d'ensemble de l'approche proposée. Celle-ci repose sur l'utilisation d'une base de connaissance qui rassemble les informations issues des deux domaines. Pour le domaine du traitement des images, nous présentons également le modèle d'analyse de documents anciens que nous proposons. Celui-ci est basé sur le paradigme de la complexité proposé par Edgar Morin, et il propose de simplifier les images en couches d'informations spécifiques (couche formes et couche textures), de décrire chacune de ces couches, et de les combiner pour obtenir un description complexe des images.

Chapitre 4

Modèle d'analyse de documents anciens à partir d'une ontologie

La numérisation de masse de documents du patrimoine a permis de constituer un gros volume de données. Ce volume représente plusieurs millions de pages numérisées non annotées, et donc difficiles à exploiter. Se dresse alors le challenge de décrire les images à l'aide d'outils d'extraction automatique d'information (outils de traitements des images), et de se servir de ces outils pour annoter automatiquement les images avec les mots-clés des historiens. Cette annotation permet d'organiser les documents (les indexer) pour permettre leur partage et leur recherche.

De nombreux algorithmes décomposent l'image en sous-parties, décrivent ces zones et leurs relations. Plusieurs approches ont été proposées dans la littérature comme nous avons pu le voir dans la partie précédente. Les historiens de leur côté, utilisent des thesaurus pour décrire le contenu des images. A partir de ces mots-clés, ils disposent d'outils leur permettant de structurer leurs bases d'images à partir d'index. Le problème majeur actuel consiste donc à mettre en relation la description d'une image telle qu'elle est perçue par les informaticiens, et la description de la même image vue par les historiens.

Nous avons donc retenu l'idée de modéliser la connaissance des historiens et celle issue des résultats de traitements d'images afin d'apporter une contribution au problème de fossé sémantique. Cette modélisation est effectuée à l'aide d'ontologies. La première repose sur le thesaurus et la connaissance des historiens. La seconde permet de représenter les informations obtenues par les traitements d'images. Dans notre cas, les documents sont décrits à l'aide d'informations visuelles, mais également à l'aide de l'agencement spatial entre les objets qui les composent. Enfin, une dernière ontologie permet de faire le lien entre les deux premières pour unifier les connaissances obtenues, et ainsi réduire le fossé sémantique qui existe entre les deux domaines. Nous présentons les bases des ontologies, ainsi que leur utilisation appliquée aux documents anciens dans ce chapitre.

4.1 Ontologies, outils et généralités

Dans le cadre de cette thèse, nous nous intéressons aux ontologies pour nous permettre de mettre en relation deux domaines. Ces domaines ne sont pas initialement liés et il est difficile de définir le lien qui les unit. Ainsi, si l'on regarde par analogie le problème de fossé sémantique entre les connaissances des utilisateurs et les connaissances des traiteurs d'images, on peut remarquer qu'il existe deux domaines de connaissance distincts qu'il serait bon de lier. Mais avant de présenter ces ontologies et comment nous proposons de réduire le fossé sémantique en utilisant les ontologies, nous présentons des rappels sur les connaissances principales des ontologies.

4.2 Logique de description et Ontologie

En philosophie, l'ontologie est « l'étude de l'être en tant qu'être, c'est-à-dire l'étude des propriétés générales de ce qui existe » [1]. L'étymologie renvoie à la « théorie de l'existence », c'est-à-dire la théorie qui tente d'expliquer les concepts qui existent dans le monde et comment ces concepts s'imbriquent et s'organisent pour donner du sens. Par analogie, le terme est repris en sciences de l'information, pour définir l'ensemble structuré des termes et concepts représentant le sens d'une source d'information. Les termes et concepts peuvent être vus comme des mots-clés de la source d'information (également appelée domaine de connaissance). L'ontologie est donc un ensemble structuré de concepts permettant de donner un sens aux informations.

Généralement, cette structuration des données s'appuie sur des logiques de description. À partir de cette structuration, des règles sont ajoutées pour contraindre la cohérence entre toutes les données, et enfin des règles d'inférence sont utilisées pour déduire des connaissances de nos données. Nous présentons ces grandes étapes dans la suite de ce chapitre.

4.2.1 Logique descriptive

Les logiques descriptives, également appelées *logiques de description*, sont une famille de langages issus de la logique du premier ordre. Les logiques de description permettent de représenter d'une manière formelle et structurée les connaissances d'un domaine particulier. Elles utilisent les notions de concept, de rôle et d'individu. Les concepts correspondent à des "classes d'éléments" et sont interprétés comme un ensemble dans un univers donné. Les rôles correspondent aux "liens entre les éléments" et sont interprétés comme des relations binaires sur un univers donné. Les individus correspondent aux éléments d'un univers donné. Les logiques descriptives décrivent deux types d'informations dans une base de connaissances [Gagnon 07] :

1. Définition tirée du site français de Wikipedia : http ://fr.wikipedia.org/wiki/Ontologie_(philosophie)

- Les informations terminologiques du domaine à l'aide de concepts. Cette première composante définit des notions basiques ou dérivées et comment elles sont reliées entre elles. Ces informations sont ***génériques*** ou ***globales***, vraies dans tous les modèles et pour tous les individus. Dans la littérature, cette composante est dénommée TBox ;
- Les informations sur les relations entre les individus, c'est à dire les instances des concepts. Cette composante définit des informations ***spécifiques*** ou ***locales***, vraies pour certains individus particuliers. Elle contient des assertions sur des individus (notées C(a) avec a l'individu), en spécifiant leur classe et leurs attributs ; et des assertions de rôles (notées R(a,b) avec a,b des individus) qui indiquent qu'il existe deux individus qui sont en relation par le rôle R. Dans la littérature, cette composante est appelée ABox.

Ces deux composantes sont les éléments principaux qui composent des logiques descriptives. Celles-ci sont également composées de règles qui permettent de déduire de l'information, c'est le mécanisme d'inférence lié à chaque logique de description. Les types d'inférences réalisées sur les logiques permettent de vérifier la cohérence dans les modèles, et les règles diffèrent entre la TBox et la ABox. Nous présentons ces règles :

- Les inférences réalisées sur la TBox permettent quatre types de contrôle [Baader 03] :
  - la ***satisfiabilité*** : qui revient à vérifier si pour chaque concept, il peut exister au moins un individu membre de cette classe. Par exemple, définissons une classe "homme", une classe "femme", une classe comme étant à la fois une sous-classe des classes "homme" et "femme" et spécifions que ces deux classes sont disjointes (c'est-à-dire qu'aucune entité ne peut à la fois être un homme et une femme), on se retrouve alors avec un concept inconsistant.
  - la ***subsomption***, qui consiste à déduire qu'une classe est une sous-classe d'une autre classe, même si cela n'est pas déclaré explicitement dans la base de connaissances. Par exemple si on spécifie que "humain" est une sous-classe de "animal", on peut déduire qu'une mère est un animal. De manière formelle, cela revient à vérifier que tout individu de la sous-classe est bien membre de la classe ;
  - l'***équivalence*** : deux concepts C et D sont équivalents, relativement à la TBox, si et seulement si D est une sous-classe de C et C est une sous-classe de D ;
  - la ***disjonction*** : deux concepts C et D sont disjoints relativement à la TBox si leur intersection est vide. De manière plus formelle, leur intersection est équivalente à l'ensemble vide.
- Les inférences réalisées sur la ABox, c'est à dire au niveau factuel, peuvent être vues en quatre problèmes [Baader 03] :
  - ***Cohérence*** : une ABox est cohérente par rapport à une TBox si l'ensemble des assertions est consistant. En d'autres termes, si tous les individus déclarés comme instance d'une classe peuvent réellement être une instance de cette classe, et similairement, si les relations déclarées entre deux individus sont réellement possible. Par exemple, si un individu est une mère, c'est-à-dire une instance de la classe mère. Comme la TBox spécifie qu'une mère doit avoir au moins un enfant,

la ABox devra contenir au moins un autre individu, et une relation indiquant qu'il est un de ses enfants ;
- ***Vérification d'instance*** : vérifier par inférence si une assertion C(a) d'un individu est vraie pour tout modèle d'une ABox et d'une TBox. Autrement dit, vérifier qu'une instance d'un concept peut avoir lieu ;
- ***Vérification de rôle*** : vérifier par inférence si une assertion R(a, b) est vraie pour tout modèle d'une ABox et d'une TBox. Autrement dit, vérifier que deux instances peuvent être mise en relation ;
- ***Problème de récupération*** : ce problème revient à retrouver le concept le plus spécifique que l'on peut associer à un individu.

4.2.2 Ontologie

L'ontologie est un ensemble structuré de concepts permettant de donner un sens aux informations. Les concepts sont organisés dans un graphe dont les relations peuvent être :
- des relations sémantiques ;
- des relations de composition et d'héritage (au sens objet).

L'objectif premier d'une ontologie est de modéliser un ensemble de connaissances dans une communauté donnée, qui peut être réel ou imaginaire. Voici un exemple basique d'ontologie qui illustre les relations d'héritage et de composition d'une ontologie :

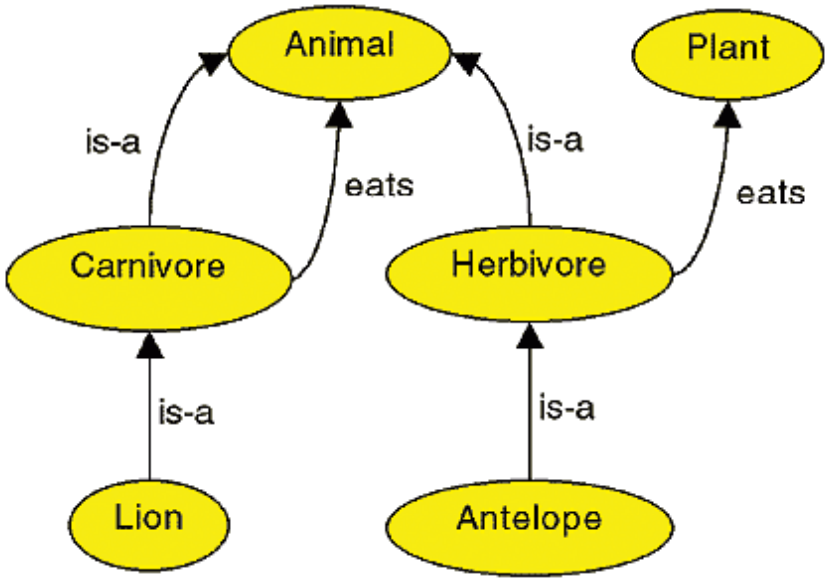


FIGURE 4.2.1: Exemple d'une ontologie sur les animaux. Elle illustre les relations de « sous-classe » et de « parent ». Les classes inférieures dans la hiérarchie représentent les relations subClassOf.

Les ontologies sont employées dans de nombreux domaines (intelligence artificielle, Web sémantique, génie logiciel, informatique biomédicale, architecture de l'information, ...) afin de représenter la connaissance d'un monde ou d'une certaine partie de ce monde. Les ontologies sont généralement composées :

- d'individus : les objets de base ;
- de classes : ensembles, collections, ou types d'objets ;
- d'attributs : propriétés, fonctionnalités, caractéristiques ou paramètres que les objets peuvent posséder et partager ;
- de relations : les liens que les objets peuvent avoir entre eux ;
- d'évènements : changements subis par des attributs ou des relations.

D'après Gruber [Gruber 93], cinq critères permettent de mettre en évidence des aspects importants d'une ontologie :

1. Clarté : chaque terme de l'ontologie doit être défini de manière claire et de manière non équivoque. Les définitions doivent être objectives, indépendantes du contexte et les plus complètes possibles (à l'aide de formalisme et de conditions nécessaires et suffisantes). Toutes les définitions doivent appartenir au langage naturel ;
2. Cohérence : rien qui ne puisse être inféré de l'ontologie ne doit entrer en contradiction avec les définitions des concepts, y compris celles qui sont exprimées en langage naturel. Une contradiction apparaît lorsque deux affirmations, idées, ou actions s'excluent mutuellement ;
3. Extensibilité : Les extensions qui pourront être ajoutées à l'ontologie doivent être anticipées. Il doit être possible d'ajouter de nouveaux concepts sans avoir à toucher aux fondations. Les fondations d'un ouvrage assurent la transmission et la répartition des charges (poids propre et surcharges) de l'ontologie. En d'autres termes, il doit être possible de définir de nouveaux termes pour des usages particuliers basés sur le vocabulaire existant sans avoir à modifier les définitions de l'existant ;
4. Déformation minimale due au langage : La création des concepts doit reposer sur les connaissances indépendamment du langage utilisé. Une déformation d'encodage a lieu lorsque le langage utilisé pour la création d'une ontologie influe la conceptualisation (un concept donné peut être plus simple à définir d'une certaine façon pour un langage d'ontologie donné, bien que cette définition ne corresponde pas exactement au sens initial) ;
5. Un engagement ontologique minimal : Une ontologie doit définir un vocabulaire pour décrire un domaine (si possible de manière complète). Contrairement aux bases de connaissances, une ontologie ne doit pas être en mesure de fournir systématiquement une réponse à une question arbitraire sur le domaine. Une ontologie correspond à la théorie la plus faible couvrant un domaine ; elle ne définit que les termes nécessaires pour partager la connaissance liée à ce domaine [2].

2. Le lecteur intéressé par des connaissances plus poussées sur l'engagement ontologique minimal pourra se référer à la notion de *upper-ontology* qui consiste à définir des concepts très généraux pour qu'ils restent vrais au travers de différents domaines

4.2.3 Les Descriptions d'ontologie

Les ontologies sont définies suivant un formalisme précis. Afin de proposer des implémentations de ce formalisme, plusieurs langages de description d'ontologie ont été proposés. Nous présentons les principaux ci-dessous.

4.2.3.1 Resource Description Framework (RDF/RDFS)

Resource Description Framework (RDF) est un modèle de graphe destiné à décrire de façon formelle les ressources Web et leurs métadonnées, de façon à permettre le traitement automatique de telles descriptions. Pour ce faire, RDF procède par une description de savoirs (données tout comme métadonnées) à l'aide de triplets, qui associe :

1. **Un sujet** qui représente la ressource à décrire ;
2. **Un prédicat** qui représente un type de propriété applicable à cette ressource ;
3. **Un objet** qui représente une donnée ou une autre ressource : c'est la valeur de la propriété.

Un document RDF ainsi formé correspond à un multigraphe orienté étiqueté. Chaque triplet correspond alors à un arc orienté dont le label est le prédicat, le nœud source est le sujet et le nœud cible est l'objet (voir figure pour un exemple de triplet). Dans un graphe, chaque triplet représente l'existence d'une relation entre les choses symbolisées par les noeuds qui sont joints. En annotant des documents non structurés et en servant d'interface pour des applications et des documents structurés (par exemple des bases de données, des systèmes de gestion électronique de documents, etc), RDF permet une certaine interopérabilité entre des applications échangeant de l'information non formalisée et non structurée sur le Web. Les documents RDF peuvent être écrits suivant différentes syntaxes.

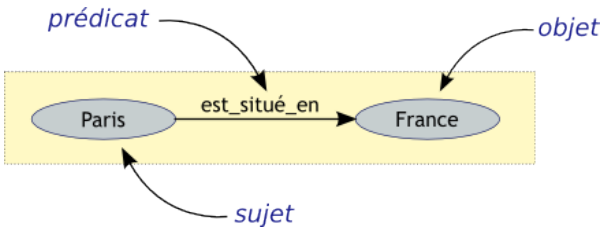


FIGURE 4.2.2: Exemple de triplet RDF

RDF Schema ou RDFS

est un langage extensible de représentation des connaissances. Il appartient à la famille des langages du Web sémantique publiés par le W3C. RDFS fournit des éléments de base pour la définition d'ontologies ou vocabulaires destinés à structurer des ressources RDF.

Il définit la manière dont sont stockées et décrites les données enregistrées au format RDF

RDF peut être représenté de diverses manières : la syntaxe RDF/XML ou la syntaxe Notation3 (ou N3) :

- **RDF/XML** est une syntaxe, définie par le W3C, pour exprimer un graphe RDF comme un document XML. Les descriptions RDF pourraient être représentées de façon assez naturelle grâce au formalisme XML adapté à RDF. La syntaxe RDF/XML permet de représenter des modèles très riches (mais les fichiers générés sont volumineux, même pour des descriptions assez simples), mais la définition des règles de raisonnement sur les taxonomies ne peut se faire que de façon trop limitée. De plus, cette syntaxe souffre de l'absence de la notion de modularité, de la portée globale plutôt que locale des propriétés, de l'impossibilité de déclarer des classes disjointes, des définitions de classes complexes à l'aide des expressions ensemblistes, et des restrictions de cardinalité ;
- **Notation 3 ou N3** : conçue par Tim Berners-Lee, N3 représente les descriptions RDF sous forme de suite de triplets et possède l'avantage d'avoir un formalisme RDF/XML plus concis. La syntaxe N3 offre l'avantage d'être plus facile à lire et à écrire que la syntaxe XML. Par ailleurs, des outils, comme CWM, existent pour convertir des fichiers N3 en fichiers RDF/XML.

RDF et RDFS permettent de définir, sous forme de graphes de triplets, des données ou des métadonnées. Ils permettent donc de mettre en place un schéma qui décrit les données. Cependant, de nombreuses limitations bornent la capacité d'expression des connaissances établies à l'aide de ces modèles. On peut citer, par exemple, l'impossibilité de raisonner ou de mener des raisonnements automatisés sur les modèles de connaissances établis à l'aide de RDF/RDFS. De même, aucune contrainte ne permet de garantir la cohésion globale du système puisque seules les données sont décrites dans un modèle RDF-DRFS. Afin de proposer une réponse à ce problème le langage OWL est apparu.

4.2.3.2 Web Ontology Language (OWL)

OWL est un langage XML fondé sur la syntaxe de RDF/XML proposé par le consortium W3C [McGuinness 04]. Contrairement au couple RDF/RDFS qui apporte à l'utilisateur la capacité de décrire des classes et des propriétés, OWL intègre, en plus, des outils de comparaison des propriétés et des classes (identité, équivalence, contraire, cardinalité, symétrie, transitivité, disjonction, etc). Ainsi, OWL offre aux machines une plus grande capacité d'interprétation du contenu Web que RDF et RDFS, grâce à des contraintes et des règles. Ces contraintes permettent de restreindre l'ajout d'instances dans le modèle. Prenons un exemple : soit les objets "*Humain*", "*Homme*" et "*Femme*". Le langage RDF/RDFS va permettre de définir que ces objets sont des classes (rdfs :Class), et que "*Homme*" et "*Femme*"sont des sous-classes de "*Humain*". Par contre, aucune contrainte

ne permet de préciser qu'une instance de la sous-classe "*Homme*" ne peut hériter de la sous-classe "*Femme*". C'est ce qu'apporte le langage OWL. En ajoutant des contraintes dans le modèle, on pourra spécifier qu'une instance de la classe "Humain" ne peut hériter qu'une fois (soit en Homme, soit en Femme).

Le langage OWL se décline en trois sous langages offrant des capacités d'expression croissantes et destinés à des communautés différentes d'utilisateurs : OWL-Lite, OWL-DL et OWL-Full [Lacot 05] :

- **OWL Lite** : est le sous langage de OWL le plus simple. Il est destiné aux utilisateurs qui ont besoin d'une hiérarchie de concepts simple. OWL Lite est adapté, par exemple, aux migrations rapides depuis d'anciens thesaurus [McGuinness 04] ;
- **OWL DL :** est plus complexe que OWL Lite, permettant une expressivité bien plus importante. OWL DL est fondé sur la logique descriptive (d'où son nom, OWL Description Logics), un domaine de recherche étudiant la logique, et conférant donc à OWL DL son adaptation au raisonnement automatisé. Malgré sa complexité relative face à OWL Lite, OWL-DL garantit la complétude des raisonnements (toutes les inférences sont calculables - toute demande d'extraction d'information amènera à un résultat) et leur décidabilité (leur calcul se fait en une durée finie) ;
- **OWL Full :** est la version la plus complexe d'OWL, mais également celle qui permet le plus haut niveau d'expressivité. OWL Full est destiné aux situations où il est plus important d'avoir un haut niveau de capacité de description, quitte à ne pas pouvoir garantir la complétude et la décidabilité des calculs liés à l'ontologie. OWL Full offre cependant des mécanismes intéressants, comme par exemple la possibilité d'étendre le vocabulaire par défaut de OWL.

Il existe entre ces trois sous langages une dépendance de nature hiérarchique : toute ontologie OWL Lite valide est également une ontologie OWL DL valide, et toute ontologie OWL DL valide est également une ontologie OWL Full valide.

4.2.4 L'inférence ou comment extraire de la connaissance de la base

Les logiques de description sont à la base des langages de description d'ontologies. Représenter des données à l'aide d'une logique de description permet donc de structurer des données et d'assurer une cohérence entre elles. Cependant, il n'est pas possible d'extraire de la connaissance de la "base de données" [3] directement à partir des différentes logiques descriptives. Pour répondre à ce problème, des moteurs et des règles d'inférences peuvent être proposés.

Afin de proposer des outils pour gérer la connaissance stockée dans ces ontologies, de très nombreux moteurs d'inférences ont été créés pour raisonner sur des logiques de description (Racer, Pellet, FaCT, FaCT++, Bossam, Hermit, ...). Ils permettent de raisonner sur des descriptions sémantiques ou sur des descriptions topologiques (relations spatiales entre objets). Nous présentons ci-dessous les principaux moteurs utilisés dans

3. Cette base est couramment appelée base de connaissance dans la littérature

ces deux approches.

Les règles d'inférences, quant à elles, permettent de raisonner sur la base de connaissance pour en extraire de nouvelles connaissances (implicites). L'objectif de règles est donc d'enrichir le champ des connaissances en les déduisant des informations présentes dans les concepts (inférence sémantique), ou des concepts eux-mêmes (inférence spatiale).

4.2.4.1 Inférence sémantique

L'inférence sémantique repose sur des algorithmes de résolution issus de la logique du premier ordre (clauses de Horn [René Cori 03]). Les mécanismes d'inférences reposent sur deux propriétés importantes :

- La subsomption, qui désigne une relation hiérarchique entre des concepts, dans les logiques de description. Cette notion est proche de la relation « est impliqué par » en logique classique, ou encore « contient » en logique ensembliste ;
- La satisfiabilité : qui doit assurer qu'une réponse puisse être apportée à toute requête.

Pour exprimer des requêtes à proposer aux moteurs d'inférences, différents langages de requête ont été proposés :

- **SWRL [Motik 05] :** pour *Semantic Web Rule Language* est un langage de règles pour le Web sémantique, combinant le langage OWL-DL et le langage RuleML - *Rule Markup Language (Unary/Binary Datalog.* SWRL est un langage qui enrichit la sémantique d'une ontologie définie en OWL, et qui permet de manipuler des instances de classes par des variables. SWRL ne permet pas de créer des concepts ni des relations, il permet simplement d'ajouter des relations suivant les valeurs des variables et la satisfaction de règles ;
- **SPARQL (protocole SPARQL et langage de requête RDF) [SPA 11] :** devenu une recommandation W3C le 15 janvier 2008, le langage SPARQL définit la syntaxe et la sémantique nécessaire à l'expression de requêtes sur une base de données de type RDF et la forme possible des résultats. SPARQL est adapté à la structure spécifique des graphes RDF, et s'appuie sur les triplets qui les constituent. Il s'inspire du langage SQL (langage de requête adapté aux bases de données relationnelles) pour la syntaxe et ses fonctionnalités. Le langage propose deux types de requête : soit interrogative pour extraire de l'information (extraction d'un sous-graphe), soit constructive pour ajouter de l'information (ajout d'un sous-graphe au graphe interrogé).

Afin d'inférer de la connaissance sur les ontologies, plusieurs moteurs d'inférences ont été créés pour manipuler de manière un peu plus aisée les langages de requêtes. Nous présentons les plus connus ci-dessous.

- **RACER [Haarslev 03] :** littéralement *Renamed ABox and Concept Expression Reasoner*, est un moteur d'inférence très répandu. Commercialisé par Racer Systems

GmbH & Co. KG, il a été fondé en 2004 par des chercheurs qui travaillaient à l'université de Hambourg. Ce moteur peut-être utilisé pour gérer des ontologies décrites en OWL ou en RDF. Ce moteur d'inférence possède également son propre langage de requête nRQL (new Racerpro Query Language) pour interroger les ontologies sur la ABox et la TBox ;

- **Pellet [Sirin 07] :** le moteur Pellet est plus récent. C'est un des projets du MIND-SWAP Group, un groupe de recherche sur le Web sémantique de l'université du Maryland. Il est disponible en OpenSource et offre des évolutions fréquentes. Pellet travaille sur des ontologies décrites en RDF ou OWL et permet les requêtes avec RDQL et SPARQL sur la ABox et la TBox. Écrit en Java, ce moteur est compatible OWL Lite, OWL DL et OWL Full. Il permet de raisonner sur les logiques de description ou sur les instances de concepts. Il y a deux façons différentes d'utiliser Pellet dans un programme de Jena : soit en utilisant directement l'interface Pellet (fortement recommandé), ou en utilisant Jena DIG interface (non recommandé). L'interface Pellet est plus efficace et fournit plus d'inférences ;
- **FACCT++ [Tsarkov 06] :** est un nouveau raisonneur en C++ conçu par des chercheurs de l'université de Manchester (Royaume-Uni). Il fournit une plate-forme d'expérimentation de nouveaux algorithmes d'aide à la décision et utilise un grand nombre d'optimisations améliorant les performances, notamment celles introduites dans le système précédent : FACT. Un algorithme de raisonnement à base de tableaux (semblables à ceux utilisés pour raisonner avec le langage d'ontologie OWL) peut ensuite traiter de manière répétée les entrées de la liste jusqu'à ce qu'elle soit vide. FACT++ est distribué sous licence publique GNU ;
- **DIG [Baader 03] :** l'interface DIG fournit un accès uniformisé aux raisonneurs sur des logiques de descriptions. L'interface définit un protocole simple (basé sur HTTP PUT / GET) avec un schéma XML qui décrit un langage de conception et les opérations d'accompagnement.

L'interface n'est pas conçue comme un cahier des charges mais fournit plutôt une liste minimale d'opérations qu'il faut proposer (par exemple, satisfiabilité, subsomption, ...). De nombreux moteurs d'inférences (CEL, FACT++, Pellet, RacerPro, ...) font partie de ce groupe, ainsi qu'un certain nombre d'éditeurs d'ontologies (Protégé, SWOOP, Jean, ...).

4.2.4.2 Inférence spatiale

Les mécanismes d'inférence sémantique permettent d'extraire de la connaissance du contenu des concepts et de leurs rôles. Dans le cas où une ontologie décrit des régions et les relations spatiales entre ces régions (comme des relations d'inclusions ou de superpositions), les mécanismes d'inférences sémantiques ne permettent pas d'encoder directement la connaissance existante entre des régions non directement connectées dans la base de connaissance. C'est pour cela que des mécanismes d'inférence spatiale ont été développés. Ils permettent de déduire de la connaissance sur les relations entre des

concepts non-immédiatement connectés. L'inférence spatiale propose donc des solutions pour déduire de la connaissance entre les différentes régions à partir de leurs relations.

Dans le cas d'images de documents, il est important de pouvoir identifier la position d'une forme dans l'image (par exemple la lettre d'une lettrine se trouve au centre de l'image). Afin de permettre cette analyse spatiale, nous avons mis en place des propriétés qui permettent de décrire les relations spatiales entre les formes et un quadrillage de l'image. Ces relations spatiales découlent de l'algèbre *Region Connection Calculus* (RCC-8). De manière analogue aux relations d'Allen permettant la définition des relations entre des intervalles temporels ou des spatiaux [Allen 83, Malki 02], le *Region Connection Calculus* est un formalisme logique destiné à la représentation des relations entre des régions spatiales. Ce formalisme a été proposé par Randell *et al.* [Randell 92], amélioré dans [Alboody 10], et permet de savoir comment deux régions spatiales [4] x et y sont connectées.

La figure 4.2.3 illustre un ensemble de huit relations entre deux régions x et y (RCC-8). Ces relations présentent l'avantage d'être exhaustives et disjointes deux à deux. L'algèbre RCC-8 permet dès lors de définir n'importe quelle relation spatiale entre deux régions, et il n'existe qu'une seule relation RCC-8 valide entre deux régions quelconques. Dans notre cas, nous avons étudié ces relations entre les formes extraites précédemment et les sous-parties de l'image présentées dans la figure 8.4.1. Elles sont présentées ci-dessous :

- *Region* **disconnectedFrom (DC)** *Part* : la région n'est pas connectée à la sous-partie ;
- *Region* **externallyConnectedTo (EC)** *Part* : la région touche tangentiellement la sous-partie ;
- *Region* **partiallyOverlaps (PO)** *Part* : la région recouvre partiellement la sous-partie ;
- *Region* **equalsTo (EQ)** *Part* : la région est égale à la sous-partie ;
- *Region* **tangentialProperPartOf (TPP)** *Part* : la région est incluse dans la sous-partie et la touche tangentiellement ;
- *Region* **nonTangentialProperPartOf (NTPP)** *Part* : la région est incluse dans la sous-partie ;
- *Region* **hasTangentialProperPart (TPPi)** *Part* : la sous-partie est incluse dans la région et la touche tangentiellement ;
- *Region* **hasNonTangentialProperPart (NTPPi)** *Part* : la sous-partie est incluse dans la région.

A partir de ces relations, et afin de faciliter le calcul des relations pour en déduire des connaissances, différents moteurs d'inférence spatiale ont été développés. Nous présentons les plus connus.

- **Geo-swrl [Kammersell 07] :** Geo-swrl est un raisonneur écrit en Java pour l'inférence spatiale. Il fournit les opérateurs de base de JTS (JTS Topology Suite est

4. Nous considérons ici qu'une région spatiale correspond à un ensemble de points dans un plan. Celle-ci est délimitée des autres régions spatiales par une frontière continue

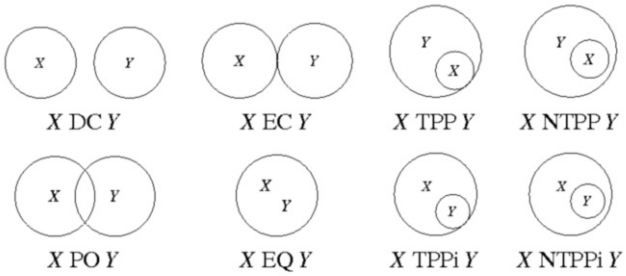


FIGURE 4.2.3: Relations spatiales utilisées pour observer l'agencement des formes dans l'image. Ces relations sont heritées de l'algèbre RCC-8.

une API 2D qui permet de manipuler des prédicats et des fonctions spatiales : `http://www.vividsolutions.com/jts/jtshome.htm` - The Geospatial web). Bien que Geo-swrl permette à l'utilisateur d'inclure des calculs de relations spatiales dans ses requêtes, sa capacité de représentation et de traitement des données, ainsi que sa documentation, restent pour le moment limitées.

- **Oracle Spatial [Kothuri 07] :** Oracle spatial permet une gestion simple et efficace des données de type géométrique. C'est une librairie de fonctions et de procédures qui permet d'accéder à une base de données Oracle, de l'analyser, et de la classer. Bien que conçu pour gérer des figures en 3 ou 4 dimensions, les fonctionnalités spatiales ne sont développées que pour les deux premières dimensions.
- **Pellet Spatial [Stocker] :** version étendue du raisonneur Pellet OWL, ce moteur de raisonnement permet de vérifier la consistance d'un ensemble de relations spatiales exprimées avec le formalisme *Region Connection Calculus* (RCC-8), d'en inférer de nouvelles à partir des relations fournies, ainsi que de répondre à des requêtes SPARQL combinant à la fois des relations RCC-8 et des relations RDF/OWL. Deux raisonneurs ont été implémentés dans PelletSpatial, le premier utilisant une traduction des relations RCC-8 en axiomes de classes OWL, et le second basé sur la table de composition RCC-8. Enfin, PelletSpatial ne peut pas déterminer la relation existante entre deux régions à partir de leurs coordonnées, mais nécessite que cette information soit contenue dans la ABox.

Nous venons de faire un rappel sur les bases des ontologies et les principales étapes de leur création. Nous allons maintenant présenter comment nous pouvons adapter cet outil à notre problématique.

4.3 Ontologie du domaine historien

Les ontologies reposent sur un formalisme qui permet de représenter, d'extraire et de déduire des connaissances issues du domaine historien. Dans le cas des lettrines, les

historiens s'intéressent à certains points particuliers, et souhaitent pouvoir annoter les images manuellement et naviguer dans cette base de connaissance. Nous présentons, dans cette section, le modèle d'analyse des lettrines du point de vue des historiens que nous proposons.

4.3.1 Définition et Description des lettrines par l'approche des historiens

Après différentes discussions et échanges avec les historiens, nous avons pu constater que les lettrines, du point de vue des historiens, peuvent être vues comme une superposition de quatre éléments [Jimenes 08]. La figure 4.3.1 illustre les différentes couches que l'on peut trouver dans une lettrine, et l'annexe B présente les formes que peuvent prendre ces couches en détails. Ces couches sont :

1. **La lettre** : Souvent située au centre de la lettrine, une lettre peut-être identifiée à plusieurs niveaux de détails. Il est possible d'identifier :
   a) la lettre elle-même (A, B, C, ...) ;
   b) l'alphabet utilisé (latin, grec, hébreu, ...) ;
   c) la couleur de la lettre (noire, blanche, ...) ;
   d) la police utilisée (romain, gothique, ...).
2. **Le motif** : Il correspond aux "objets" de l'image qui ont un but ornemental. Un motif peut soit représenter des scènes historiées (personnages, animaux, architecture, paysage, ...) et dans ce cas il sera figuratif, soit représenter des éléments récurrents (souvent des fleurs) et dans ce cas le motif sera dit décoratif ;
3. **Le fond** : Il peut-être noir, blanc, hachuré, criblé ou indéfini ;
4. **La cadre** : c'est la limite extérieure (le contour) de la lettrine. Il peut ne pas apparaître, ou être constitué de filets (1, 2 ou 3). Dans certains cas particuliers, le cadre est épais et peut-être vu comme un encadrement, ou un motif "troué" avec une lettre à l'intérieur (cas des passe-partout).

4.3.1.1 Les concepts et les propriétés associées

Partant de cette taxonomie, les historiens du Centre d'Etude Supérieur de la Renaissance (CESR) ont annoté manuellement un ensemble de 4288 images de lettrines. A partir des propriétés identifiées par les historiens, il est aisé de définir la *terminology box* (T-Box) d'une ontologie. Un concept a été construit pour chaque élément, et pour toutes ses variations possibles. La hiérarchie de classe obtenue est présentée dans l'image 4.3.2.

Une fois les éléments de base des lettrines identifiés, nous avons défini des classes pour attribuer un rôle à chaque concept. Les propriétés permettent d'assurer la cohérence de l'ontologie, pendant que les rôles permettent d'extraire de la connaissance. Voici la liste des rôles et des propriétés :

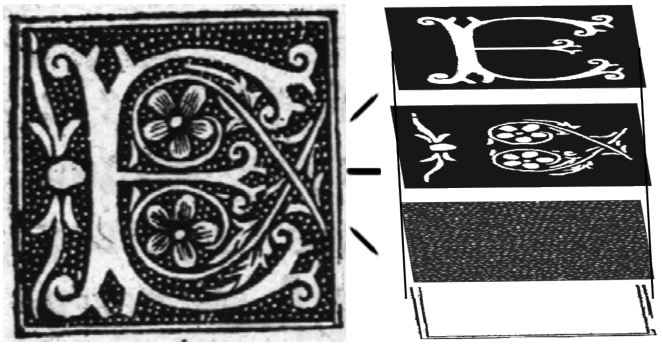


FIGURE 4.3.1: Superposition de couches d'informations dans les documents anciens : exemple d'une lettrine

- Lettrine **aUneLettre** : Indique si la lettrine comporte ou non une lettre ;
  - Lettre **aUneIdentification** Lettre : indique la lettre identifiée (lettre A, B, C, ...)
  - Lettre **aUneCouleur** : indique la couleur de la lettre de la lettrine
  - Lettre **aUnAlphabet** Alphabet : indique l'alphabet associé à la lettrine
  - Lettre **aUnePolice** Police : indique la police utilisée pour créer la lettrine
- Lettrine **aUnMotif** Motif : Indique si une lettrine a un motif
- Lettrine **aUnFond** Fond : Indique si une lettrine a un fond
- Lettrine **aUnCadre** Cadre : Indique si une lettrine a un cadre

4.3.1.2 Requêtes sur le modèle métier

L'annotation réalisée par les historiens nous a permis de peupler l'ontologie à l'aide d'assertions (peuplement de la A-Box plus précisément). Cette ontologie, base de connaissance du domaine historien, peut être utilisée pour classer les lettrines. Afin de vérifier cette possibilité, nous avons réalisé un comparatif entre des requêtes réalisées sur une base de donnée relationnelle classique, et des requêtes sur la base de connaissance. Les requêtes possibles, à partir d'une base de donnée relationnelle, sont de deux types possibles :

- Les requêtes complètes : tous les champs sont renseignés. Un exemple serait : « Tous les V Blancs, Latin, Romain, Hachuré, Figuratif avec un filet » ;
- Les requêtes incomplètes : où certains champs ne sont pas renseignés. Un exemple serait : « Toutes les L hachurés ».

Après avoir réalisé différentes requêtes sur la base de donnée relationnelle, et la base de connaissances, nous avons constaté que les résultats obtenus sont identiques. Ceci nous permet de vérifier la cohérence entre toutes nos données et la pertinence de nos modèles.

L'ontologie du domaine historien peut également être utilisée pour guider la recherche de lettrines à partir des descripteurs obtenus par traitement d'image. C'est en ce point que les ontologies deviennent intéressantes puisqu'elles permettent de relier deux infor-

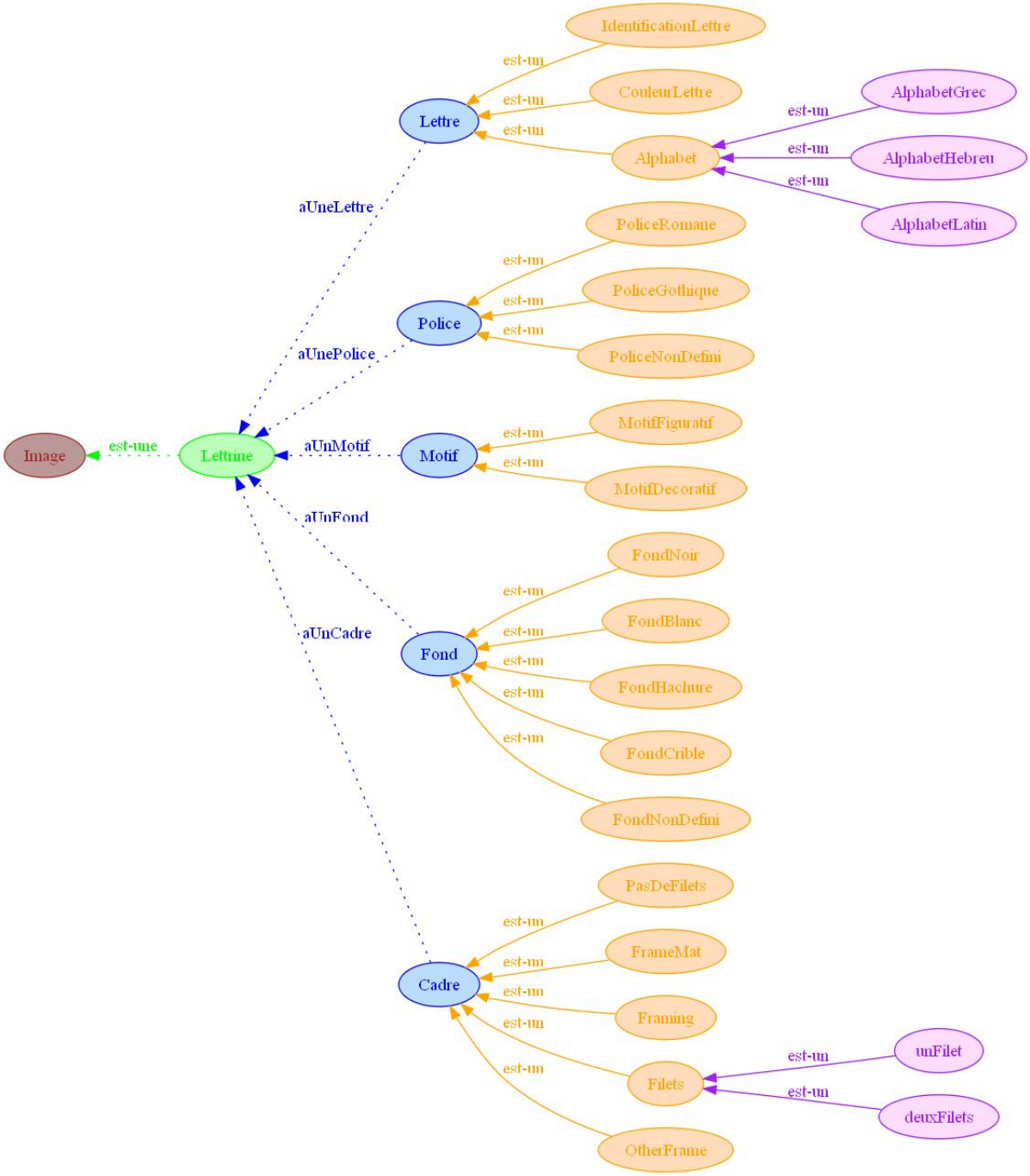


FIGURE 4.3.2: Hiérarchie de concepts utilisés dans l'ontologie du domaine Historien

mations issues de domaines différents. L'ontologie qui représente la connaissance des historiens permet de modéliser le sens des images, tandis que l'ontologie qui représente les résultats obtenus par les traitements d'image permet de décomposer les éléments constituant chaque image. Nous avons donc deux modèles qui représentent d'un côté les régions des images, et de l'autre le sens donné à ces régions. Dans la suite de cette partie, nous présentons l'ontologie associée aux traitements d'images, et son utilisation de manière liée à l'ontologie de la connaissance des historiens. C'est l'utilisation conjointe de ces deux ontologies qui nous permet de réduire le fossé sémantique entre connaissance métier et outils d'extraction d'information.

4.4 Ontologie des traitements d'images

Tout comme nous avons défini un modèle qui permet de représenter la connaissance des historiens sur les lettrines, nous proposons un modèle d'analyse et d'extraction de connaissances sur les lettrines basé sur des traitements d'images. L'intérêt de ce modèle repose sur sa capacité à hiérarchiser les connaissances issue des traitements d'images, et à en déduire de nouvelles implicitement. Nous le présentons dans cette section.

4.4.1 Lettrines, images complexes - approche informatique

Les images que nous traitons datent du début de l'imprimerie. Ces images étaient produites à l'aide de tampons en bois et composées de traits pour représenter les ombres et les perspectives. Ces traits forment des textures particulières, et les méthodes usuelles d'extraction et d'analyse de textures ne sont pas adaptées à ce type d'images.

Les images de documents anciens peuvent être vues comme des images complexes (composées de zones uniformes, de traits, ...). Pour reprendre un extrait d'Edgar Morin [Morin 96], le principe de simplicité impose de disjoindre, le principe de complexité enjoint de relier, tout en distinguant. Toutes les approches scientifiques jusqu'à la moitié du XXème siècle cherchaient à simplifier les problèmes compliqués pour les résoudre. Là où un problème compliqué va nécessiter une simplification sans chercher à respecter son environnement, un problème complexe nécessitera des allers-retours entre description certaine par simplification et description incertaine de son contexte.

Afin de prendre en compte cette complexité, nous proposons une approche qui consiste, contrairement aux approches cartésiennes, à simplifier une image pour la décrire et à recombiner ces différentes descriptions. Nous avons défini un modèle de traitement qui prend en compte la complexité des images de documents, et qui permet de gérer la connaissance qui en découle.

4.4.2 Modélisation de lettrines : concepts et propriétés de l'ontologie

Afin d'obtenir une base de connaissance qui rassemble les informations sur les régions extraites des images, nous avons défini une ontologie des traitements d'images. A l'intérieur de cette ontologie, les formes sont représentées par des concepts, et des propriétés leur sont associées pour les décrire. Comme nous souhaitons pouvoir analyser les relations spatiales des régions de l'image par rapport à un quadrillage de l'image, les concepts qui représentent les formes doivent permettre d'extraire de l'information sur l'organisation spatiale des concepts. C'est cette organisation qui permet de conserver la complexité et le contexte de la description de nos images (où se situe l'information ?).

De nombreux standards d'analyse de la topologie dans les images existent dans la littérature (grammaire 2D [Mas 10], graphes topologiques[Coustaty 11], graphes d'adjacence des régions [Bodic 09], ...), mais tous ne permettent pas de raisonner facilement sur l'organisation spatiale des régions par rapport à un quadrillage, tout en conservant la complexité. De plus, afin de lier cette connaissance à l'ontologie, nous avons choisi de nous appuyer sur le standard *Simple Feature Interface Standard* (SFS) développé par le consortium Open GIS [Goodchild 99] qui est directement intégrable dans les ontologies.

L'approche que nous proposons repose sur l'utilisation d'une ontologie puisqu'elle fournit un cadre formel et explicite de conceptualisation des traitements. Ce modèle est ainsi lisible et manipulable par les humains et les machines, et peut-être contraint en fonction du cas d'utilisation. Cette approche repose sur un processus en trois étapes :

1. Tout d'abord, les images sont simplifiées pour séparer les différents types d'informations qui la composent (identifier les différents éléments complexes qui composent l'image) ;
2. A partir de chaque image simplifiée, les zones d'intérêts (zones de formes, zones de traits, ...) sont segmentées pour pouvoir identifier chaque zone à retenir dans une image. Parmi l'ensemble des zones retenues, seules celles qui ont une surface significative sont conservées. Cela permet d'assurer que les régions peuvent être considérées comme prégnantes d'un point de vue perceptuel, et que l'information issue des descripteurs extraits dans l'étape suivante restent pertinente ;
3. Enfin, toutes les formes marquées précédemment sont décrites à l'aide de descripteurs adaptés à la typologie des zones extraites.

Partant des étapes du processus d'extraction d'information dans les images, nous proposons la hiérarchie de concepts suivant :

Image (de lettrine) : L'image de lettrine en elle même ;

ImageMeasurementUnit : Précise l'unité de mesure utilisée ;

ImageReferenceSystem : Référentiel utilisé dans le système pour indiquer les coordonnées des pixels/formes ;

ColorReferenceSystem : Le référentiel colorimétrique utilisé dans les traitements (Niveaux de gris, RGB, HSV, ...) ;

SurFace : Classe abstraite du standard SFS, derivée en :

ImagePartition : Image correspondant à la grille des sous-parties (voir figure 8.4.1) ;

Region : Chaque région extraite avec l'algorithme présenté ci-dessus ;

Centroid : Coordonnées du centre de gravité de chaque forme ;

RegionSet : ensemble des régions extraites d'une image ;

Pour rendre le calcul des relations spatiales entre les formes directement dans l'ontologie, les concepts *Region* et *ImagePartition* héritent du concept *SurFace* défini dans le standard SFS. Cette héritage permet l'utilisation des relations définies dans l'algèbre RCC-8 [Renz 02]. La hiérarchie de concepts obtenue à partir de cette modélisation est présentée dans la figure 4.4.1.

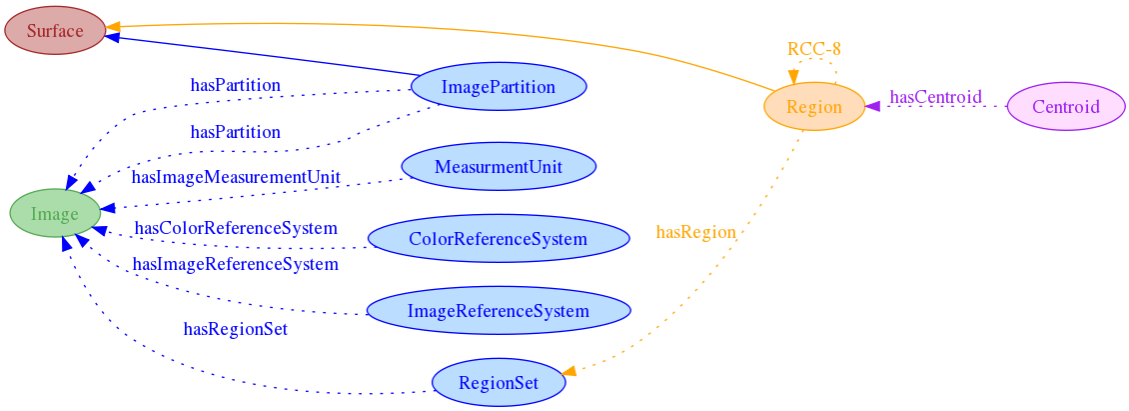


FIGURE 4.4.1: Hiérarchie de concepts utilisés dans l'ontologie du domaine traitement des images

Tout comme pour l'ontologie associée aux connaissances des historiens, des propriétés associées aux images, aux régions et aux relations spatiales ont été ajoutées pour enrichir l'ontologies. Les propriétés associées aux images sont :

- *Image* **hasImageMeasurementUnit** *MeasurementUnit* : unité utilisée dans les différentes mesures obtenues par les traitements d'images ;
- *Image* **hasImageReferenceSystem** *ImageReferenceSystem* : référentiel utilisé dans les traitements ;
- *Image* **hasColorReferenceSystem** *ColorReferenceSystem* : espace colorimétrique utilisé dans les traitements ;
- *Image* **hasRegionSet** *RegionSet* : récupère les informations sur l'ensemble des régions d'une image ;

- *Image* **hasPartition** *Partition* : récupère les informations sur le quadrillage défini précédemment ;
- *Image* **hasHeight**, et *Image* **hasWidth** renseigne les dimensions de l'image.

Afin de pouvoir récupérer l'information extraite automatiquement des formes, nous avons définis les propriétés suivantes :

- *RegionSet* **hasRegion** *Region* : récupère les informations sur les régions d'une image ;
- *Region* **hasCentroid** *Centroid* : récupère le centre de gravité de la forme est bien renseigné dans la base de connaissance ;
- *Region* **hasX** and *Region* **hasY** : récupère les coordonnées du centre de gravité ;
- *Region* **hasAire** : récupère l'aire de la forme ;
- *Region* **hasExcentricite** : récupère l'excentricité de la forme ;
- *Region* **hasGreyMean** et *Region* **hasGreySTD** : récupère la moyenne et l'écart-type des niveaux de gris de la forme ;
- *Region* **hasEuler** : récupère la valeur du nombre d'Euler de la forme.

Enfin, pour pouvoir peupler cette ontologie, nous avons développé une approche d'extraction de zones dans les lettrines. Les images de documents anciens étant des images complexes, nous nous sommes appuyés sur l'idée de complexité d'Edgar Morin [Morin 96] pour les décrire en respectant un cadre le plus générique possible. La figure 4.4.2 présente le schéma général de cette étape qui propose une simplification des images pour en extraire une connaissance adaptée. Une fois cette connaissance acquise, l'idée est de la re-complexifier en combinant les différents types d'informations extraits tout en tenant compte des liens qui les unissent (relations topologiques entre les zones).

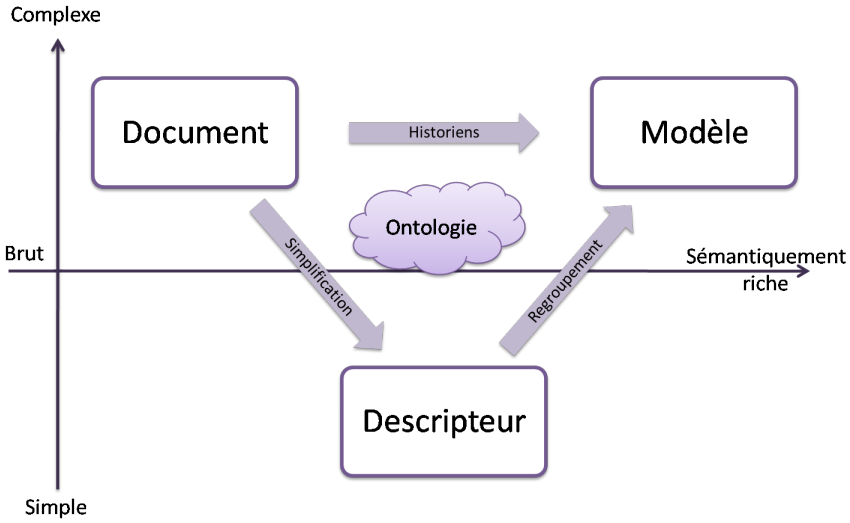


FIGURE 4.4.2: Description globale de l'analyse des documents anciens

Synthèse du chapitre

Le second chapitre de cette partie présente en détails notre modèle d'analyse de documents anciens à partir d'une ontologie. Après un rappel des outils et généralités liés aux ontologies, un rappel du cadre formel des ontologies a été rappelé. Ce cadre permet de présenté les différents formalismes de représentation des ontologies existants, leurs avantages et leurs inconvénients, et les différents mécanismes d'inférences (inférence sémantique / inférence spatiale) qui leurs sont associés. Partant de ce formalisme, nous présentons dans un premier temps les concepts et les propriétés associés à l'ontologie associée aux connaissances des historiens. Puis dans un second temps, nous présentons les concepts et les propriétés retenus pour décrire les connaissances complexes issues des traitements d'images.

Troisième partie

Réalisations pratiques et résultats

Chapitre 5

Étapes pour la réalisation du modèle d'analyse

Nous rappelons que l'approche globale que nous proposons dans ce manuscrit propose de modéliser, à l'aide d'ontologies, les connaissances des historiens, et celles des traitements d'images, pour constituer une base de connaissances. Cette base est alors utilisée pour réduire le fossé sémantique, et permettre aux historiens de réaliser des requêtes qui mêlent les connaissances des deux domaines.

Nous avons défini dans le chapitre précédent l'ontologie des traitements d'images de documents anciens, qui repose sur l'extraction de régions d'intérêt à partir de couches d'informations spécifiques, et de leur description à l'aide de caractéristiques adaptées. Cette idée repose sur l'idée d'analyse de la complexité définie par Edgar Morin [Morin 96], et nécessite dans un premier temps de simplifier l'image pour en extraire des couches d'informations spécialisées. Partant de chacune de ces couches, il devient nécessaire d'en extraire des régions d'intérêts et de les décrire.

Dans un soucis de généricité, nous proposons un processus générique en trois étapes applicable à chacune des couches : 1) Extraction de motifs adaptés à la couche ; 2) Définition d'une description pertinente liée au contenu ; 3) Extraction de régions d'intérêts dont on peut caractériser le contenu (description des éléments extraits) et leur agencement (contextualisation de ces éléments).

Nous présentons dans cette partie comment nous avons appliqué ce modèle aux lettrines, images graphiques du début de l'imprimerie. Ces images sont complexes et composées de zones de formes (qui représentent la lettre, les personnages, ...), et de zones de traits (caractéristiques de l'époque). Nous présentons donc, dans un premier temps, les spécificités des images. Puis dans un second temps, nous présentons un bref état des lieux des méthodes existantes pour simplifier des images, et nous développerons l'approche retenue : la décomposition de Meyer.

Cette décomposition nous permet d'obtenir deux couches d'informations, une ***couche forme*** et une ***couche texturée***, à partir desquelles nous venons extraire les formes et

les traits qui composent les lettrines. Chaque couche est alors traitée dans un chapitre, au travers duquel nous expliquons comment nous venons extraire des connaissances. Cette connaissance doit permettre de retrouver des éléments sémantiques recherchés par les historiens. Il nous a donc paru essentiel de ne plus travailler au niveau du pixel, mais au niveau de motifs de plus haut niveau qui représentent l'information élémentaire recherchée. Ainsi, la première étape consiste à extraire les motifs élémentaires de chaque couche d'information. Puis, partant de ces motifs, nous proposons de décrire de manière pertinente le contenu de l'image à l'aide d'une signature spécifique. Enfin, ces motifs sont également utilisés pour identifier des régions d'intérêts. Celles-ci sont utilisées d'une part pour définir une description structurelle de l'image, et d'autre part dans l'ontologie des traitements d'images.

Enfin, proposer un système de recherche d'images impose de savoir les comparer. Pour chaque couche d'information, nous proposons donc une méthode qui permette de les comparer, et d'évaluer la pertinence de la connaissance extraite. C'est sur cette base que repose l'ontologie finale qui permet à un utilisateur de rechercher une image à partir de son contenu.

5.1 Images de documents anciens - description des lettrines

Parmi les différents types d'images graphiques de documents anciens énoncés dans [Delalandre 09], ce travail s'intéresse plus particulièrement aux lettrines. Elles correspondent à des images très utilisées dans les ouvrages et très réutilisées au cours du temps. Une lettrine est une lettre ornementale qui débute un chapitre ou un paragraphe. Les images de documents anciens, et plus particulièrement les lettrines, possèdent des spécificités liées au début de l'imprimerie (XVème et XVIème siècles). Les approches classiques de la littérature caractérisent et indexent les images (naturelles ou de documents) à partir de signatures statistiques ou spatiales [Duda 00]. A partir de ces signatures, des propriétés fréquentielles et structurelles sont extraites pour décrire le contenu des images.

Dans le cas des images de documents anciens, qui possèdent des propriétés particulières (images binaires créées par l'homme, pour l'homme, et composées de traits - voir partie 2.2.2 pour plus de détails), ces outils sont inefficaces car inappropriés. En effet, la plupart des approches classiques utilisent la couleur, et ne sont pas robustes aux dégradations qui apparaissent dans les documents anciens (voir chapitre 2.2.1 pur plus de détails). Ces documents présentent la particularité d'être composés de formes (les personnages par exemple qui composent une image, la lettre dans les lettrines, ...) qui correspondent à des zones uniformes en niveau de gris, et de traits créés par les tampons en bois utilisés lors de l'impression (utilisés pour définir des zones d'ombrages, ou donner un effet de relief aux images). On peut donc tout de suite voir que de par la nature de ces documents,

deux types d'informations peuvent être étudiés : les zones uniformes (qui correspondent aux formes), et les zones texturées (qui correspondent aux traits).

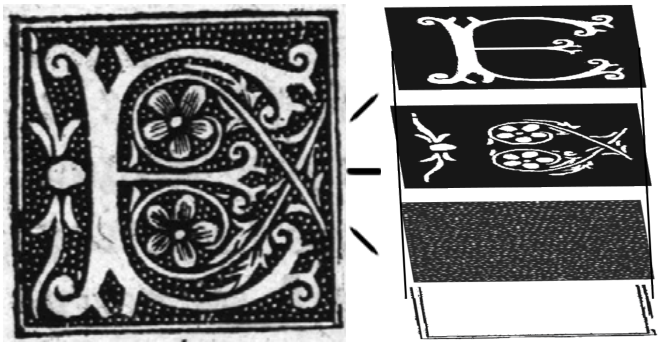


FIGURE 5.1.1: Superposition de couches d'informations dans les documents anciens : exemple d'une lettrine

Selon notre point de vue, celui des historiens, et comme évoqué dans le chapitre précédent, les images de documents anciens peuvent être vues comme une superposition de différentes couches d'information, chacune avec ses spécificités. Dans le cas des lettrines, on peut voir apparaître une superposition de trois couches insérées dans un cadre (voir figure5.1.1 pour un exemple). Ces couches sont :

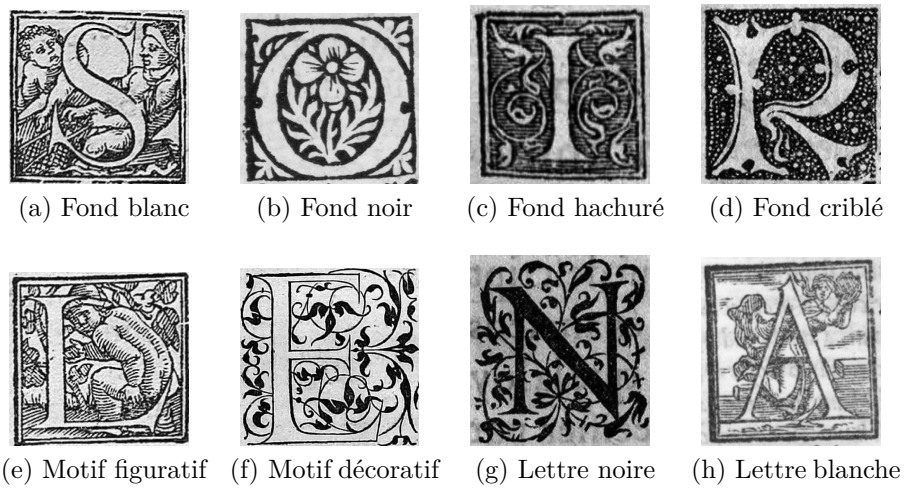
- Le fond : correspond à l'arrière plan et peut-être plein (noir ou blanc), hachuré ou criblé ;
- Le motif : correspond aux ornements. Il peut être décoratif ou figuratif ;
- La lettre : est un élément clef qui peut-être noir ou blanc et de différentes polices.

Enfin, le cadre peut-être composé d'aucun, un ou deux filets. Les couches et leurs caractéristiques sont détaillées dans l'annexe B.

Les images issues du début de l'imprimerie peuvent donc être vues comme des images composées de multiples couches d'informations. Chacune de ces couches possède des caractéristiques particulières que nous proposons de décrire pour notre processus d'indexation. La complexité de nos images repose sur un mélange de zones uniformes et de zones texturées. Simplifier ces images revient alors à séparer nos images selon deux étapes successives. Premièrement, séparer les images d'un point de vue fréquentiel pour séparer les zones lisses des zones texturées, puis dans un deuxième temps, séparer les images d'un point de vue objet pour extraire des zones d'intérêt.

L'approche proposée repose sur une ontologie qui permet de structurer et de hiérarchiser la connaissance extraite sur les images. Nous partons d'un modèle qui consiste à voir les images comme un ensemble de régions d'intérêts. L'ensemble de ces régions, munies chacune des informations qui leurs sont propres, permet de déduire de la connaissance sur l'image. Dans le cadre des images de documents anciens [1], qui sont des images

1. et plus particulièrement celui des lettrines



(a) Fond blanc (b) Fond noir (c) Fond hachuré (d) Fond criblé

(e) Motif figuratif (f) Motif décoratif (g) Lettre noire (h) Lettre blanche

FIGURE 5.1.2: Exemples de fond, de motif, et de couleur de lettre dans les lettrines

complexes, nous proposons un traitement en deux grandes étapes : tout d'abord une simplification de l'image en couches qui contiennent soit les régions homogènes, soit les régions texturées. Puis dans un second temps, nous utilisons chacune de ces couches pour extraire des zones d'intérêts. Ces zones sont alors décrites à l'aide de caractéristiques statistiques et de relations spatiales. La connaissance extraite de chaque couche est alors combinée avec celle de la couche voisine.

Ce chapitre présente l'ensemble des étapes que nous avons retenues et développées pour décrire les images. Afin de suivre le paradigme d'Edgar Morin [Morin 96], ces étapes proposent de simplifier les images pour en extraire une information la plus pertinente possible, puis de re-complexifier cette description en combinant les différentes informations extraites. La suite de ce chapitre se présente sous la forme suivante : tout d'abord nous présentons comment nous simplifions les images, pour séparer les informations liées aux formes des informations liées aux traits. Puis, dans une seconde partie, nous présentons la méthodologie retenue pour extraire les éléments de ces couches, et les caractériser. De par la nature des documents que nous étudions (images de documents créés par l'homme, pour l'homme, et avec des éléments sémantiques), nous avons fait le choix de ne pas travailler au niveau du pixel, mais d'extraire une information de plus haut niveau, et de la décrire. Ainsi, pour ce qui est de formes, nous travaillons avec des motifs de taille 3x3 [2] puisque nous recherchons des zones uniformes. Enfin, pour ce qui est des textures, ces images sont composées de traits caractéristiques de l'époque. Nous travaillons donc avec des traits, et nous les décrivons pour caractériser le contenu de chaque image.

Concernant la description des images, nous avons pu voir dans le chapitre **??** qu'il

2. le choix de la forme et de la taille du motif est expliquée par la suite

existe deux grandes catégories de description des images dans la littérature : celles qui décrivent statistiquement le contenu pixellaire des images, et celles qui décrivent la topologie des régions qui composent une image. Afin d'étudier le paradigme sur la complexité au maximum, nous décrivons les images suivant ces deux types d'approches. Enfin, une dernière section présente comment nous re-complexifions la description de nos images en combinant les différentes descriptions que nous présentons sur chaque couche.

Extraire de l'information et la décrire est une étape clé dans la description des images. Il nous a paru évident de valider nos choix, et nous avons choisi d'évaluer notre méthodologie, et les différentes étapes qui la composent, selon trois types de critères :

1. Une première évaluation porte sur la qualité de l'extraction de l'information qui compose chaque couche, en la confrontant à une vérité terrain. Cela nous permet ainsi de vérifier que l'information que nous retenons pour chaque couche est proche de ce que recherchent les utilisateurs finaux ;
2. Une seconde évaluation vise à vérifier que la description de l'information extraite est pertinente en utilisant des approches statistiques de types *Bag of Words*. Les approches classiques de la littérature, en recherche d'images par le contenu, visent à résumer les images en recherchant la fréquence d'apparition de mots-clés visuels dans les images. Ceci nous permet de vérifier que la description proposée reste pertinente pour comparer les images ;
3. Une dernière évaluation porte sur l'évaluation de l'intégration de l'information spatiale dans la description des images en utilisant des graphes. Les informations présentes dans les lettrines se trouvent à des endroits bien particuliers. Le lieu même de ces éléments est porteur d'information. Afin de mettre en avant cette idée, nous vérifions que les zones extraites permettent également de comparer les images.

Le schéma de la figure 5.1.3 présente les étapes de ce chapitre.

5.2 Simplification des images en couches

Nous avons pu voir dans la section 2.3.1 que les images de documents anciens sont des images complexes. Ceci est vérifié dans le cas des lettrines, puisqu'elles correspondent à des images qui mêlent différents types d'informations (des zones uniformes et des zones texturées).

La plupart des approches récentes de la littérature tentent de décrire les images (de scènes naturelles ou de documents) en utilisant des caractéristiques propres à un type d'information particulier [Torralba 08, Yang 08, Rusiñol 10, Jouili 10, Kisku 10] (le chapitre 1.4 présente un état de l'art de différentes caractéristiques adaptées à différents types de contenus). Les caractéristiques s'attardent donc à décrire un type de contenu

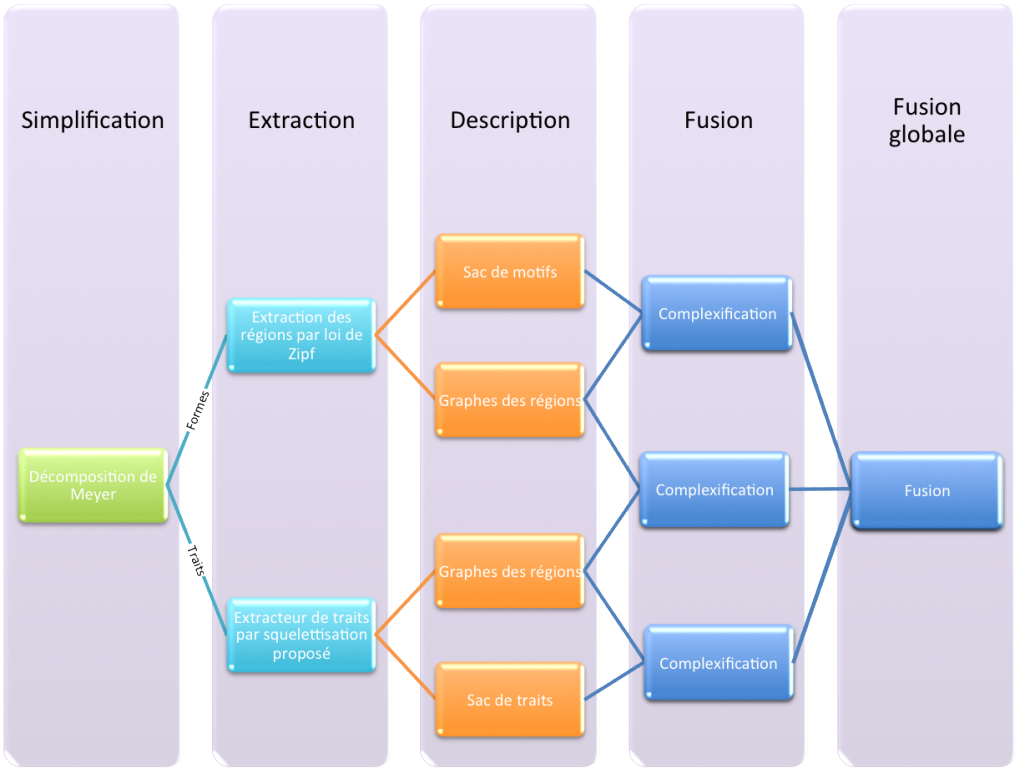


FIGURE 5.1.3: Étapes d'extraction et de description du modèle proposé

en particulier, et il ne parait pas cohérent de décrire une image composée de formes et de textures à l'aide de caractéristiques propres à un seul type d'information.

Afin de pouvoir étudier ces images, nous avons décidé d'utiliser le paradigme de la complexité défini par Edgar Morin. Celui-ci vise à simplifier les images pour pouvoir les décrire en utilisant des caractéristiques propres à chaque type d'information. La description finale de l'image repose alors sur la complexification des différentes descriptions utilisées. La simplification d'image doit permettre de segmenter les zones de l'image qui appartiennent à un type d'information particulier (les formes, les textures, ...). La simplification des images peut-être vue comme un processus de segmentation des images ou seuls les éléments respectant un critère particulier sont conservés. Dans la suite de cette section, nous présentons différentes approches existantes dans la littérature, les problèmes liés à ces approches, et nous présentons la méthode que nous avons retenue.

5.2.1 Différentes approches

La simplification d'une image consiste à séparer les informations qu'elle renferme en fonction de leur type. Les premiers outils de simplification ayant existés reposaient sur des processus de segmentation où l'on ne retenait alors que les zones qui respectaient un critère d'uniformité [Rosenberger 99]. Ces outils peuvent être vus comme des approches

locales, puisqu'elles reposent sur l'application d'une fenêtre glissante. Ces approches permettent de caractériser chaque type de région pour les segmenter par la suite.

Une seconde catégorie de méthodes, les approches globales, s'appliquent sur l'image entière et permettent de générer des *couches d'images*. Ces couches correspondent à une nouvelle image ne contenant que les pixels qui correspondent à une catégorie d'information. Ces approches correspondent généralement à des décomposition et permettent de séparer une image pour obtenir un couche par type d'information recherchée.

Nous présentons des méthodes associées à ces approches ci-dessous. La littérature est très riche dans ce domaine, et nous ne prétendons pas être exhaustif. Toutes les approches que nous présentons ci-après sont des méthodes de type descendantes, c'est à dire qu'elles cherchent à identifier la classe d'appartenance des différentes zones (zone uniforme ou zone texturée) en partant de l'image entière.

5.2.1.1 Approches locales

Les approches locales consistent à appliquer une fenêtre glissante sur l'image, permettant de définir si les éléments de la fenêtre sont localement uniformes ou texturés. L'attribution de la zone résulte du calcul d'un critère qui renseigne sur la complexité des éléments à l'intérieur de la fenêtre. De nombreux critères ont été proposés dans la littérature, et tous étudient l'information présente à l'intérieur de la fenêtre d'étude. L'un des plus répandus de la littérature a été proposé dans [Rosenberger 99], et il étudie le niveau de gris du pixel central par rapport à son voisinage. Une fois le critère calculé pour chaque pixel, une classe lui est associée (pixel de fond ou pixel de forme / pixel uniforme ou pixel texturé / ...).

Nous pouvons également citer les attributs qui permettent d'identifier le contenu des fenêtres à partir de la fréquence et de la répétition de motifs (attributs de texture classiques). Les méthodes les plus couramment utilisées reposent sur des attributs de cooccurrences [Haralick 73], des attributs de longueur de plage [Liao 94], ou des attributs issus d'histogrammes locaux[Haese-Coat 93].

Enfin, une dernière catégorie d'approche locale repose sur la recherche d'éléments saillants du point de vue de la perception visuelle humaine. Généralement connues sous le nom de cartes de saillances [Itti 98, Silva 10], ces méthodes consistent à caractériser chaque zone de l'image en fonction de l'attention que l'oeil humain pourrait lui accorder. Plus l'oeil humain consacre de temps à cette zone, plus elle est complexe et donc texturée.

Comparatif

Afin d'évaluer la pertinence de ces différentes méthodes faces aux documents anciens, nous proposons ci-dessous un tableau (5.1) qui permet de comparer les méthodes proposées ci-dessus (critères de complexité, attributs de fréquences de motifs, cartes de saillances) selon trois critères.

- Temps exécution : Ce premier critère sert d'indication dans l'optique de grandes masses de données. Les attributs de complexité proposent des implémentations rapides quand ils sont optimisés mais restent relativement lourds à calculer. Les attributs de fréquences sont malheureusement les plus chronophages puisqu'ils imposent de nombreux parcours des images selon différentes orientations ou fréquences. Enfin, les cartes de saillances, sont originellement pensées pour traiter des flux vidéos en temps réels. Elles proposent donc des implémentations en temps d'exécution rapides.
- Généricité : Nous entendons par généricité la capacité d'une méthode à s'adapter à tout type d'images ou de documents sans avoir à modifier ses paramètres via une interaction humaine. Les attributs de complexité ont été créés pour indiquer si un pixel appartient au fond ou à un objet, tout comme les cartes de saillances permettent de détecter les objets visuellement attractifs pour l'oeil humain. Ces deux catégories reposent généralement sur une modélisation de l'information qui figure à l'intérieur de l'image (fond/forme, zones lisses/zones texturées, ...). La méthodologie reste donc générique (filtrage de l'image par une fenêtre glissante généralement), mais il faut adapter les filtres en fonction des images que l'on souhaite traiter, et de l'information que l'on souhaite extraire. Les attributs de fréquences quant à eux dénombrent le nombre d'occurrences d'un motif dans une image. Ils ne nécessitent pas de paramétrage particulier et restent donc les attributs les plus performants face à une grande variabilité d'images ou de documents.
- Adaptation de la méthode aux documents anciens : Tous les attributs présentés n'ont pas directement été créés pour être utilisés sur des images de documents anciens. Comme présenté dans le critère généricité, ils sont adaptables pour la plupart aux documents anciens mais nécessitent une connaissance de la part d'un expert sur l'outil et sur les images pour proposer des adaptations de ces méthodes à la topologie des documents anciens (recherche d'éléments caractéristiques des documents anciens, résistances aux variations dues au temps, traits plus ou moins larges, ...).

En conclusion, nous pouvons dire que ces méthodes ne sont pas adaptées aux documents anciens puisqu'elles sont orientées vers l'extraction d'informations propres aux images naturelles. Ces images sont généralement décrites en utilisant les couleurs et les orientations principales qui les composent.

5.2.1.2 Approches globales

Par opposition aux approches locales, les approches globales proposent de séparer les différentes zones de l'image en filtrant celles-ci à l'aide de critères particuliers généralement basés sur une recherche d'homogénéité ou de discontinuités [Kermad 95]. Ces critères cherchent à mettre en avant des zones particulières (zones uniformes, zones texturées, ...).

Un premier critère repose sur la fréquence d'apparition des motifs composant l'image. Développé dans [Pareti 06a], tous les motifs de taille 3x3 sont extraits et ordonnés selon

	Attributs du complexité	Attributs de fréquence	Carte de saillance
Temps exécution			
Généricité			
Adapté aux documents anciens			

TABLE 5.1: Tableau comparatif des méthodes présentées

leur fréquence d'apparition dans l'image. Une courbe de Zipf est alors calculée à partir des ces fréquences, et une approximation de la courbe résultante à l'aide de trois fonctions affines permet de séparer les pixels de l'image en trois classes (une par fonction). Du point de vue perceptuel humain, la première partie de la courbe correspond aux motifs les plus fréquents, tandis que les deux fonctions restantes permettent de faire ressortir les pixels appartenant aux contours et aux zones peu fréquentes.

Enfin, le dernier type d'approche repose sur la décomposition d'images en différentes composantes homogènes. Ce dernier type d'approche est un domaine de recherche très actif ces 20 dernières années. Les méthodes les plus connues sont la décomposition de Wold[Francos 93] qui permet de décomposer des images texturées (séparation en deux parties : un champ déterministe et un champ stochastique). On peut également citer les travaux de [Meyer 01, Vese 04, Starck 05, Aujol 06a, Hamidi 10] qui permettent de séparer une image en deux composantes : une partie uniforme (grandes zones lisses de l'image d'un point de vue signal), et une partie texturée. De manière plus généralisée, des méthodes de séparation de sources, comme l'analyse en composantes indépendantes [Davies 04, Koldovsky 06] et l'analyse en composantes morphologiques [Starck 04, Fadili 10] proposent de décomposer un signal. L'idée consiste alors à projeter les signaux (des images dans le cadre de ce manuscrit) dans des espaces particuliers. Chacun de ces espaces pourra être vu comme un filtre permettant de ne retenir qu'un type d'information.

Dans le cadre de nos travaux, nous avons préféré nous orienter vers une méthode définie par [Meyer 01, Aujol 05, Aujol 06a, Dubois 08, Hamidi 10] car elle permet de séparer l'image en trois couches : les zones lisses, les zones texturées, et le bruit. Ainsi, il nous est possible de séparer les zones uniformes des images de documents, des zones de traits, tout en réduisant l'impact du vieillissement sur les images. De manière conforme aux attentes des utilisateurs, les zones lisses correspondent aux formes recherchées (person-

nages, lettres, fleurs, ...), tandis que les zones de traits permettent de caractériser l'arrière plan des images. Nous présentons cette méthode en détails dans la section suivante.

5.2.1.3 L'approche utilisée - Décomposition de Meyer

Les images sont très souvent composées d'informations de natures diverses et variées. Ainsi, dans une même image, on pourra retrouver des zones avec peu de variations de niveaux de gris (des zones lisses), et des zones plus oscillantes (comme les textures ou le bruit). De nombreux travaux récents en traitements des images proposent de décomposer une image en couches, chaque couche se limitant à un type d'information (couches homogènes). Les images de documents anciens sont composées de zones lisses (les formes, comme la lettre ou les personnages présents dans les lettrines), de zones texturées (les zones de traits), et de bruit dû aux dégradations des documents dans le temps. Il parait donc important de séparer ces informations dans des couches pour les étudier séparément.

Plusieurs approches ont proposé de décomposer une image en couche. Ainsi, Sloven Dubois présente dans sa thèse [Dubois 10] différentes méthodes, comme l'analyse en composantes morphologiques, la décomposition de Meyer, ou les ondelettes. Surapong *et al.* [Uttama 05] ont quant à eux proposé d'utiliser la décomposition de Wold, pour séparer les zones lisses des zones texturées des images. Nous avons orienté notre choix sur une décomposition de Meyer, puisque celle-ci présente l'avantage d'être générique (applicable à différents types d'images, dont celles de documents), n'impose pas d'avoir à choisir les ondelettes mères à utiliser dans la décomposition, et propose une couche bruit qui récupère toute l'information qui n'a pu être récupérée dans les couches formes et textures.

Partant de l'observation d'une image dégradée f, le but de la décomposition de Meyer est de retrouver une image idéale u, dégradée par une matrice de bruit v et rendue floue par une matrice A. Le bruit est souvent considéré comme additif, et la matrice A est souvent considéré comme un opérateur linéaire de déformation de l'image (opérateur de flou). Cette décomposition peut s'écrire de manière générale sous la forme :

$$f = Au + v \tag{5.2.1}$$

Ce modèle de décomposition est un problème inverse généralement mal posé. Une approche classique consiste à introduire un terme de régularisation qui induit une solution unique pour le système. Ce problème de minimisation a déjà été appliqué à travers de nombreuses implémentations, comme la restauration, la déconvolution, le débruitage, la segmentation ou l'inpainting d'images.

Les premières études ont porté sur la restauration et le débruitage d'image. Mais d'après l'idée d'Yves Meyer [Meyer 01], et dans le cadre de l'implémentation de la minimisation de la variation totale définie par Rudin *et al.* [L.Rudin 92], la décomposition d'images en une couche géométrique et une couche oscillante (texture) apparaît être

une approche utile et intéressante pour l'analyse d'images. Dans le cadre de l'analyse d'images de documents anciens, nous cherchons à séparer les zones uniformes indépendamment des zones texturées, tout en s'abstenant des problèmes d'acquisition comme le bruit.

Les images de documents anciens sont complexes et riches en terme d'information. Elles ont besoin d'être simplifiées. Ces images sont principalement composées de traits, qui rendent les méthodes classiques d'extraction de textures inopérantes. Partant de l'idée de Meyer et Rudin, nous nous sommes basés sur l'approche similaire de Dubois et Lugiez [Dubois 08, Hamidi 10] qui repose sur une série de projection. Celle-ci sépare l'image originale en différentes couches d'informations. La décomposition retenue consiste à décomposer l'information contenue dans l'image en une couche géométrie (les zones lisses de l'image), et une couche texture (les zones oscillantes de l'image).

Couche géométrie

La première couche d'information u est définie dans $BV(\Omega)$ qui correspond au sous-espace des fonctions $u \in L1(\Omega)$ de telle sorte que la quantité suivante, appelée variation totale de u, soit finie :

$$J(u) = sup\left\{\int_{\Omega} u(x) div(\xi(x)) dx\right\} \tag{5.2.2}$$

Cette première couche contient la structure de l'image, c'est à dire les grandes zones géométriques et uniformes de l'image.

Couche texture

Cette seconde couche v est définie sur G, sous-espace introduit par Meyer pour détecter les motifs oscillants (les textures et le bruit). Ce sous-espace, $J*$, est composé des fonctions f qui peuvent s'écrire sous la forme :

$$f = div(g) \tag{5.2.3}$$

c'est à dire qu'il correspond à l'espace des divergences de champs de vecteurs. Dans le cas d'images, cet espace est limité à deux dimensions, et l'on obtient alors :

$$div(g) = \frac{\delta g_1}{\delta_1} + \frac{\delta g_2}{\delta_2} \tag{5.2.4}$$

g_1 et g_2 sont définis dans $L^{\infty}(\Omega)$. Avec $\|g\|_{L^{\infty}} = sup_{t\in[a,b]}|g(t)|$, la norme suivante peut être considérée sur G :

$$\|v\|_G = inf\left\{\|g\|_{L^{\infty}(\Omega,R^2)} | v = div(g)\right\} \tag{5.2.5}$$

De manière intuitive, cette seconde couche contient les zones oscillantes de l'image,

c'est à dire la texture. L'opérateur divergence peut-être vu comme une mesure de la variation des niveaux de gris. Si une zone est uniforme, sa mesure sera nulle. Elle ne sera donc pas conservée dans cette couche.

Couche bruit

Enfin, la dernière couche obtenue par cette décomposition permet de mettre en exergue le bruit, c'est à dire les zones oscillantes très rapidement dans l'image. Cette couche w est définie sur E, espace dual qui permet de modéliser les motifs oscillants très rapidement. Soit $\dot{B}^1_{1,1}$ l'espace homogène de Besov usuel, E correspond alors à l'espace dual associé à $\dot{B}^1_{1,1}$, appelé espace de Banach :

$$E = \dot{B}^{\infty}_{-1,\infty} \tag{5.2.6}$$

La décomposition de Meyer est un processus itératif qui permet d'extraire les détails d'une image. Le signal de l'image est projeté dans des espaces particuliers (J, $J*$ and B) pour conserver uniquement les parties intéressantes. Une partie résiduel du signal :

$$\frac{1}{2\alpha}\|f-u-v-w\|_{L^2} \tag{5.2.7}$$

est ajoutée au calcul de la minimisation de la fonctionnelle pour récupérer tous les éléments qui ne sont pas récupérés au cours des projections dans les espaces. Contrairement aux approches par transformée en ondelettes, les contours sont préservés au cours de la projection dans l'espace des zones géométriques. Ceci est important pour la suite de nos traitements puisqu'il nous permet de conserver la séparation nette des formes (lettre, visages, membres des personnages). Cette séparation nette est primordiale pour l'extraction de ces formes à l'aide d'un algorithme d'extraction de composantes connexes. Le lecteur intéressé par l'étude théorique de cette décomposition et ses espaces associés pourra se référer aux articles suivants : [Aujol 05, Aujol 06a, Aujol 06b].

Ce modèle de décomposition repose sur la minimisation de la fonctionnelle F suivante :

$$\inf_{(u,v,w)\in X^3}(F(u,v,w)) = J(u) + J*\left(\frac{v}{\mu}\right) + B\left(\frac{w}{\lambda}\right) + \frac{1}{2\alpha}\|f-u-v-w\|_{L^2} \tag{5.2.8}$$

avec X l'espace euclidien $\mathbb{R}^{N*M}$ avec $N*M$ les dimensions de l'image ; J, la projection dans l'espace de variation totale, associée à l'extraction de la composante géométrique de l'image (également appelée couche forme de l'image) ; $J*$ la projection dans l'espace de Meyer, associée à l'extraction de la composante oscillante (appelée couche texture) ; et B, la projection dans l'espace de Besov associée à l'extraction de la composante bruit (appelée couche bruit). Les paramètres α, μ et λ permettent de contrôler l'importance que l'on veut donner à chacune de ces projections.

Afin de minimiser cette fonctionnelle, nous avons utilisé l'algorithme de projection de

Algorithme 5.1 Algorithme de décomposition d'images en niveaux de gris

A l'étape n :

1. u et v ont été préalablement calculés, nous estimons :

$$\widetilde{w} = P_{\delta B_E}(f - u - v)$$

2. ensuite, nous calculons :

$$\widetilde{v} = P_{\mu B_G}(f - u - w)$$

3. et finalement nous obtenons :

$$\widetilde{u} = f - u - \widetilde{v} - \widetilde{w} - P_{\lambda B_G}(f - v - w)$$

Cette opération est répétée jusqu'à ce que :

$$max(|\widetilde{u} - u|, |\widetilde{v} - v|, |\widetilde{w} - w|) < \varepsilon$$

Aujol et Chambolle [Aujol 05]. La projection P telle que définit par Chambolle dans l'espace λBG [3]est notée : $P_{\lambda B_G}(f)$. Cette projection est calculée par un algorithme itératif. Cet algorithme commence avec $P^0 = 0$, et pour chaque pixel (i, j) et à chaque étape $n + 1$, on obtient :

$$P_{i,j}^{n+1} = \frac{P_{i,j}^n + \tau \left(\triangle div(P^n) - \frac{f}{\lambda} \right)_{i,j}}{1 + \tau |\triangle div(P^n) - \frac{f}{\lambda}|_{i,j}} \tag{5.2.9}$$

D'après [Chambolle 05], une condition suffisante pour assurer la convergence de l'algorithme est : $\tau \leq \frac{1}{8}$. Pour résoudre la fonctionnelle de l'équation 5.2.8, les auteurs proposent l'algorithme 5.1. Dans le cadre de cet algorithme, la partie résiduelle de la fonctionnelle est incluse dans la couche bruit.

Dans [Aujol 05], les auteurs ont remplacé $P_{\delta B_E}(f - u - v)$ par $f - u - v - WST(f - u - v, \delta)$ où $WST(f - u - v, \delta)$ correspond à un seuillage doux en ondelettes de $f - u - v$ en utilisant le seuil δ tel que :

$$WST(f - u - v, \delta) = \begin{cases} d_i^j - \delta * sign(d_i^j) & si \quad |d_i^j| > \delta \\ 0 & si \quad |d_i^j| \leq \delta \end{cases} \tag{5.2.10}$$

avec d_i^j les valeurs des coefficients d'ondelettes, j la résolution, et $i \in \{x, y, xy\}$.

Dans le cadre de ce manuscrit, nous avons utilisé les paramètres par défaut de la décomposition de Meyer tels que définis dans [Aujol 05, Dubois 08].

Couches obtenues et interprétations

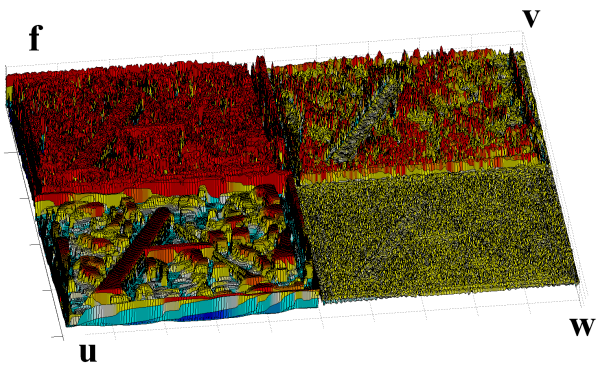
En nous appuyant sur l'implémentation de l'algorithme de Aujol et Chambolle proposé dans [Dubois 08], nous obtenons les trois couches de la décomposition de Meyer. Ces trois couches, dans le cas d'images de documents anciens, peuvent être vues comme :

3. $\lambda B_G = \{f \in G | \|f\|_G < \lambda\}$

- Une couche forme, qui correspond aux zones de l'image avec de faibles variations de niveaux de gris. Cette couche permet de mettre en évidence la géométrie de l'image, c'est à dire les zones de l'image qui correspondent aux formes ;
- La couche oscillante, qui correspond aux éléments répétitifs de l'image. Dans notre cas, cette couche retient les textures (zones de traits) ;
- La couche des oscillations rapides, qui correspond au bruit dans l'image. En fait, cette couche contient tous les éléments qui n'ont pas été retenus dans les deux premières. Cette couche est constituée du bruit, du texte appartenant au verso, et des dégradations dues aux effets du temps. Comme le but de cette thèse est d'indexer les images de documents anciens, tout en étant robuste à différents bruits, nous n'utiliserons pas cette couche par la suite. Elle ne sera utilisée que comme prétraitement.



(a)



(b)

FIGURE 5.2.1: Ensemble des couches obtenues avec la décomposition de Meyer. La sous-figure (b) présente les niveaux de gris de l'image de la sous-figure (a) en 3 dimensions (le niveau de gris correspond à la troisième dimension). La couche forme (U) apparaît en bas à gauche, tandis que la couche texture (V) apparaît en haut à droite. On voit clairement que les informations importantes des lettrines sont séparées au cours de cette décomposition

La figure 5.2.1 présente un exemple de résultat obtenu sur une lettrine. La lettrine est présentée dans son état d'origine, pendant que les différentes couches sont présentées sous forme de courbes en 3 dimensions. La représentation des niveaux de gris en trois dimensions permet de mettre en avant la séparation des zones oscillantes et des zones lisses. Des exemples de couches obtenues en 2D sur es lettrines sont également présentés dans la figure 5.2.2.

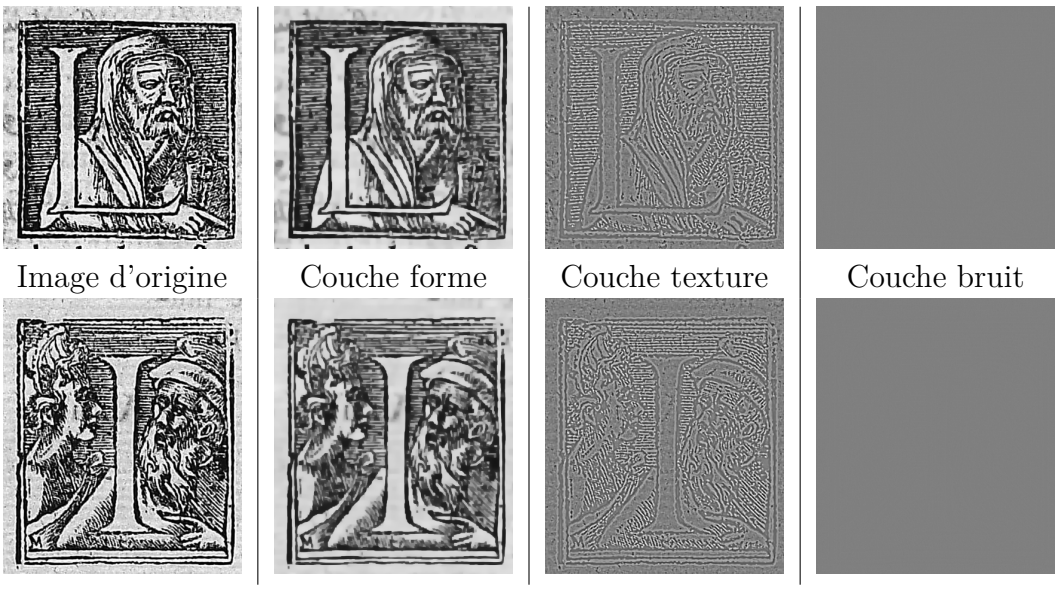


FIGURE 5.2.2: Exemples de couches obtenues sur des lettrines via la décomposition de Meyer

Synthèse du chapitre

Cette troisième partie présente en détails les étapes de notre modèle (extraction de motifs, caractérisation spécifique, extraction de régions). Nous commençons donc par une analyse des images de lettrines pour identifier leurs caractéristiques du point de vue du traiteur d'images. Ces images sont complexes puisque composées de plusieurs couches d'informations (formes et traits). Nous présentons donc, dans la seconde partie de ce chapitre, comment simplifier en couches ces images pour isoler chaque type d'information qu'elle renferme. Après une description des catégories de simplification existantes, nous présentons en détails l'approche retenue : la décomposition de Meyer. Celle-ci présente l'avantage de débruiter l'image, tout en extrayant des couches d'informations conformes aux attentes des utilisateurs.

À partir de cette décomposition, nous obtenons deux images simplifiées. La couche forme contient tous les pixels liés aux zones uniformes des lettrines (la lettre, les personnages, ...), tandis que la couche texture contient tous les pixels liés aux traits. À partir

de cette décomposition, nous proposons d'extraire les zones correspondantes à la lettre ou aux personnages de la couche forme, tandis que les zones de traits seront extraites de la couche texture. Nous allons maintenant présenter les méthodes utilisées pour extraire des zones de ces images et les décrire pour enrichir la base de connaissance.

Chapitre 6

Cas de la couche forme

6.1 Extraction d'information à partir des formes de l'image

Maintenant que l'image originale a été décomposée, nous allons nous intéresser plus particulièrement à l'image obtenue dans la couche forme. Toutes les variations rapides de niveaux de gris ont été lissées et elle n'est plus composée que de zones relativement uniformes.

Si l'on observe des images de documents anciens (voir figure 5.1.2 pour des exemples), on peut remarquer que les zones de formes correspondent à une lettre, des objets décoratifs (des fleurs par exemple), ou des personnages. Extraire les régions des images de documents anciens revient à identifier les zones lisses de l'image. Cependant, ces éléments peuvent apparaître en blanc ou en noir, et ont subi les effets du temps. Cette identification impose donc également d'être invariant aux dégradations existantes, et à la couleur des formes recherchées. Extraire ces formes n'est donc pas une chose facile, et les méthodes traditionnelles d'extraction de formes peuvent être de trois catégories :

1. Soit elles partent d'une première partition de l'image, qui est ensuite modifiée en divisant ou regroupant des régions, et on parle alors de méthodes de type décomposition/fusion (ou split and merge en anglais) ;
2. Soit elles partent de quelques régions, qui sont amenées à croître par incorporation de pixels jusqu'à ce que toute l'image soit couverte, et on parle alors de méthodes par croissance de régions ;
3. Des méthodes fondées sur la modélisation statistique conjointe de la régularité des régions et des niveaux de gris de chaque région existent également.

Les algorithmes de type décomposition/fusion exploitent les caractéristiques propres de chaque région (surface, intensité lumineuse, colorimétrie, texture, etc.). On cherche des couples de régions candidates à une fusion en fonction du critère de similarité et de leur agencement spatial. On fusionne alors les couples de régions avec une valeur commune,

et on réitère jusqu'à ce que les caractéristiques de l'image remplissent une condition prédéfinie (nombre de régions, luminosité, contraste, texture générale donnée, ...). On peut par exemple citer les algorithmes d'extraction de composantes connexes dans les images[Haralick 92].

Les algorithmes par croissance de régions partent d'un premier ensemble de régions, qui peuvent être calculées automatiquement (par exemple, les minima de l'image), ou fournies par un utilisateur de manière interactive. Les régions grandissent ensuite par incorporation des pixels les plus similaires suivant un critère donné, tel que la différence entre le niveau de gris du pixel considéré et le niveau de gris moyen de la région. Les algorithmes de segmentation par ligne de partage des eaux [Gonzalez 09], développés dans le cadre de la morphologie mathématique, appartiennent à cette catégorie, tout comme les algorithmes de remplissage par diffusion [Treuenfels 94], ou encore de croissance de régions [Mehnert 97].

Les algorithmes fondés sur une modélisation statistique conjointe des régions et des niveaux de gris, notamment ceux s'appuyant sur les Champs de Markov Cachés, reposent sur la minimisation d'une fonction de vraisemblance (ou énergie). Cette fonction prend simultanément en compte la vraisemblance de l'appartenance du pixel à une région considérant son niveau de gris, et les régions auxquelles appartiennent les pixels voisins. Cette fonction effectue un compromis entre la fidélité à l'image initiale et la régularité des régions segmentées.

Toutes ces méthodes ne permettent pas de répondre aux problèmes des images de documents anciens puisqu'elles imposent soit de travailler sur des images binaires, soit d'avoir une connaissance a priori sur l'information recherchée (la couleur de la lettre, définir les germes, ...).

On peut remarquer que les formes importantes des documents anciens correspondent à de grandes zones uniformes. Une grande zone uniforme correspond à une zone composée plusieurs fois du même motif. On peut donc supposer qu'une analyse fréquentielle des motifs qui composent ces formes nous permette d'extraire ces formes. Plusieurs approches de la littérature ont porté sur l'analyse fréquentielle du contenu des images (auto-corrélation [Journet 08c], analyse des motifs qui composent l'image [Pareti 08], décomposition de Wold [Surapong 08], ...). Parmi toutes ces approches, nous nous sommes particulièrement intéressés à celle qui repose sur une loi de Zipf. Celle-ci extrait les motifs qui composent une image (donc moins sensibles aux variations qui touchent directement le pixel) et permet de ne retenir que les zones composées des motifs les plus fréquents, sans connaissance à priori sur l'image (elle ne nécessite pas de connaître la couleur de la lettre ou des formes qui composent l'image, et toutes les formes n'ont pas besoin d'être de la même couleur). Nous présentons l'extraction des motifs, et comment nous obtenons la loi de Zipf à partir des images ci-dessous.

6.1.1 Extraction des motifs de l'image - Loi de Zipf

Comme mentionné précédemment, la simplification en utilisant une décomposition de Meyer nous permet d'obtenir une couche forme. Cette couche est alors considérée pour extraire les formes de l'image [1]. Extraire les formes d'une image consiste à identifier dans une image tous les pixels qui partagent une information commune, et à les annoter avec une étiquette commune. Travailler au niveau du pixel pose plusieurs problèmes difficiles (sensibilité au bruit, non-prise en compte du contexte du pixel, ...).

Afin d'avoir une méthode robuste aux bruits (bruits dus aux dégradations liées au temps, à l'usure des documents, et à l'acquisition de ces documents), aux changements d'apparence (objets blancs sur fond noir, et vice-versa), et qui tienne compte du contexte de chaque pixel (pixel d'une forme, d'un contour, ...), nous avons choisi une approche qui travaille au niveau de motifs qui composent l'image. Plus exactement, nous nous intéressons à la fréquence de ces motifs. En effet, des zones homogènes présentes un grand nombre de motifs redondants, sans avoir à définir au préalable les motifs que nous recherchons.

Les méthodes utilisées dans l'analyse de textes visent à extraire des séquences de caractères qui, une fois assemblés, forment des mots de différentes longueurs. La fréquence de ces motifs (sous-ensemble de lettres) est donc utilisée pour caractériser le document. Chaque document est résumé par le sous-ensemble des motifs les plus fréquents qui le composent. En analyse d'images, de nombreux travaux ont porté sur la recherche de mots visuels caractéristiques des images. Ces approches (dénommées Bag of Words dans la littérature anglophone), extraient des mots visuels dans les images et les indexent en énumérant leur fréquence d'apparition [Deselaers 08]. Ces mots visuels peuvent correspondre à des motifs pré-définis, ou une autre approche proposée par Pareti *et al.* [Pareti 06a] consiste à caractériser une image combinant une loi de Zipf avec les motifs de l'image. Les motifs correspondent à l'ensemble des masques de taille $n*m$ [2], et l'image est ainsi décrite par le calcul de la fréquence de ses motifs.

Une approche par une loi de Zipf présente l'avantage de ne pas rechercher des motifs particuliers et de ne retenir que l'information importante dans une image ou une base d'images. Nous avons donc fait le choix de retenir cette approche pour décrire nos images de formes.

6.1.1.1 Loi de Zipf

La loi de Zipf [Zipf 49] est une loi empirique formulée il y a 60 ans qui repose sur une loi puissance. Si l'on possède des phénomènes décrits par un ensemble de motifs organisés topologiquement (notés $M_1\, M_2\, \ldots\, M_n$), et de leurs fréquences d'apparitions (notés $N_1\, N_2\, \ldots\, N_n$), cette loi infère que la distribution des fréquences d'apparitions des

1. par forme, nous entendons les grandes zones lisses de l'image comme la lettre ou des membres des personnages présents dans les lettrines.

2. $n$et m étant des paramètres de l'approche

motifs ordonnée par ordre de fréquence, est corrélée au rang de ces motifs. Cette loi peut s'écrire sous la forme :

$$N_\sigma(i) = k * i^a \tag{6.1.1}$$

$N_\sigma(i)$ représente le nombre d'occurrences du motif apparaissant au $i_{ème}$ rang, k et a étant des constantes. Cette loi puissance est caractérisée par la valeur de l'exposant a ; k étant plus étroitement lié à la longueur de la séquence de symbole étudiée. La relation de corrélation entre le rang et la fréquence des motifs n'est pas linéaire, mais une simple transformation conduit à une relation linéaire entre le logarithme de N et le logarithme du rang. Si l'on représente de manière ordonnée l'ensemble des motifs sur une courbe, on obtient une courbe de Zipf. La figure 6.1.1 présente la courbe associée au livre « Ulysses » de James Joyce. Dans cet exemple, les motifs utilisés sont les mots du texte, et chaque point correspond à la fréquence d'apparition de ce mot dans le texte.

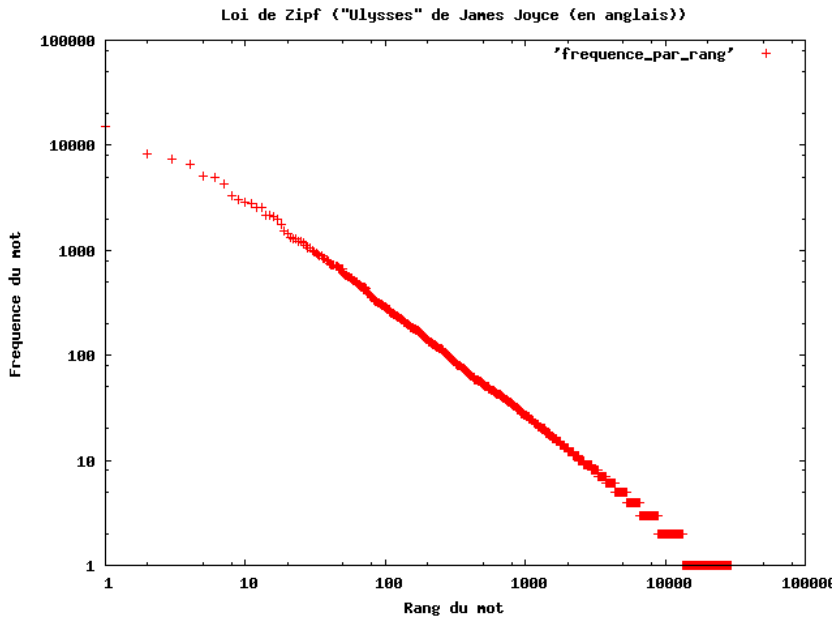


FIGURE 6.1.1: Graphique log/log de la fréquence d'apparition des mots par rapport à leur rang dans "Ulysses" de James Joyce (image issue de wikimedia)

La valeur de l'exposant a peut être facilement estimée à partir du coefficient directeur de la droite correspondant à la régression linéaire de la courbe 2D : $(\log(i), \log(N_\sigma(i)))$, avec $i \in [1; n]$. Comme nous pouvons le constater dans l'exemple présenté dans la figure 6.1.1, cette courbe peut être estimée par une fonction linéaire.

6.1.1.2 Extraction des motifs dans les images

L'analyse de données avec une loi de Zipf en 1 dimension se limite à l'analyse de l'ensemble des symboles successifs. Par exemple, l'analyse d'un texte consiste à analyser la fréquence des mots qui le composent (les mots étant les motifs du texte). L'enchainement de ces symboles forme des motifs identifiables et quantifiables. Dans le cas des images, ces motifs doivent être remplacés par des masques respectant la topologie de l'espace à 2 dimensions dans lequel les données sont plongées. Nous avons choisi d'utiliser des masques de taille 3x3, voisinage couramment utilisé en traitement d'images. Cette taille de motif présente l'avantage de prendre en compte le voisinage du pixel tout en restant relativement robuste face au bruit (plus un motif est grand, plus il peut-être affecté par une variation due au bruit). De plus, Rudolf Pareti a montré dans sa thèse [Pareti 08] que des motifs de taille 3x3 sont un bon compromis entre quantité d'information extraite et pertinence de celle-ci.

Le principe général d'analyse d'images à partir d'une loi de Zipf reste le même qu'en dimension 1. Tout d'abord, le nombre d'occurrences de chaque motif est calculé. Cependant, chaque image est encodée sur 256 niveaux de gris. Il existe donc 256 symboles possibles pour décrire chaque pixel, et théoriquement 256^9 (environ $4,7.10^{21}$) motifs possibles. Cette valeur est beaucoup plus grande que le nombre de pixels contenus dans une image [3]. Si tous les motifs d'une image n'apparaissent qu'une seule fois, aucun modèle fiable ne pourra être déduit, et l'analyse statistique établie perd tout son sens. Il devient donc nécessaire de restreindre le nombre de motifs perçu pour donner un sens au modèle. Le codage de chaque pixel devient donc décisif en lui-même. Le but du codage est donc de trouver la méthode d'encodage la plus appropriée pour définir les index capables de distinguer chaque élément constituant d'une lettrine.

Des études ont montré qu'une loi de Zipf pouvait s'appliquer au cas des images avec différentes méthodes d'encodages [Caron 03, Pareti 08]. Dans notre cas, nous cherchons un encodage qui permette de distinguer les images que nous étudions (c'est-à-dire, les images de documents anciens). Plus précisément, nous étudions les lettrines qui ont été numérisées en niveaux de gris. Chaque pixel est donc encodé sur 8 bits (256 niveaux de gris) et ce codage permet d'encoder l'intensité de chaque pixel.

Afin de réduire cette information, c'est à dire l'intensité des niveaux de gris pour chaque pixel, deux approches sont possibles :

- réduire le nombre de symboles utilisés pour décrire l'intensité de chaque pixel (réduire le nombre de niveau de gris pour chaque pixel) ;
- réduire la taille du masque mis en jeu dans le parcours de l'image.

Dans le cadre de ces travaux, nous avons préféré orienter notre choix sur la préservation de la taille des motifs plutôt que de préserver le nombre de niveaux de gris. En effet, si l'on conserve le nombre de niveaux de gris, cela implique de réduire la taille

3. Par exemple, une image de 640x480 pixels contient seulement 304 964 motifs, et les lettrines sont souvent limitées à des dimensions inférieurs à 400 pixels.

des motifs. Or, les images de documents anciens étaient initialement conçues en noir et blanc. Les variations de niveaux de gris décrivent donc plus l'évolution du document à travers le temps (les effets du vieillissement), que l'information qu'ils renferment. Pour ces raisons, ces travaux s'attarderont plus sur la recherche de motifs particuliers que sur la recherche de différences particulières de niveaux de gris.

Une approche simple pour réduire le nombre de niveaux de gris, consisterait à n'utiliser que k niveaux de gris pour caractériser le niveau d'intensité des pixels. Le plus souvent, un petit nombre de niveaux de gris suffit pour observer une image tout en conservant ses détails les plus importants. Étant donné qu'une quantification en k classes égales mènerait à un résultat peu fiable, nous nous sommes inspirés de [Pareti 06a], et nous avons choisi d'utiliser une méthode de classification des niveaux de gris en k classes via l'utilisation d'un algorithme des $k-moyennes$ [McQueen 67]. En effet, l'algorithme des $k-moyennes$ est une méthode dont le but est de diviser des observations en K partitions (clusters) dans lesquelles chaque observation appartient à la partition avec la moyenne la plus proche. Les centres des moyennes ne sont pas fixés par avance, et cela nous permet d'avoir une quantification adaptée aux images à dominante blanche, noire, ou encore très bruitées. Nous avons expérimenté cet algorithme avec différentes valeurs de k and nous avons décidé de ne conserver que 3 niveaux de gris. Ceci reste cohérent vis-à-vis de l'apparence des images de documents. En effet, les documents issus du début de l'imprimerie étaient originellement composés de deux couleurs : le blanc et le noir. Le blanc était issu du papier, tandis que que le noir était issu de l'encre. Le troisième niveau découle du vieillissement de ces documents. Avec le temps, des zones grises sont apparues (étalement de l'encre, diffusion des zones de traits lisses pour devenir des zones grisâtres, ...), et peuvent naturellement être rassemblées dans une classe commune. La figure 6.1.2 présente des exemples d'images réduites à 3 niveaux de gris en utilisant l'algorithme des $3-moyennes$.

Une fois l'image réduite à trois niveaux de gris, nous venons extraire l'ensemble des motifs qui la composent. Une fois ces motifs extraits, il est possible d'étudier leur fréquence d'apparition (dans une image ou dans la base), et ainsi décrire les images de documents anciens.

6.1.2 Sac de motifs et Recherche par le contenu - Approche statistique

La description d'une image de forme est un domaine ancien et très actif en recherche, et de nombreux descripteurs d'images ont été présentés préalablement dans ce manuscrit (voir section 1.3 pour plus de détails). La plupart des approches récentes en recherche d'images par le contenu utilisent des approches statistiques pour décrire les images [Lazebnik 06, Torralba 08, Guillaumin 10]. Parmi l'ensemble des méthodes existantes, les approches par sacs de mots (ou *bag of words* en anglais) sont les plus répandues

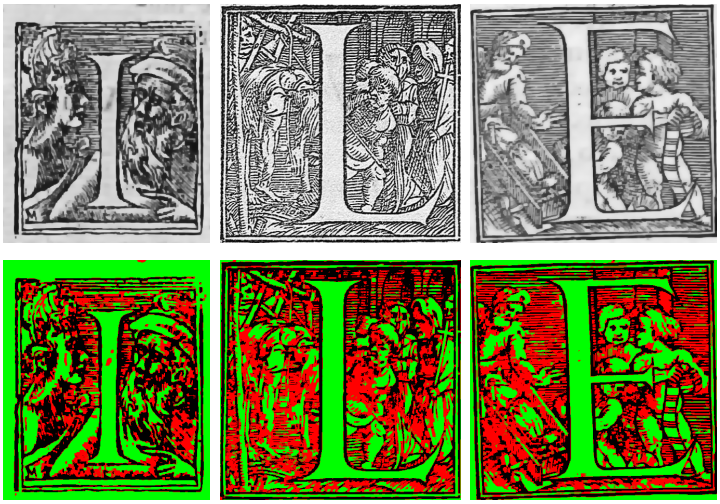


FIGURE 6.1.2: Exemple d'images de la couche forme de Meyer en 256 niveaux de gris (1ère ligne) et en 3 niveaux de gris (2nde ligne). Dans le cas de la seconde ligne, les 3 niveaux ont été représentés par les couleurs rouge, verte et noire pour améliorer le contraste entre les niveaux.

actuellement [Lazebnik 06, Nguyen 09, Guillaumin 10]. Ces approches sont calquées sur les outils de recherche de texte [Baeza Yates 99, Sivic 09]. Ils décrivent une image en détectant la redondance de motifs particuliers, et permettent d'obtenir de bons résultats.

Cette technique est issue des outils de fouille de texte, et consiste à résumer le contenu d'un document ou d'une image en recensant le nombre d'apparitions de ces mots particuliers. Dans le cas d'un texte, les mots sont naturels et il suffit de dénombrer le nombre d'apparitions des mots qui le composent. Dans le cas des images, toutes les approches de la littérature extraient des « *mots visuels* » [Lazebnik 06, Niebles 08]. Ces mots correspondent à des motifs remarquables dans les images. Un motif est remarquable soit parce qu'il correspond à un motif recherché particulier (recherche d'un visage par exemple), soit parce qu'il est caractéristique d'une image (ou d'une classe d'image). Dans cette section, nous présentons donc comment nous avons utilisé les motifs extraits précédemment pour décrire les images de formes. Cette section se décompose naturellement en trois étapes : 1) Recherche des motifs remarquables à retenir pour décrire les images ; 2) Description des images à partir des motifs sélectionnés ; 3) Mesure de similarité entre images.

6.1.2.1 Recherche de motifs remarquables

Les motifs extraits des lettrines sont composés de 9 pixels (motifs de taille 3x3) sur des images en 3 niveaux de gris. Nous avons donc un maximum de 19683 motifs possibles pour décrire nos images. Il n'est pas envisageable de conserver tous ces motifs pour décrire une image pour deux raisons principales : premièrement, comparer des vecteurs

de dimensions supérieur à 19 000 serait lourd en temps et en espace ; deuxièmement, de nombreux motifs ne sont pas pertinents pour chaque image. Il est donc important de réduire le nombre de motif à retenir pour chaque image pour rendre la description plus pertinente. Afin de réduire le nombre de motifs, nous décrivons chaque motif pour définir sa pertinence au vue d'une image et de la base d'images, puis nous sélectionnons les plus pertinents.

Caractérisation des motifs

Une fois que les différents motifs ont été extraits des images, nous allons les considérer comme un mot visuel potentiel. Une méthode classique en fouille de texte, pour comparer des documents et détecter les mots les plus pertinents, s'appuie sur le modèle $TF-IDF$. Le but de $TF-IDF$ (pour *Term Frequency - Inverse Document Frequency*) est de pondérer chaque motif en fonction de sa fréquence dans une image d'une part, et dans la base d'images d'autre part. Ce modèle permet ainsi de détecter les mots visuels les plus pertinents pour décrire une image. Si l'on considère l'ensemble de tous les motifs possibles dans une image I : $|PATT_I|$, et la fréquence du $k_{ème}$ motif (c'est à dire le nombre d'occurrences de celui-ci dans l'image I) : $|Patt_{k,I}|$, le *Term Frequency* du $k_{ème}$ motif dans l'image I correspond à :

$$TF_{Patt_{k,I}} = \frac{|Patt_{k,I}|}{|PATT_I|} \tag{6.1.2}$$

On peut voir ce terme comme l'importance que l'on donne à un motif comparativement à tous les autres motifs présents dans l'image.

Le second terme, IDF (*Inverse Document Frequency*) mesure la rareté d'un motif dans la base d'image. C'est à dire que plus un motif sera rare dans la base, plus cette valeur sera importante. Il permet donc de connaître les motifs qui sont spécifiques à quelques rares images. Cette valeur est estimée en utilisant toute la base d'image (ou du moins toutes les images présentes lors de l'apprentissage). En considérant un ensemble de $|I|$ images dans la base (la $j_{ème}$ image de la base est notée i_j), l'IDF d'un motif $Patt_k$ est alors obtenu comme suit :

$$IDF_{Patt_k} = log\frac{|I|}{|i_j : Patt_i \epsilon i_j|} \tag{6.1.3}$$

On peut donc interpréter cette valeur comme étant le nombre d'image dans la base contenant un motif. À partir de ces deux valeurs, il est possible de calculer le coefficient $TF-IDF_{i,j}$ d'un motif j dans une image i :

$$TF-IDF_{i,j} = TF_{i,j} * IDF_j \tag{6.1.4}$$

Chaque motif est ainsi pondéré en fonction de sa fréquence d'apparition dans une image, et dans la base.

Sélection des motifs les plus pertinents

Afin d'observer les valeurs de TF-IDF, nous les avons ordonnées par ordre décroissant et nous avons construit la courbe de Zipf associée (courbe qui représente la « fréquence » (au sens de TTF-IDF) des motifs en fonction de leur rang. Un exemple de courbe obtenue sur une image de document ancien est présenté dans la figure 6.1.3. On peut remarquer que seule une petite partie des motifs présents dans une image présente une valeur de $TF - IDF$ élevée (relativement aux valeurs des autres motifs). C'est à dire que seule une petite partie des motifs d'une image sont pertinents pour la décrire du point de vue du critère TF-IDF.

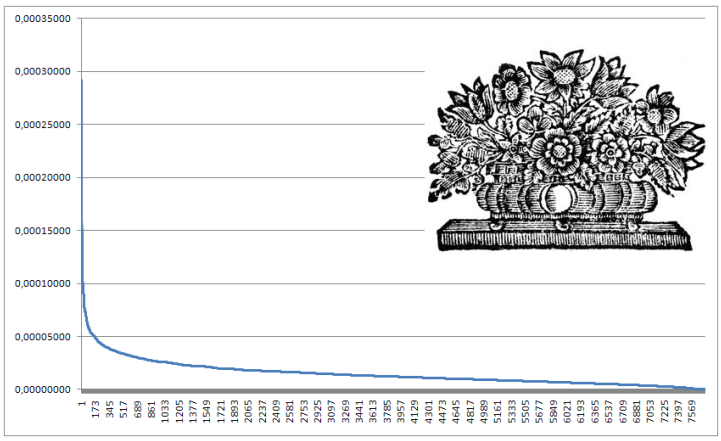


FIGURE 6.1.3: Exemple de courbe TF-IDF obtenue sur une image de document ancien

Si l'on observe la courbe de plus près, on peut remarquer qu'elle n'est pas linéaire mais plutôt qu'elle est similaire à une exponentielle décroissante. La partie gauche de la courbe correspond aux motifs (et donc aux régions) de l'image avec une forte valeur de TF-IDF. Un motif avec une valeur élevée correspond à un motif fréquent dans l'image, et qui n'est pas présent dans toutes les images de la base. En d'autres mots, les motifs qui appartiennent à la partie gauche de la courbe apparaissent fréquemment dans l'image, et lui sont donc spécifiques. Il parait donc cohérent de ne retenir que les motifs les plus à gauche pour décrire le contenu d'une image. De plus, comme nous appliquons cette méthode sur la couche forme de Meyer, nous pouvons supposer que les motifs les plus fréquents décrivent le style de la lettrine sans se limiter uniquement aux grandes formes de l'image (qui sont naturellement ignorées par le critère $TF - IDF$).

Partant de ce constat, nous avons décidé de ne garder que les motifs possédant une valeur de $TF - IDF$ située dans la partie de gauche de la courbe. Pour ce faire, nous retenons les motifs qui ont un $TF - IDF$ supérieur à $t\%$ de la valeur max de $TF - IDF$ que l'on peut trouver dans chaque image. Après avoir testé plusieurs valeurs de t sur une base de 900 lettrines, nous avons constaté que la valeur optimale de t est de 0,20 (20% de la valeur maximale de TF-IDF que l'on peut trouver dans une image).

L'intérêt majeur de cette sélection repose sur le fait qu'elle est adaptée à chaque image, et que contrairement aux approches classiques de sélection de caractéristiques, nous ne sélectionnons pas le même sous-ensemble de motifs pour toute la base, mais plutôt le meilleur sous-ensemble pour chaque image. Ainsi, pour chaque image, seuls les motifs les plus fréquents, c'est à dire ayant un $TF - IDF$ supérieur à 20% de la fréquence maximale, sont conservés. Le sous-ensemble retenu est caractéristique de l'image, et sa sélection est adaptative.

6.1.2.2 Description des images à partir des motifs sélectionnés

Une fois les motifs sélectionnés, chaque image est décrite par un vecteur composé des valeurs de TF-IDF de chaque motif retenu. Ainsi, chaque image est caractérisée par un vecteur de taille variable correspondant aux $TF - IDF$ des motifs sélectionnés. Afin de comparer deux images, il devient nécessaire de définir une mesure de dis-similarité entre deux vecteurs de motifs. En effet, ces motifs peuvent être de taille différente, et composés de motifs différents. Nous présentons ci-après comment nous comparons les vecteurs de motifs entre-eux.

6.1.2.3 Mesure de similarité entre images

Chaque image est décrite par un vecteur qui contient la liste des motifs qui la caractérisent le mieux (au sens du critère TF-IDF). Cette sélection créée des vecteurs de tailles différentes et composés de motifs différents. Nous avons donc défini une mesure de similarité entre deux vecteurs pour comparer les images. Cette mesure entre deux vecteurs (v_1 et v_2) intègre différents aspects sous-jacents au vecteur TF-IDF, afin de décrire le plus finement possible la ressemblance entre les images. Cette mesure repose sur un triplet d'informations : < Longueur relative ; Taux de recouvrement ; Distance entre motifs >. Voici en détails la signification de chacune de ces mesures :

1. **Longueur relative** : afin d'intégrer la similarité entre deux vecteurs, nous calculons tout d'abord un rapport entre la longueur des deux vecteurs à comparer :

$$R_L = \frac{\min(L_{v_1}, L_{v_2})}{max(L_{v_1}, L_{v_2})} \tag{6.1.5}$$

   avec L_{v_1}et L_{v_2} qui correspondent aux longueurs respectives des deux vecteurs à comparer. Plus cette mesure sera faible, plus les vecteurs auront des tailles différentes, et les images seront considérées dis-similaires ;

2. **Taux de recouvrement** : la seconde mesure comptabilise le nombre de motifs qui sont présents dans les deux vecteurs. Soient $Patt_1$ et $Patt_2$ l'ensemble des motifs des deux images à comparer, le taux de recouvrement R_S est obtenu comme suit :

$$R_S = \frac{\|P_i : P_i \epsilon Patt_1 \cap Patt_2\|}{max(L_{v_1}, L_{V_2})} \tag{6.1.6}$$

Ce rapport indique si deux vecteurs contiennent les mêmes motifs, c'est à dire si les images sont composées des mêmes motifs ;

3. **Distance entre motifs** : Enfin, afin d'obtenir une comparaison plus précise entre les images composées des mêmes motifs, nous calculons une distance euclidienne entre les motifs qui apparaissent dans les deux vecteurs. Soit $Patt_{12}$ l'ensemble de motifs qui sont communs à $Patt_1$ et $Patt_2$ et $L_{Patt_{12}}$ le nombre de motifs en commun, la distance correspond à :

$$D_{Patt} = \sum_{i=1}^{L_{Patt_{12}}} \frac{\sqrt{(TF-IDF_{Patt_1}(i)} - TF-IDF_{Patt_2}(i))^2}{L_{Patt_{12}}} \tag{6.1.7}$$

La distance est enfin normalisée en la divisant par la valeur maximale de $TF-IDF$ qui est trouvé dans la base d'image.

Les deux premières mesures indiquent la similarité entre les vecteurs. R_L et R_S ont une valeur comprise $\epsilon\,[0;1]$, avec 0 correspondant à deux dis-similaires, tandis qu'une valeur de 1 indique une concordance parfaite. Enfin, la distance euclidienne entre les motifs indique si deux vecteurs sont semblables, c'est à dire qu'une valeur égale à 0 indique que les vecteurs sont identiques (et vice-versa).

Nous combinons donc ces trois valeurs dans une mesure globale de similarité. Considérant deux images i_1 et i_2, la mesure de similarité est définie telle que :

$$Sim_{i_1,i_2} = (1 - (R_L * R_S)) * D_{Patt}) \tag{6.1.8}$$

Partant de cette mesure de similarité, il nous est possible de comparer des images et de mesurer leurs ressemblances. Nous évaluons cette approche dans la partie validation de ce chapitre.

6.1.3 Segmentation de l'image et extraction de couches - Approche topologique

Nous rappelons que le modèle d'analyse des documents anciens que nous avons présenté dans le chapitre 5, consiste à définir les motifs élémentaires à partir desquels il est possible de décrire le contenu de l'image, et d'extraire des zones d'intérêts. Nous venons de voir dans la section 6.1.1.2 comment il est possible d'extraire des motifs élémentaires qui représentent les formes. Une sélection particulière de ces motifs, vis-à-vis de chaque image et de l'ensemble de la base, nous a permis de décrire les images par des sacs de motifs.

Afin de respecter notre modèle, il est nécessaire de définir une méthode qui permette d'extraire des régions d'intérêts à partir des motifs extraits directement de l'image. C'est ce que nous présentons dans cette section.

Si l'on trace l'ensemble des fréquences ordonnées des motifs d'une image sur un graphe en deux dimensions, on peut constater que l'on obtient une courbe de Zipf similaire à celle présentée dans la figure 6.1.4. Comme l'explique Rudolf Pareti dans sa thèse[Pareti 08], la partie gauche de la courbe est liée aux zones les plus fréquentes, c'est à dire les objets de l'image, tandis que la partie droite correspond aux motifs les moins fréquents dans l'image.

En observant de manière plus approfondie les courbes associées à ces images, on peut remarquer qu'elles ne sont pas toujours linéaires. On peut donc en déduire que la loi de Zipf n'est pas vérifiée pour l'ensemble des motifs. Néanmoins, un certain nombre de lignes droites peut être constaté. Cela signifie que plusieurs structures apparaissent dans l'observation de la distribution des motifs. Ceci est tout à fait naturel puisque différents éléments complexes apparaissent dans une lettrine, comme la lettre et le fond. Une loi de Zipf ne peut pas estimer de manière fidèle la distribution des motifs, cependant un mélange de ces lois peut-être considéré. En fonction de la méthode utilisée pour coder l'information contenue dans les pixels, ces zones (c'est à dire les zones de l'image) peuvent être extraites pour être décrites.

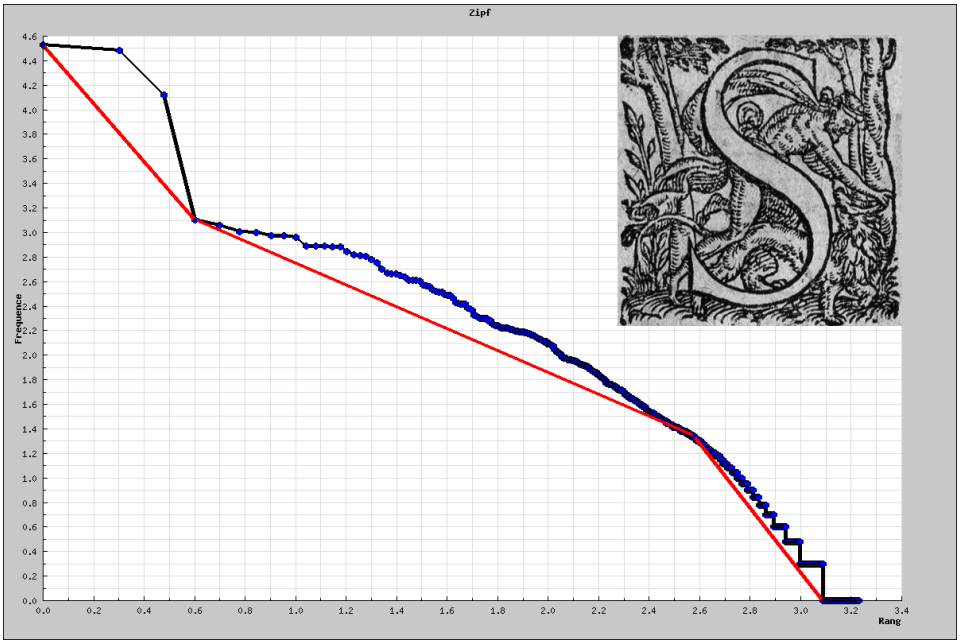


FIGURE 6.1.4: Exemple de courbe associée à la Loi de Zipf de l'image présentée ci-contre

Ces trois courbes reflètent trois types de motifs que l'on peut retrouver dans les lettrines. Ceci a été observé et mis en évidence dans [Pareti 08], et validé dans toutes nos expérimentations. Ces trois fonctions linéaires permettent de séparer les motifs très fréquents, des motifs moyennement et peu fréquents. Des motifs fréquents correspondent à des zones qui apparaissent souvent dans une image. Dans le cas des lettrines, ceci correspond aux zones uniformes des lettrines (la lettre, les personnages, ...). Ceci peut

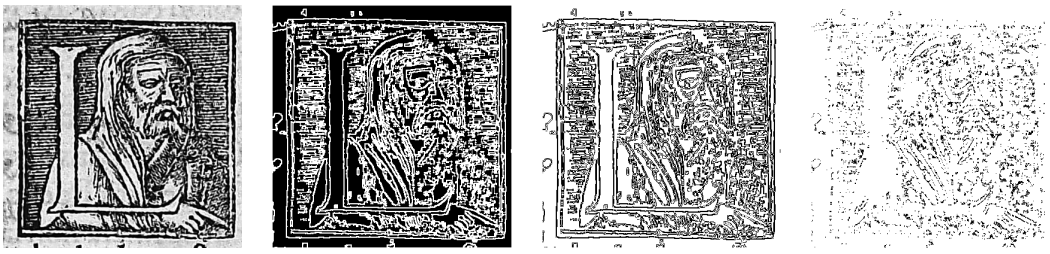
facilement être déduit de l'idée qu'un objet doit être suffisamment grand pour que l'oeil humain lui porte de l'attention. Les motifs moyennement et peu fréquents quant à eux sont plus difficiles à interpréter. Rudolf Pareti, dans sa thèse, propose de voir ces motifs comme étant ceux qui caractérisent les contours et l'intérieur des contours. L'avantage de l'extraction des zones fréquentes, c'est qu'elle ne se limite pas à l'extraction des formes d'une couleur ou qui suivent un motif particulier. Seule leur fréquence pondère leur importance.

Une des possibilités pour estimer ces trois fonctions linéaires consiste à utiliser l'approche des moindre carrés récursifs. Comme les points ne sont pas espacés régulièrement, les points de la courbe sont ré-échantillonnés le long de l'axe des abscisses avant d'estimer ces fonctions. Il est alors possible de récupérer, à partir de l'image, les pixels qui contribuent aux différentes zones de la courbe de Zipf. Tous ces pixels sont assemblés dans une image que nous appelons couche de Zipf. L'image est décomposée en plusieurs couches. Les motifs de la première couche (partie gauche de la courbe) correspondent aux zones fréquentes, tandis que la partie droite de la courbe correspond aux motifs moins fréquents (textures et contours de l'image).

Cette décomposition permet d'extraire à la fois de l'information sur les régions elles-mêmes, mais également de l'information sur la structure des contours. C'est ce que nous appelons la décomposition de Zipf d'une image, telle que présentée dans 6.1.9, avec I_0 qui correspond à l'image originale, et L_n à la $n^{ième}$ couche avec $n \in [1;3]$. Dans ce cas, la somme utilisée est une somme exclusive, c'est à dire que l'information contenue dans une couche n'est pas présente dans les autres.

$$Z(I_0) = L_1 + L_2 + L_3 \tag{6.1.9}$$

Les différentes zones associées aux 3 pentes détectées sont automatiquement extraites en utilisant un processus récursif. Un exemple de résultat est présenté dans la figure 6.1.5.



(a) Image originale (b) Couche n°1 : zones fréquentes (c) Couche n°2 : zones moyennement fréquentes (d) Couche n°3 : zones peu fréquentes

FIGURE 6.1.5: Exemple de couches obtenues avec la décomposition de Zipf

Un point de rupture de la courbe va permettre de définir le début (ou la fin) de chaque

fonction linéaire utilisée pour l'approximation de la loi. Un point de rupture est défini comme le point le plus éloigné de la ligne droite qui relie les deux extrémités de la courbe. Il faut noter que la courbe associée à l'image ne peut être estimée par une seule fonction linéaire, c'est à dire une seule loi de Zipf. En effet, chaque image comporte un mélange de plusieurs phénomènes qui sont mis en évidence à chaque itération. Plusieurs loi puissance sont impliquées, à partir desquelles plusieurs valeurs d'exposants sont calculées.

La méthode peut donc caractériser non seulement l'image dans son apparence globale, mais aussi sa composition. Dans le cas des lettrines, nous pouvons par exemple noter que la première couche de Zipf comprend la lettre et les grandes formes de l'image (morceaux de personnes dans le cas d'images figuratives par exemple). La seconde et la troisième couche contiennent plus les contours et petits motifs et ramification des décorations. Nous nous intéressons donc à la première couche qui permet d'extraire les éléments les plus structurants (les grandes formes) de l'image, but de cette section.

Notre méthode présente l'avantage de ne pas être guidée par le niveau de gris des formes, ni par une connaissance à priori des images. Ceci permet donc d'extraire les zones les plus fréquentes de l'image, zones visuellement intéressantes pour l'oeil humain puisque redondantes. Ce sont les principales raisons qui nous ont orientés vers le choix de ce modèle qui repose sur la fréquence plutôt que sur l'aspect général des motifs.

6.1.3.1 Extraction des régions de formes

Jusqu'à présent, nous avons présenté deux décompositions qui permettent de simplifier l'image (décomposition de Meyer) puis de séparer les informations contenues dans l'image en fonction de leurs fréquences d'apparition (décomposition selon une loi de Zipf). Les deux méthodes de décomposition utilisées présentent deux objectifs bien distincts : la première approche considère l'image comme une superposition de trois signaux différents (approche basée sur la décomposition de Meyer), pendant que la seconde considère l'image comme une juxtaposition d'objets avec des significations différentes (Loi de Zipf).

On peut remarquer que de par la nature même des lettrines, la lettre correspond à une grande zone uniforme au centre de l'image qui peut-être noire ou blanche. Extraire la lettre nécessite donc d'être capable de détecter les grandes zones de l'image indépendamment de leur couleur. Il en va de même pour tous les objets qui composent une lettrine, qui correspondent à de grandes zones uniformes de l'image, et qui de plus ne sont pas forcément de la même couleur que la lettre. Extraire les formes d'une lettrine consiste donc à segmenter l'image d'origine tout en étant :

- robuste aux changements de niveaux gris (dégradation dues au temps) ;
- invariant aux changements de couleurs utilisées pour créer les formes ;
- insensible aux textures et zones de traits.

C'est donc l'utilisation conjointe de ces deux approches qui permet l'extraction des formes. La première décomposition permet d'avoir un signal plus pur, tandis que la seconde permet un partitionnement de l'image. L'extraction des formes consiste alors à

l'extraction des composantes connexes résultantes des deux transformations précédentes. La figure 6.1.6 présente les résultats d'extraction des formes d'une lettrine à partir de la Loi de Zipf uniquement, puis via l'utilisation combinée de la décomposition de Meyer et de la Loi de Zipf. On constate clairement que les formes extraites en partant de la couche forme de Meyer sont beaucoup plus nets, et mieux séparés des éléments du fond de l'image.



(a) Image d'origine (b) Zones les plus fréquentes de l'image d'origine selon la loi de Zipf (en noire) (c) Zones les plus fréquentes de la couche forme de Meyer, selon la Loi de Zipf (en noire)

FIGURE 6.1.6: Exemple de résultat obtenu en appliquant la décomposition de Zipf soit sur l'image d'origine, soit sur la couche forme de Meyer. On peut observer que les composantes connexes extraites à partir de la couche forme de Meyer sont plus régulières et plus grandes. Ceci permet d'obtenir une meilleure compréhension de l'image du point de vue visuel

De plus, on peut remarquer que les plus grandes composantes connexes extraites peuvent être associées à des objets sémantiques puisque par exemple, la plus grande composante connexe correspond à la lettre ou une partie du visage. Cependant, la lettre ne correspond pas toujours à la plus grande composante connexe, et distinguer les différentes formes extraites impose d'analyser ces formes en elles-mêmes pour en déduire des critères plus élaborés. C'est pour cela que nous avons décidé de décrire chacune des régions que nous avons extraites à l'aide de caractéristiques de formes. Nous présentons l'étape de description ci-dessous.

6.1.3.2 Description des régions de formes

Une fois que les formes ont été extraites, c'est à dire que nous avons extrait les composantes connexes de l'image pré-traitée, nous obtenons une liste de régions. En observant toutes ces formes dans la figure 6.1.7(c), nous pouvons voir que les formes les plus importantes ont des caractéristiques particulières (du point de vue de leur taille, de leur position, de leur centre de gravité et de leur excentricité). En réalisant une sélection de ces composantes connexes basée sur ces caractéristiques, il devient possible d'extraire

des régions d'intérêts dans les lettrines. Un exemple de zones extraites peut-être observé dans la figure 6.1.7(d).



(a) (b) (c) (d)

FIGURE 6.1.7: Exemples d'images et de représentations des différents traitements : a) Image originale ; b) Couche forme de Meyer (zones avec de faibles variations de niveaux de gris) ; c) Formes segmentées en utilisant la loi de Zipf (les formes apparaissent en bleue dans cette image) ; d) Plus grandes composantes connexes retenues

Chaque zone extraite est considérée comme une forme. Certaines de ces formes correspondent à des éléments importants qu'il faut extraire pour pouvoir reconnaître l'image. Une sélection simple (telle que retenir la plus grande composante connexe) ne permet pas d'obtenir toujours le même élément. Par exemple, la plus grande composante connexe correspond parfois à la lettre, parfois à un élément du fond, parfois au cadre, sans pouvoir décider réellement de ce qui est extrait. Afin de décrire de manière plus fine les formes, nous avons décidé d'utiliser des descripteurs de formes. Il en existe de très nombreux dans la littérature (voir section 1.4.1.1), et ceux-ci permettent de décrire statistiquement le contenu des formes, soit de décrire leurs contenus topologiquement. Dans cette seconde catégorie de descripteurs, on peut retrouver des descripteurs qui permettent de décrire l'aire d'une forme, son nombre de trous, ou encore son orientation principale.

Dans le cadre de ce manuscrit, nous cherchons à représenter les objets qui composent une lettrine pour permettre aux utilisateurs finaux de retrouver des images semblables à celles qu'ils recherchent. Pour arriver à ce but, nous utilisons un modèle qui extrait de l'information des images, et propose de lier cette connaissance à celle des historiens (typiquement, retrouver une lettrine figurative, *ie* qui contient des personnages, ou bien retrouver la lettre dans une lettrine). Il devient alors nécessaire d'enrichir les formes

extraites à l'aide d'informations liées à la topologie de ces formes (où se situent-elles dans l'image ? Quelle est leur aspect ? ...) pour pouvoir en déduire une connaissance de plus haut niveau (c'est une forme ovale située au centre de l'image par exemple). L'idée globale étant de décrire de manière relativement simple et robuste les formes. C'est la combinaison de ces formes qui permettra de décrire l'image en elle-même.

Dans le cadre de ce manuscrit, nous avons choisi d'enrichir les formes en les décrivant à l'aide de caractéristiques classiques de la littérature. Ces caractéristiques ont été choisies pour leur généricté et leur rapidité de calcul (caractéristiques qui doivent être réutilisables pour d'autres bases d'images, et calculables rapidement sur plusieurs centaines d'images). Voici le vecteur de caractéristiques que nous avons retenu :

$$(C_G\,;\, Aire\,;\, Exc\,;\, NdG_{moy}\,;\, NdG_{ect}\,;\, Eul\,;\, Part) \tag{6.1.10}$$

Avec :

- C_G : coordonnées du centre de gravité de la forme, pour la localiser dans l'image ;
- $Aire$: nombre de pixels composant l'image, pour connaître sa taille ;
- Exc : excentricité de la forme, telle que définie dans [Pratt 07]. L'excentricité correspond au rapport entre les rayons de l'ellipse englobante (avec r_M le plus grand, et r_m le plus petit rayon de l'ellipse englobante), et permet d'avoir une information sur l'allongement de la forme :

$$Exc = \frac{r_M - r_m}{r_M + r_m}$$

- NdG_{moy} : niveau de gris moyen de la forme, pour avoir une information sur la tendance colorimétrique de la forme ;
- NdG_{ect} : écart-type des niveaux de gris de la forme, pour avoir une information sur sa régularité ;
- Eul : nombre d'Euler de la forme, tel que défini dans [Pratt 07]. Initialement, le nombre d'Euler correspond à une différence entre le nombre de forme et le nombre de trous dans l'image. Comme dans notre cas, nous étudions des composantes connexes indépendamment les unes des autres, nous calculons le nombre d'Euler sur une région à la fois. Il peut donc être réduit à (H correspond au nombre de trou) :

$$Eul = 1 - H$$

- $Part$: dernière partie du vecteur, celui-ci correspond à une liste de booléens qui indiquent si une forme possède des pixels dans une des sous-parties du masque défini dans la figure 8.4.1. Les sous-parties résultent d'une division de l'image en neuf parties fixes (la figure 8.4.1 présentent un exemple de découpage possible, où les sous-parties sont numérotées de 1 à 9). Les booléens associés à cette partition sont utilisés pour localiser les formes dans l'image. Dans le cas d'images de documents, il est important de connaître la position d'un objet dans l'image. En effet, ceux-ci comportent une structure implicite liée au fait que nous traitons des images créées

par l'homme pour l'homme. Par exemple, la lettre d'une lettrine est située au centre, elle correspondra donc à une forme qui n'aura pas de recouvrement avec une zone proche d'un bord.

Il est important de noter qu'il est tout à fait envisageable d'ajouter d'autres caractéristiques au vecteur retenu pour adapter la description à d'autres types d'images, ou d'autres cas d'utilisations.

6.1.3.3 Construction de la structure topologique

Maintenant que nous avons extrait et décrit l'ensemble des régions de la couche forme, il nous faut les intégrer dans un un modèle structurel qui décrit à la fois les différents composants de l'image et les relations qui les lient. L'intérêt d'un modèle structurel repose la richesse de sa représentation, très élevée comparativement à celle des vecteurs caractéristiques [Bunke 01]. Parmi les structures de données qui sont largement utilisées dans la reconnaissance de formes structurelles, Salim Jouili dans sa thèse [Jouili 11] distingue 3 catégories : les chaînes, les arbres et les graphes. Du point de vue algorithmique, les graphes sont des structures de données généralisantes de toutes les structures de données [Borgwardt 07, Riesen 09], en particulier, les chaînes et les arbres. En outre, un vecteur peut être aussi représenté par un graphe où les noeuds correspondent aux éléments du vecteur. Par conséquent, les travaux sur les graphes sont applicables sur les autres structures de données.

La représentation d'objets sous forme de graphes offre trois avantages majeurs par rapport aux vecteurs caractéristiques. Premièrement, elle offre une représentation plus expressive que les vecteurs à travers la description explicite de la structure de l'objet par l'ensemble de ses arcs et de leur pondération. Deuxièmement, le nombre de nœuds et des arêtes n'est pas limité a priori et peut être adapté à la complexité de chaque objet. Troisièmement, les structures de graphes sont généralement faciles à interpréter et plus facilement lisibles pour un humain.

Grâce à ces propriétés, les graphes ont connu une large utilisation dans différents domaines scientifiques [Conte 04]. La reconnaissance d'images en est un exemple puisque la représentation sous forme de graphe a prouvé sa flexibilité dans de grandes variétés de types d'images (documents anciens, schéma électriques et architecturaux, images de scènes naturelles, images médicales, ...). Tous ces travaux ont utilisé les graphes pour obtenir une représentation qui préserve l'information topographique de l'image ainsi que les relations entre ses composantes (ses régions).

Dans la littérature, plusieurs travaux [Bunke 01, Jouili 08, Jouili 09b] discutent la représentation d'images sous forme de graphe. Dans notre cas, à partir des régions qui ont été extraites, il nous est possible de construire un graphe de voisinage où les noeuds correspondent aux régions d'intérêts, et un arc relie une région à la région la plus proche. Chaque noeud est alors décrit par l'ensemble des caractéristiques présentées précédemment, et chaque arc indique la distance entre les noeuds qu'il relie. Nous nous sommes appuyés sur cette démarche pour décrire les régions extraites par notre approche.

Nous tenons à préciser que nous parlons de relations topologiques dans notre modèle, et que ces relations topologiques s'appuient sur l'algorithme RCC-8. Cet algorithme permet principalement de savoir si deux régions se touchent, et si oui de quelle manière (recouvrement, inclusion, ...). Les régions qui sont extraites sont des composantes connexes, c'est à dire que deux régions sont par définition déconnectées. Il n'y aurait donc aucun intérêt à utiliser ce type d'algorithme dans la description des graphes de nos images, puisque nous cherchons à connaître la position d'une région par rapport à une autre. Nous avons donc utilisé la distance entre les régions pour caractériser leur agencement relatif.

Ainsi, chaque image est décrite par un graphe, où chaque noeud représente une région d'intérêt, et les arcs relient les régions les plus proches en qualifiant leur relation (distance entre ces régions).

6.2 Validation de l'approche sur la couche forme

Afin d'évaluer la pertinence de notre chaîne d'extraction de régions de formes, nous avons réalisé deux types d'expérimentations. Une première consiste à comparer notre approche avec une vérité terrain. Celle-ci doit permettre de constater que notre approche est capable de détecter toutes les formes qui pourraient intéresser un utilisateur final (un être humain). La seconde série d'expérimentations a consisté à évaluer le pouvoir discriminant des formes et de leurs descriptions. Ainsi, nous avons cherché à regrouper toutes les images qui ont des descriptions similaires, et nous évaluons la pertinence du clustering ainsi réalisé. Nous présentons ces deux séries d'expérimentations ci-dessous.

6.2.1 Comparaison avec une vérité terrain

Afin de valider précisément les résultats de la segmentation obtenue, nous avons défini deux critères objectifs. Le but de cette évaluation est d'obtenir une information quantitative sur les résultats en utilisant des critères semblables à ceux utilisés dans l'évaluation des moteurs de recherche (rappel-précision). Nous avons donc généré une vérité terrain manuelle sur un sous-ensemble de 45 images de lettrines, où l'utilisateur devait indiquer les zones de l'image qui contiennent des formes prégnantes. Nous sommes conscient de la taille modérée de cette base d'évaluation (45 images), mais l'annotation manuelle d'images est une tâche fastidieuse.

Un exemple de masques obtenus à partir de cette vérité terrain peut-être observé dans la figure 6.2.1. Dans ces masques, un pixel noir correspond à une région contenant une (ou des) forme(s) pendant qu'un pixel blanc correspond au fond (zone ne contenant pas de forme).

Évaluer la qualité de l'extraction des formes n'est pas chose facile, et afin d'apporter une solution à ce problème, plusieurs systèmes d'évaluation de performances ont été développés dans la littérature [Girard 10, Rusiñol 09b, Delalandre 10]. Ils sont tous issus

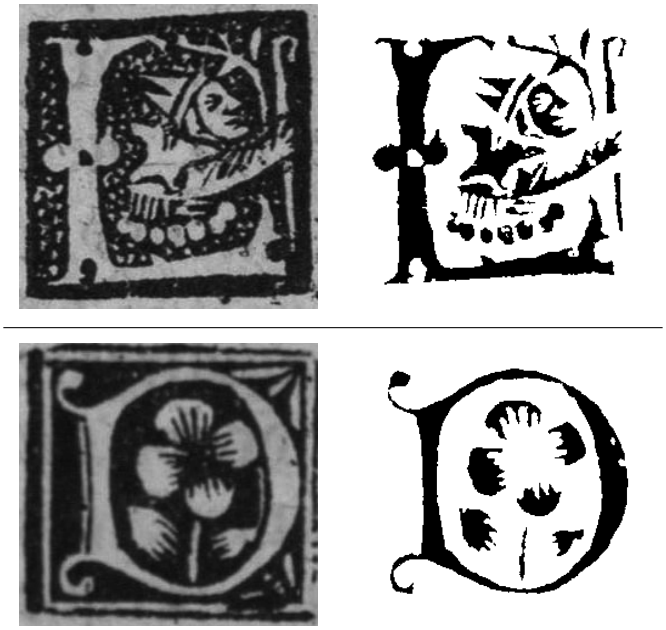


FIGURE 6.2.1: Exemple de lettrines avec les masques obtenus après saisie de la vérité terrain. Les masques (tâches noires) correspondent aux régions de formes dans l'image

du domaine de la recherche d'information, et un bon état de l'art de ces méthodes, spécialement appliquées aux images graphiques et de documents, peut-être trouvé dans la thèse de Marcal Rusinol [Rusiñol 09a]. Parmi les méthodes les plus connues, nous pouvons citer la courbe ROC, la Précision moyenne, ou encore la F-mesure. Tous ces critères découlent de deux critères principaux : le rappel et la précision. Ces deux critères permettent d'identifier la pertinence de l'information renvoyée par le système. Afin d'évaluer la pertinence des formes extraites via notre méthode, nous avons décidé d'utiliser deux critères similaires.

Le premier critère, le **rappel**, indique le pourcentage de régions issues de la vérité terrain qui contiennent des formes extraites automatiquement. Un rappel de 100% signifie que pour toutes les régions de la vérité terrain, une forme a été automatiquement extraite par l'approche décrite ci-dessus. A contrario, une valeur de rappel égale à 0 signifie qu'aucune forme n'a été extraite dans les régions marquées par la vérité terrain.

Le second critère, la **précision**, représente le pourcentage de surface de la vérité terrain qui est recouverte par des formes extraites automatiquement. Nous tenons à préciser que les régions de la vérité terrain ont été annotées manuellement, et que des erreurs infimes peuvent apparaître de part et d'autre des zones, ce qui réduit la valeur de ce critère.

Le tableau 6.1 présente les valeurs de rappel et de précision obtenues sur un sous-ensemble annoté de 45 images. On peut remarquer que notre système obtient un très bon résultat en rappel, puisque l'on peut constater que toutes les régions annotées dans

la vérité terrain contiennent une forme extraite automatiquement. La seconde valeur, la précision, est relativement correcte puisqu'un peu plus de 70 % des zones détectées correspondent à des zones annotées. Les 29% de zones qui ne sont pas détectées peuvent s'expliquer par des phénomènes liés aux dégradations des images par le temps. Ainsi, bien que l'image soit débruitée, des erreurs de segmentation subsistent, et les formes extraites ne sont pas toujours bien lisses et régulières (les contours ne sont pas aussi nets que dans la vérité terrain, et des trous peuvent apparaître dans les zones détectées). Enfin, la vérité terrain annote des régions de formes, et celles-ci sont liées aux zones de traits. La frontière entre ces zones, dans la vérité terrain comme dans les images extraites automatiquement, n'est pas toujours parfaite et induit des erreurs dans la précision de l'extraction et de l'évaluation. Un exemple de zones extraites automatiquement est présenté dans la figure6.2.2.

	Rappel	Précision
Approche proposée	92,4 %	84,8 %

TABLE 6.1: Rappel et précision obtenus avec notre méthode d'extraction de traits sur un sous-ensemble de 45 lettrines accompagnées de leur vérité terrain. Le critère de rappel permet de mettre en avant la capacité de notre approche à extraire les traits marqués dans la vérité terrain, pendant que la précision permet de connaître le pourcentage de surface des régions de la vérité terrain qui est couverte par les traits extraits automatiquement



(a) Image originale (b) Régions annotées dans la vérité terrain (en noirs) (c) Régions extraites automatiquement (en noirs)

FIGURE 6.2.2: Exemple d'image ou l'extraction automatique ne permet pas de couvrir l'ensemble de la surface des zones annotées dans la vérité terrain

Finalement, à partir de cette première évaluation, nous pouvons déduire que notre système est capable d'extraire les formes prégnantes, du point de vue des utilisateurs finaux, dans les lettrines. Cependant, afin d'être exhaustif dans l'évaluation de notre approche, nous proposons d'évaluer la qualité de l'extraction et de la description de formes en comparant nos images (évaluation du pouvoir discriminant de notre approche).

6.2.2 Évaluation des mots visuels obtenus sur les formes de l'image

Cette deuxième évaluation a pour but de vérifier la validité de notre approche, et ainsi prouver qu'elle permet d'obtenir des résultats pertinents. Nous rappelons que chaque image de forme est décrite en utilisant les valeurs de $TF - IDF$ des motifs qui la composent (motifs ayant une valeur de TF-IDF supérieur à 20% de la valeur maximale de TF-IDF de chaque image). En utilisant la mesure de similarité définie précédemment, nous avons mené des expérimentations afin d'évaluer la pertinence de notre approche de « *sacs de motifs* ». Nous avons cherché à évaluer notre méthode sur deux bases différentes pour vérifier que les résultats obtenus sont génériques et utilisables avec d'autres images que des lettrines. Nous présentons donc les résultats obtenus avec des lettrines dans un premier temps, et avec des images de documents anciens dans un second temps.

6.2.2.1 Expérimentations avec des images de lettrines

Les expérimentations ont été menées sur une base de 358 lettrines. Cette base d'image a été utilisée par Pareti et al. dans [Pareti 08] pour évaluer la pertinence de leurs caractéristiques issues de la loi de Zipf. Les lettrines sont issues du Centre d'Études Supérieures de la Renaissance (CESR) de Tours. Ces images de lettrines ont été annotées par des experts pour indiquer leur « style ». Les styles que l'on peut retrouver dans la base sont illustrés dans la figure 6.2.3.

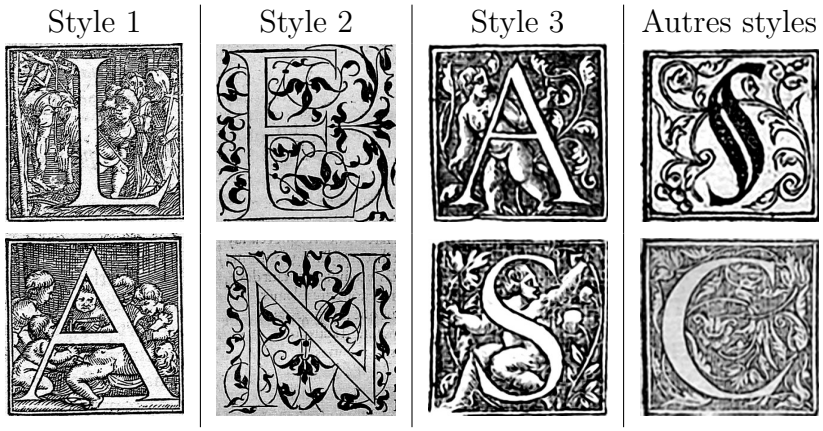


FIGURE 6.2.3: Exemples de lettrines pour chaque style présent dans la base d'images utilisées par [Pareti 08]. Le style numéro 1 correspond aux lettrines composées de personnages dans une fond hachuré, le style numéro 2 est composé des lettrines avec un fond blanc et décoratif, le style numéro 3 aux lettrines avec un fond noir et des personnages, et enfin le dernier style contient toutes les lettrines qui n'appartiennent pas aux trois premiers styles

Nous avons utilisé 10 images par style en apprentissage. Cet apprentissage porte sur

K	Total
1	63,4 %
3	64,6 %
5	70,1 %

TABLE 6.2: Taux de reconnaissances obtenus avec notre méthode sur la base de lettrine utilisée dans [Pareti 06a] pour différentes valeurs de K dans l'algorithme des K-Plus Proches Voisins

environ 10% de la base d'image (40 images) et sert au calcul des IDF. Nous avons volontairement limité l'apprentissage à un faible taux pour mettre notre système en situation semblable à un processus d'indexation de grandes masses de données. Le reste de la base (soit 318 images) a été utilisé dans l'étape de reconnaissance. Pour chaque image, nous calculons alors sa valeur de similarité avec toutes les autres de la base. Nous appliquons ensuite un algorithme des K-Plus Proches Voisins [Cover 67] pour retrouver les images les plus similaires à l'image requête. Les résultats obtenus sont présentés dans le tableau 6.2. Les résultats sont présentés pour trois valeurs de K différentes utilisées dans le classifieur des K-Plus Proches Voisins. Dans le cas ou $K = 1$, le taux de reconnaissance correspond aux images pour lesquelles le système a renvoyé une image du même style. Pour $K = 3$ et $K = 5$, avons appliqué un processus de vote à la majorité [Hechenbichler 04], et sélectionné la classe à attribuer à l'image requête.

Les résultats montrent que l'approche proposée permet d'obtenir des résultats corrects sans pouvoir toutefois décider de manière sûre du style de la lettrine. Cela peut s'expliquer par le fait que décrire une image par des motifs particuliers nécessite qu'elle contienne des motifs qui lui soient propres. Si une image ne contient que des motifs qui apparaissent dans plusieurs images de classes différentes, la distance entre cette image et les autres de la base ne permettront pas de définir clairement sa classe d'appartenance.

Une deuxième explication qui justifie ces taux provient du fait que l'on décrit une image à partir de statistiques issues des motifs qui la composent sans conserver leur position dans l'image. La perte d'information spatiale induit des erreurs et deux images symétriques (c'est à dire composées des mêmes motifs mais à des endroits différents) induit forcément des erreurs dans les résultats de classification. Ce premier résultat prouve que la méthode proposée permet de caractériser globalement les images de documents anciens, mais qu'une description basée uniquement sur une description statistique globale des images n'est pas suffisante.

6.2.2.2 Expérimentations avec des images de documents anciens

Afin d'étendre nos expérimentations, et de prouver que notre méthode peut-être appliquée à un type d'image plus générique, nous avons réalisé une seconde série d'expérimentations. Cette-ci porte sur une basse de 100 images appartenant à 9 classes différentes. Des exemples de ces classes est présenté dans la figure 6.2.4.

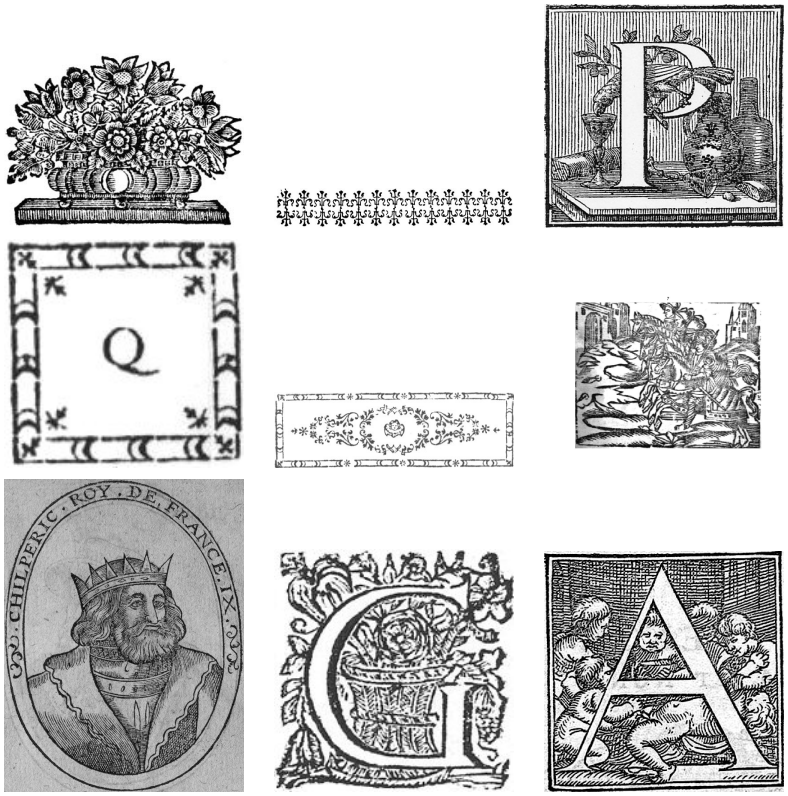


FIGURE 6.2.4: Exemples d'images de documents anciens composées de formes, appartenant aux 9 classes utilisées dans la seconde évaluation

Nous avons appliqué le même protocole expérimental sur cette seconde base d'image, et nous avons également utilisé un classifieur de type $K - PlusProchesVoisins$ avec un vote à la majorité pour les valeurs de K supérieur à 1. Les taux de reconnaissances obtenus sont présentés dans le tableau 6.3.

K	Taux de reconnaissances
1	62 %
3	83,3 %
5	85,6 %

TABLE 6.3: Taux de reconnaissances obtenus sur la base d'images diverses de documents anciens

Nous présentons également la courbe rappel-précision obtenue sur cette base dans la figure 6.2.5. Comme nous pouvons le constater, les premiers éléments renvoyés par notre système sont très pertinents. De plus, si nous choisissons de renvoyer un ensemble de 30 images à l'utilisateur final, on peut observer que le système est capable d'obtenir une

précision supérieure à 75 %, c'est à dire qu'en moyenne il n'y a que 7 mauvaises images sur 30.

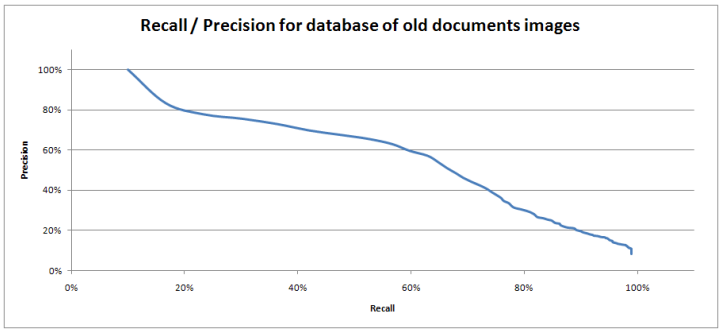


FIGURE 6.2.5: Courbe rappel-précision obtenue sur la base d'images diverses de documents anciens

En conclusion, on peut dire qu'analyser les images de documents anciens à partir d'une approche statistiques spécifiques aux formes donne de bons résultats. Cependant, cette approches ne permet pas de connaître précisément les éléments qui composent une image (où se situent les formes ? Quelle est leur aspect ?). Afin de répondre à ce problème, nous avons proposé d'extraire les formes prégnantes des images et de les décrire (d'après leur aspect et leur agencement topologique). Nous présentons l'évaluation relative à cette approche ci-dessous.

6.2.3 Évaluation de la description obtenue à l'aide des graphes de régions

Nous venons de voir qu'une approche statistique par recherche des motifs les plus pertinents permet d'obtenir des résultats corrects. Cependant, ce type d'approche n'intègre pas les informations sur les régions qui composent l'image, et sur leurs relations topologiques. Nous avons donc extrait des régions de la couche forme, et nous avons décrit leur position et leur taille.

Afin de valider l'approche proposée pour extraire les régions uniformes des lettrines, ainsi que le choix de nos caractéristiques, nous avons cherché à comparer les images pour rechercher des images similaires. De nombreux travaux de la littérature actuelle comparent des images en comparant les graphes qui leurs sont associés. On peut ainsi citer les travaux réalisés par [Jouili 08, Sidere 09, Raveaux 10, Jouili 10, Raveaux 11, Bunke 11] qui associent à chaque noeud du graphe une région, et comparer deux images revient à rechercher des graphes similaires. Dans notre cas, nous nous sommes appuyés sur une méthode développée dans le cadre du projet Navidomass par Salim Jouili [Jouili 09b, Jouili 09a, Jouili 10], qui consiste à résoudre le problème d'approximation de la distance d'édition entre graphes par la moyenne des signatures des noeuds. À chaque noeud, nous associons un ensemble de valeurs qui sont :

- le degré du noeud
- les valeurs associées au noeud
- les valeurs des noeuds adjacents
- le degré des noeuds adjacents : pour donner un poids important à l'information structurelle.

Soit un graphe $G = (V, E, \alpha, \beta)$, la signature $\gamma(n_i)$ associée à chaque noeud $n_i \in V$ est définie comme suit :

$$\gamma(n_i) = \left\{\alpha_i\, \theta(n_i),\, \{\theta(n_j)\}_{\forall i,j \in E},\, \{\beta_{i,j}\}_{\forall i,j \in E}\right\} \tag{6.2.1}$$

avec :

- α_i les caractéristiques de la forme associée au noeud n_i ;
- $\theta(n_i)$ le degré du noeud n_i ;
- $\{\theta(n_j)\}_{\forall i,j \in E}$ l'ensemble des degrés des noeuds adjacents au noeud n_i ;
- $\{\beta_{i,j}\}_{\forall i,j \in E}$ l'ensemble des attributs des arcs connectés au noeud n_i.

À partir du graphe ainsi obtenu, il est indispensable de définir une distance qui permette de comparer ces signatures, et les graphes. Comme présenté dans [Jouili 10], cette distance est basée sur la mesure de recouvrement euclidienne hétérogène (Heterogeneous Euclidean Overlap Metric - HEOM) qui permet d'associer dans une même mesure des attributs numériques et symboliques. Après quoi, en utilisant ces signatures de noeuds et la distance ainsi définie, le problème de distance d'édition entre graphe peut-être reformulé en problème d'appariement de graphes qui peut être résolue par la méthode hongroise [Kuhn 10].

Afin d'expérimenter l'approche que nous proposons, nous avons cherché à évaluer la pertinence du graphe extrait pour chaque image. Étant donné que cette évaluation doit permettre de mettre en avant la pertinence des régions extraites et les qualités topologiques des graphes générés, nous avons limité les attributs des noeuds à la position du centre de gravité de la forme, et à son aire. Enfin, les arcs des graphes sont pondérés par la distance euclidienne qui sépare les centres de gravité des formes.

Nous avons utilisé la même base de 358 lettrines (voir section 6.2.2.1 pour les détails) que celle utilisée dans la première série d'expérimentations, afin de pouvoir comparer les deux approches. Nous avons donc extrait les régions de formes pour l'ensemble de la base, décrit ces régions, puis nous avons généré un graphe par lettrine. Pour chaque lettrine, nous calculons la valeur de la distance définie ci-dessus avec toutes les autres lettrines de la base. Enfin, nous appliquons l'algorithme des K-Plus Proches Voisins pour vérifier que notre description permet bien de faire la distinction entre les styles de lettrines. Tout comme pour l'évaluation précédente, nous appliquons un mécanisme de vote à la majorité pour les valeurs de K supérieurs à 1. Les résultats obtenus sont présentés dans le tableau 6.4.

On peut remarquer immédiatement que même si les résultats au rang 1 ne sont pas très bons pour les styles 2 et 3, dès que l'on augmente le nombre d'éléments renvoyés, les

K	Total
1	73,7 %
3	92,5 %
5	95,8 %

TABLE 6.4: Taux de reconnaissances obtenus sur la base de lettrines utilisées dans [Pareti 06a] en utilisant une signature structurelle sur les zones de formes, et une distance hybride entre graphes

scores augmentent. De plus, dès que notre système renvoie plus d'une image en réponse à une requête, on peut remarquer qu'il n'y a que peu d'erreurs. En moyenne, lorsqu'un utilisateur soumet une requête au système, il n'y a qu'une petite dizaine d'images, sur les 358, pour lesquelles la majorité des images renvoyées n'est pas du même style.

Il faut noter le cas particulier des images du style 2 qui n'obtient pas de bons résultats au rang 1 (plus de la moitié des images n'est pas affectée à la bonne classe). Ceci peut s'expliquer par le fait que les images de ce style ne sont composées quasiment que de zones uniformes. Du coup, de nombreuses formes sont extraites et il devient difficile de distinguer une image du même style (c'est à dire avec autant de formes avec des positions et des tailles de formes proches) d'images de styles différents. En effet, les autres images peuvent être composées de formes similaires (la même lettre par exemple, des personnages, ...) qui correspondent à un sous-graphe quasiment identique. Comme pour des valeurs de K supérieures à 1, nous procédons par un vote à la majorité, plusieurs erreurs apparaissent. Cependant, dès que K devient supérieur ou égale à 5, on peut remarquer que le système renvoie une majorité d'image associée au style 2.

Enfin, décrire l'agencement relatif des régions de formes qui composent l'image parait intéressant, mais souffre du manque de descriptions de l'aspect des régions en elles-mêmes. C'est pour cela que nous proposons dans la section suivante de complexifier les description en mélangeant l'approche statistique et l'approche structurelle.

6.2.4 Complexification des descriptions statistiques et structurelles

Nous venons de présenter des résultats de reconnaissance de lettrines en utilisant deux approches différentes. La première décrit statistiquement les images dans leur ensemble en recherchant des motifs spécifiques à chacune d'entres-elles. La seconde approche décrit les images en recensant l'ensemble des régions qui la composent, et leur position. Nous pouvons constater que les deux approches donnent des résultats corrects, mais il semble intéressant de combiner les résultats statistiques avec les résultats structurels pour obtenir une description complexe de la couche forme.

Il est possible de combiner des caractéristiques ou des résultats obtenus via deux modalités différentes selon deux types de fusion : la fusion précoce et la fusion tardive [Chowdhury 10]. La fusion précoce consiste à concaténer les vecteurs de descripteurs, et

à utiliser un classifieur unique pour classer les éléments. La fusion des information a donc lieu en amont de la classification. La fusion tardive, quant à elle, consiste à fusionner les scores correspondants à chaque modalité. Cela implique donc l'utilisation d'autant de classifieurs que de type de description, et la décision finale revient à fusionner les scores des différents classifieurs. Cette fusion est réalisée à l'aide d'une fonction explicite, ou d'un classifieur supplémentaire.

Dans notre cas, nous cherchons à combiner deux informations de types différents (un vecteur statistique comparable à l'aide d'une métrique adaptée, et des graphes topologiques des régions). Réaliser une fusion précoce est complexe puisque celle-ci manipulerait des données hétérogènes, et que cela pré-supposerait une homogénéisation (par exemple à l'aide de graphe probing) pour se ramener à des choses comparables. Le temps ne nous a pas permis d'explorer cette voie.

Nous nous sommes donc orientés vers une fusion tardive, afin de rechercher dans une banque d'image l'ensemble des images ressemblantes à la requête. Nous réalisons donc un vote à la majorité entre les résultats obtenus avec l'approche statistique et l'approche structurelle. Combiner ces deux informations nous permet de combiner des informations statistiques qui renseignent sur les motifs contenus dans l'image, avec des informations structurelles qui nous renseigne sur l'organisation spatiale des formes de l'image. Cela nous a permis d'obtenir les résultats de classification par vote à la majorité présentés dans le tableau 6.5.

K	Total
1	90,8 %
3	96,4 %
5	98 %

TABLE 6.5: Taux de reconnaissance des lettrines à partir de la couche forme en utilisant une description recomplexifiée des lettrines

On observe immédiatement que la combinaison d'une information statistique et spatiale augmente considérablement les taux de reconnaissances que l'on peut obtenir. Pour K=1, la combinaison des descriptions augmente de plus de 40 % les résultats comparativement à l'approche statistique seule.

L'augmentation des taux de reconnaissance prouve bien que le fait de combiner une information statistique et spatiale permet de complexifier la description de l'image, et d'obtenir des résultats nettement meilleurs. Ainsi, dès que notre système renvoie plus d'une image, on peut constater qu'en moyenne au plus une image parmi les réponses proposées n'est pas du même style.

Synthèse du chapitre

Ce chapitre traite de l'analyse de la couche forme de la décomposition de Meyer. Il se décompose en deux parties : la première présente comment nous venons extraire de l'information de cette couche, pendant que la seconde présente les différentes évaluation menées pour valider nos choix. L'extraction d'information reste conforme au modèle présenté précédemment, puisque nous venons tout d'abord extraire des motifs de formes qui permettent de représenter les formes recherchées par les utilisateurs. À partir de ces motifs, nous venons définir deux signatures pour caractériser les formes : une première statistique qui repose sur l'utilisation de sacs de motifs, et une seconde qui se sert des régions extraites pour définir leur agencement topologique. La seconde partie quant à elle présente tout d'abord une comparaison entre les régions extraites et une vérité terrain. Puis, dans un second temps, l'approche statistique est évaluée premièrement sur une base de lettrines, puis sur une base variée d'images de documents anciens, pour prouver sa pertinence. Enfin, une dernière évaluation sur la même base de lettrines permet de mettre en avant les capacités de la signature structurelle.

Nous venons de voir au travers des différentes évaluations menées sur la couche forme que les approches que nous proposons permettent d'obtenir des résultats corrects mais non pleinement satisfaisants. Cela permet de vérifier que l'analyse d'images complexes ne peut pas s'appuyer sur un seul type d'information. Cela nous ramène donc à notre stratégie de départ, qui consiste à étudier les images complexes à partir de différentes couches d'informations, d'analyser chacune d'entre elles, de les caractériser, pour au final re-complexifier la description. C'est ce que nous présentons dans la suite de ce manuscrit.

Chapitre 7

Cas de la couche texture

7.1 Extraction d'information à partir des traits présents dans les images

Ce travail s'intéresse aux images de documents anciens, et plus particulièrement aux images des documents du début de l'imprimerie. A l'époque, les documents étaient imprimés par pression d'un tampon en bois encré sur le papier. De part la nature des tampons (surface en bois composée de creux et de bosses), il n'était pas possible de définir de nuances de gris (indispensables pour la création d'effets de perspectives). Afin de palier à ce manque, les imprimeurs de l'époque ont utilisé des traits parallèles rapprochés pour donner une impression de nuances. Il paraît donc cohérent d'étudier ces traits pour reconnaître et identifier les images issues de documents anciens.

Nous rappelons que dans le cadre de ce manuscrit, nous nous intéressons plus particulièrement aux lettrines qui correspondent à des images très utilisées dans les ouvrages et très réutilisées au cours du temps. Une lettrine peut être vue comme une image binaire composée de traits. Quelques travaux [Bigun 96, Pareti 06b, Uttama 06, Journet 08c] ont été menés pour caractériser statistiquement des lettrines. Cependant, bien que ces traits soient caractéristiques des images de documents anciens, aucun travail à notre connaissance n'a été dédié à leur étude pour le moment.

Le modèle d'analyse complexe que nous avons défini dans le chapitre 5, repose sur une analyse des images de lettrines selon deux couches d'informations spécifiques. La première, qui repose sur l'analyse des formes qui composent les images, a été présentée dans le chapitre précédent. De manière complémentaire, nous proposons dans ce chapitre d'explorer une deuxième approche pour la caractérisation des lettrines reposant sur l'analyse des traits qui les composent. Suivant notre modèle, il convient d'appliquer une stratégie consistant à repenser l'algorithme classique d'analyse des images, qui s'appuie généralement sur des informations pixellaires, pour l'adapter aux images de traits. La méthodologie retenue consiste donc à considérer que l'information élémentaire n'est plus

le pixel, mais le trait. Le postulat amène donc à repenser toute la stratégie d'analyse en intégrant cette propriété particulière de nos images. Suivant notre modèle, les étapes de réalisation pour la couche textures sont : 1) extraction de motifs adaptés à la couche (à savoir les traits) ; 2) définition d'une description pertinente liée au contenu ; 3) Extraction de régions d'intérêts dont on peut caractériser le contenu (description des éléments extraits) et leur agencement (contextualisation de ces éléments).

Naturellement, ce chapitre s'articule de même. Après un bref état de l'art des approches dédiées aux images de traits, nous présentons notre méthode d'extraction de traits dans les lettrines. Cette méthode comprend quatre étapes : pré-traitements pour améliorer l'image, extraction des traits par une squelettisation, caractérisation des traits, et enfin réduction du nombre de traits possible dans chaque image. À partir des traits que nous avons extraits, nous proposons une description pertinente de la couche textures via une approche similaire à celle appliquée sur la couche forme (voir section 6.1.2), c'est à dire que nous décrivons les images de traits en se basant sur une approche par sacs de traits [1]. Enfin, nous venons extraire des régions de traits pour permettre une description complémentaire de l'image à l'aide d'une signature structurelle. Finalement, nous terminons par une évaluation des différentes étapes de notre approche en suivant une démarche d'évaluation similaire à celle utilisée sur la couche forme (évaluation de l'extraction des traits, évaluation de l'approche statistique, et évaluation de l'approche structurelle).

7.1.1 Extraction des traits : Un bref état des lieux

Nous avons vu dans la section 1.4.1.3, que les traits contenus dans les lettrines peuvent être vus comme des textures. Les textures sont des informations difficiles à extraire, et plusieurs méthodes de la littérature proposent de les classer et de les segmenter. Bien que nous ayons déjà traité l'extraction de traits et de textures dans la littérature, nous présentons un bref résumé de celui-ci ci-dessous.

Comme nous le précisions précédemment, les traits n'ont pas été étudiés dans la littérature à notre connaissance. Ces traits peuvent être vus comme des textures qui composent les images. Ces traits présentent des longueurs différentes (long pour représenter le fond, courts pour représenter les ombres de personnes ou d'objets), différentes orientations et sont situés à différents endroits dans l'image.

Bien que les textures soient des informations difficiles à extraire, plusieurs méthodes existent dans la littérature, et la classification et la segmentation de textures reposent sur l'extraction de caractéristiques qui leurs sont propres. La plupart des caractéristiques extraites pour représenter une texture sont définies à partir du niveau de gris des pixels. Comme présenté en détails dans la partie 1.4.1.3, les cinq principales méthodes de la

1. approche similaire aux sacs de mots utilisés dans la fouille de texte

littérature sont :

1. Les méthodes à base de noyaux ;
2. Les techniques à base de descripteurs ;
3. Les méthodes qui reposent sur des filtres particuliers ;
4. Les méthodes qui s'appuient sur la corrélation et l'auto- corrélation ;
5. Les approche qui consistent à segmenter l'image en zones homogènes.

Cependant, toutes ces méthodes présentent l'inconvénient de travailler au niveau du pixel et non du trait. Bien que celles-ci permettent une description fine des images, elles sont par ailleurs très sensibles aux dégradations qui apparaissent dans les documents anciens (bruit dû au vieillissement, à la numérisation, ...). De plus, partant de l'idée que les lettrines sont des images composées de traits, il nous parait essentiel de décrire les éléments qui ont permis de les créer. Pour ces raisons, nous avons défini notre méthode d'extraction et de caractérisation de traits.

7.1.2 Extraction des traits : Approche proposée

Afin d'extraire les traits dans les images de lettrines, nous proposons une approche qui repose sur une série de quatre traitements détaillés ci-dessous. Dans le but de conserver la cohérence globale de notre approche, tous ces traitements sont appliqués sur la couche texture obtenue via la décomposition de Meyer (voir section 5.2.1.3 pour plus de détails). Pour rappel, cette décomposition rassemble au sein d'une couche dénommée *"couche texture"* tous les éléments qui oscillent rapidement dans une image, dans notre cas les traits des lettrines. Les quatre étapes de l'approche proposée correspondent à quatre étapes classiques des approches de traitements des images : Pré-traitements, segmentations, caractérisation, fusion des éléments communs.

7.1.2.1 Pré-traitements : binarisation et débruitage

Binarisation

Nous rappelons que cette partie de l'approche porte sur la couche texture issue de la décomposition de Meyer. Celle-ci permet de mettre en exergue les zones de traits des images de documents anciens. De manière formelle, cette couche est issue de l'espace de Banach, qui correspond à l'espace des fonctions oscillantes, c'est à dire présentant un motif répétitif. D'un point de vue réalisation, cette couche est obtenue via l'utilisation d'un gradient. Ainsi, si un pixel de l'image de la couche texture correspond à une zone lisse, il aura une valeur centrée en 0, tandis qu'un pixel correspondant à une zone texturée aura pour valeur la différence de luminosité avec ses voisins. Afin de pouvoir représenter les fronts montants et descendants des fonctions oscillantes (valeurs négatives et positives du signal), cette image est recentrée sur 128, d'où son aspect grisâtre.

Dans notre cas, nous nous intéressons uniquement au fait d'avoir une information d'oscillation sans prendre en compte le fait qu'un pixel corresponde à un front montant ou descendant. Afin de binariser les images, nous réorganisons donc ces valeurs afin d'avoir les pixels des zones lisses à 0, et l'information d'oscillation (les traits) en valeurs positives supérieures à 0, à l'aide de la fonction suivante :

$$\forall i \in [0, X], \forall j \in [0, Y] \; P'(i, j) = abs(P(i, j) - 128) \tag{7.1.1}$$

avec X et Y les dimensions de l'image à traiter, $P(i, j)$ le pixel à la position (i, j) de la couche texture, et $P'(i, j)$ la nouvelle valeur du pixel dans l'image binarisée. Une fois que les valeurs des pixels de l'image ont été réaffectée, il ne nous reste qu'à binariser l'image à proprement parler. Etant donné la variabilité d'acquisition de nos images et des dégradations qui agissent dessus, nous préférons utiliser un seuil adaptatif pour chaque image qui repose sur le critère développé par Otsu [Otsu 79].

Débruitage

De part les dégradations qui ont pu affecter les images au cours du temps et des différents traitements, une opération de nettoyage est nécessaire pour ne conserver que les pixels qui correspondent aux traits dans l'image binarisée. Dans notre cas, nous considérons comme bruit toute composante connexe de taille faible isolée. Afin de supprimer ce bruit, nous appliquons une ouverture morphologique sur l'image avec un masque carré de taille 5x5. Le choix de la taille du masque a été guidée par le fait que nous considérons que toute composante connexe composée de moins de 5 pixels ne peut être considérée comme fiable pas l'extraction des caractéristiques dans la suite de l'extraction.

La figure 7.1.1 présente les différentes étapes de ce processus sur deux exemples de lettrines.

7.1.2.2 Squelettisation

Nous pouvons remarquer dans les exemples présentés dans la figure 7.1.1 que les traits qui sont retenus après l'étape de pré-traitements sont d'épaisseur variable. Nous rappelons que nous cherchons dans ce travail à ne plus travailler au niveau du pixel, mais au niveau des traits. Afin de conserver une cohérence de description entre les différents traits et les différentes images, nous squelettisons les images binarisées puisque cela nous permet d'obtenir une information préciser sur la taille et la forme de chaque trait.

Plusieurs méthodes de squelettisation ont été proposées dans la littérature[Zhang 84, Guo 89]. Toutes ces approches présentent l'inconvénient d'être sensibles au bruit présent dans les images. Nous nous sommes orientés vers un algorithme de squelettisation basé sur une transformée en distance [McAndrew 04]. Outre le fait d'être plus robuste au bruit, cet algorithme présente également l'intérêt de donner directement l'épaisseur des squelettes, et par conséquent des traits. Différentes distances peuvent être utilisées pour

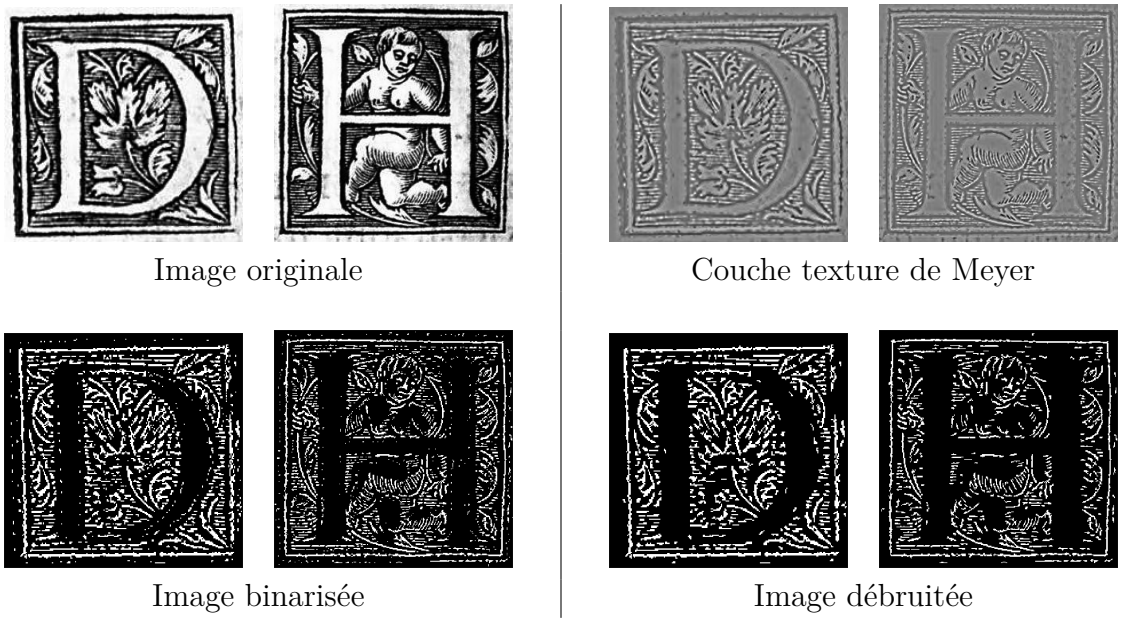


FIGURE 7.1.1: Exemples d'images de la couche texture obtenue à partir de deux lettrines. La première ligne présente l'image originale, la seconde la couche texture obtenue via la décomposition de Meyer, la troisième l'image binarisée, et enfin la dernière ligne présenté l'image débruitée

calculer la distance entre les pixels de l'image (distance de Manhattan, distance 3-4, ...). Nous avons utilisé une distance euclidienne, nous avons retenu les maxima de cette transformée pour obtenir le squelette.

Chaque composante connexe de ce squelette est alors considéré comme un trait de l'image. Nous allons maintenant les caractériser.

7.1.2.3 Caractérisation des traits

Caractériser des traits dans les images n'est pas chose courante. Ainsi, les approches les plus courantes de la littérature qui s'intéressent à des traits ont été appliquées sur des contours ou des silhouettes de formes. Un état de l'art des approches contours a été présenté dans la section 1.4.1.2. La grande majorité des approches existantes tentent de décrire l'allure du contour pour décrire une forme. Elles décrivent donc soit l'allure générale du contour (à l'aide de descripteurs statistiques), ou s'attardent sur une information intrinsèque au trait qui compose le contour (courbure, longueur, ou caractérisation des traits qui composent la forme). Les images de lettrines ont été créées par l'homme pour l'homme. Nous avons donc cherché à les étiqueter à l'aide de caractéristiques prégnantes pour la vision humaine [Livingstone 88, Graham 98, Soegaard 10], à savoir **la longueur, l'épaisseur**, **l'orientation** ou **la complexité**, et **les changements de directions** des traits.

Longueur

La longueur du squelette correspond au nombre de points qui le compose.

Épaisseur

est obtenue à partir de la transformée en distance de l'étape précédente. La squelettisation par transformée en distance utilisée [Rosenfeld 66, Breu 95] consiste à utiliser une norme infinie, et revient à ne conserver que les pics dans l'image transformée. Ces pics correspondent à la distance de chaque point du squelette au pixel de fond le plus proche. L'épaisseur du trait pour ce pixel est donc égale au double de sa valeur.

Complexité

La complexité d'un squelette doit permettre d'identifier le nombre d'embranchements qui le composent. Le coefficient de complexité C est donc obtenu avec la formule suivante :

$$C = \frac{X}{L} \tag{7.1.2}$$

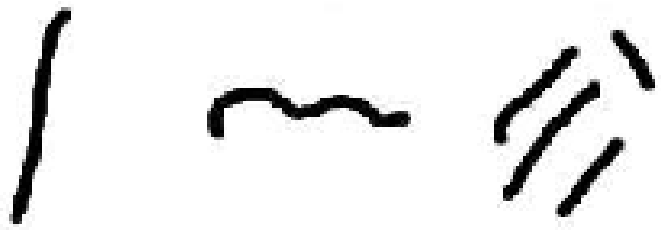
avec X le nombre de points d'embranchements du squelette, et L la longueur du squelette.

Orientation

Les traits extraits ne sont malheureusement pas des droites parfaites et l'extraction de leur orientation n'est pas simple (un trait compliqué peut raisonnablement avoir plusieurs orientations associées). Nous proposons donc une méthode d'extraction de l'orientation principale basée sur la transformation de Radon [**?**]. La transformée en Radon projette l'image sur un axe à différents angles. L'énergie de chaque axe représente l'importance donnée à une orientation. En ne conservant que l'axe de plus forte énergie, nous pouvons ainsi obtenir l'orientation principale de chaque trait. Nous conservons le sinus de cette orientation principale pour la normaliser. Ceci fonctionne aussi bien sur des images composées d'un ou plusieurs traits. Des exemples de résultats sont présentés dans la figure 7.1.2.

Changements de directions

Différents types de traits peuvent être observés dans les lettrines. Ainsi, certains seront rectilignes (comme les traits parallèles qui composent le fond), pendant que d'autres seront courbés (contours de l'image). Afin de différencier ces différents types de traits, il est important d'intégrer dans notre caractérisation des images une information sur les changements de directions. La méthode la plus répandue pour identifier les changements



(a) OP = 170 degrés (b) OP = 83 degrés (c) OP = 140 degrés

FIGURE 7.1.2: Orientation Principale (OP) obtenue pour chaque image

de direction du squelette le codage de Freeman [Freeman 61]. Partant d'un pixel, cette chaîne indique la direction à suivre pour atteindre le pixel suivant. Afin de la rendre invariante aux changements d'orientation et aux différentes longueurs de squelettes, nous avons utilisé une approximation polynomiale, avec un polynôme de degré 16, du codage de Freeman relatif. Le codage de Freeman relatif permet de définir l'orientation d'un pixel par rapport à son prédécesseur, tandis que l'approximation polynomiale résume le codage par aux 16 coefficients du polynôme utilisé. Nous avons sélectionné le degré du polynôme de façon à ce qu'il puisse coder de manière pertinente la variabilité des traits rencontrés dans les images de lettrines. Enfin, dans le cas ou le squelette est composé de plusieurs branches (voir figure 7.1.3 pour un exemple), nous utilisons la branche la plus proche de l'orientation principale et nous codons ses variations.



(a) Sous-partie d'une image composée d'un squelette à plusieurs branches (en rouge au milieu de l'image)

(b) Détection de la branche la plus proche de l'orientation principale pour le calcul du codage de Freeman

FIGURE 7.1.3: Exemple de squelette composé de plusieurs branches, et détection de la branche principale

7.1.2.4 Clustering intra-image

Une fois les traits caractérisés, nous pouvons remarquer qu'il existe de très nombreux traits dans les images (parfois plus de 1000 traits sont extraits dans une image). Cela

implique un grand nombre de caractéristiques pour chaque image et pour l'ensemble de la base, avec les problèmes dimensionnels qui y sont liés (temps de calculs, représentations creuses, faible représentativité de chaque type, ...).

De plus, si l'on observe une image de lettrine, ou plus généralement une image de document ancien, on peut remarquer qu'il n'existe qu'un nombre limité de traits (ceux-ci sont issus de tampons en bois). Il parait donc logique de chercher à regrouper les traits qui composent une image pour faire émerger les grandes tendances qui composent une image. C'est ce que nous réalisons avec ce clustering des traits de chaque image.

Le choix d'un algorithme de classification non-supervisé paraît naturel puisqu'il ne nécessite aucune connaissance a priori sur le nombre de classe de traits que l'on recherche dans les images. De plus, cet algorithme doit pouvoir être appliqué à plusieurs images composées de traits très différents (en nombre et en type).

De nombreuses méthodes de clustering existent dans la littérature [Theodoridis 08], comme le Classifieur Hiérarchique Ascendant (CHA), ou bien l'algorithme Espérance-Maximisation (E-M) définit par Dempster *et al [Dempster 77]*. L'intérêt principal de ce dernier repose sur le fait qu'il ne nécessite aucun paramètre (le CHA impose la définition d'un critère de coupe pour savoir à quel niveau de la hiérarchie il faut s'arrêter). Le principe de l'algorithme E-M est d'associer à un individu la classe dont il est le plus proche. A chaque fois qu'un individu est ajouté, les centres de gravités des différentes classes sont recalculés pour s'adapter à la nouvelle organisation de l'espace.

Après avoir normalisé toutes nos caractéristiques, nous avons utilisé une distance euclidienne entre les vecteurs des différents traits qui composent chaque image. À partir de ces distances, il nous est possible d'appliquer l'algorithme E-M et ainsi de créer des classes de traits propres à chaque image. Ce sont ces classes de traits qui vont nous servir de base pour définir des mots-visuels (pour l'approche statistique) pour l'ensemble de la base d'image, ou encore pour définir des zones de traits à extraire (dans l'approche topologique).

7.1.3 Approche par sacs de traits

Nous venons de voir que la première étape appliquée sur la couche texture nous permet d'extraire les traits des images, de les décrire, et d'en déduire les grands types de traits présents dans chaque lettrine. Par analogie avec la méthode présentée sur la couche forme, nous souhaitons décrire nos images de traits à partir des mots visuels, c'est à dire des motifs de traits les plus représentatifs de nos images.

Afin d'élaborer une méthode qui permette de décrire les images à partir des traits qui les composent, nous devons définir les mots visuels basés sur les traits les plus pertinents qui composeront le dictionnaire (c'est à dire la liste des classes de traits qui caractérisent les images). Cela consiste à utiliser une classification automatique sur les caractéristiques de traits des lettrines pour regrouper les traits similaires. Notre but est de constituer des groupes de vecteurs de caractéristiques pertinents qui représentent les différents types

de traits. Ces groupes seront considérés comme les motifs visuels à rechercher dans les images [Fei-Fei 09]. Nous avons ici la notion de "sacs de traits".

Pour regrouper les traits au sein de groupes qui partagent des vecteurs descriptifs semblables, nous utilisons le même algorithme de classification non-supervisée qu'à l'étape précédente. Les classes de traits obtenues constituent le dictionnaire de mots visuels, c'est à dire les sacs de traits qu'il faut rechercher dans les images. La description de chaque lettrine correspond alors à l'histogramme des fréquences de chaque mot du dictionnaire. Comparer deux image consiste à calculer la distance entre les histogrammes des images de la base. Le schéma de la figure 7.1.4 présente cette démarche dans son ensemble.

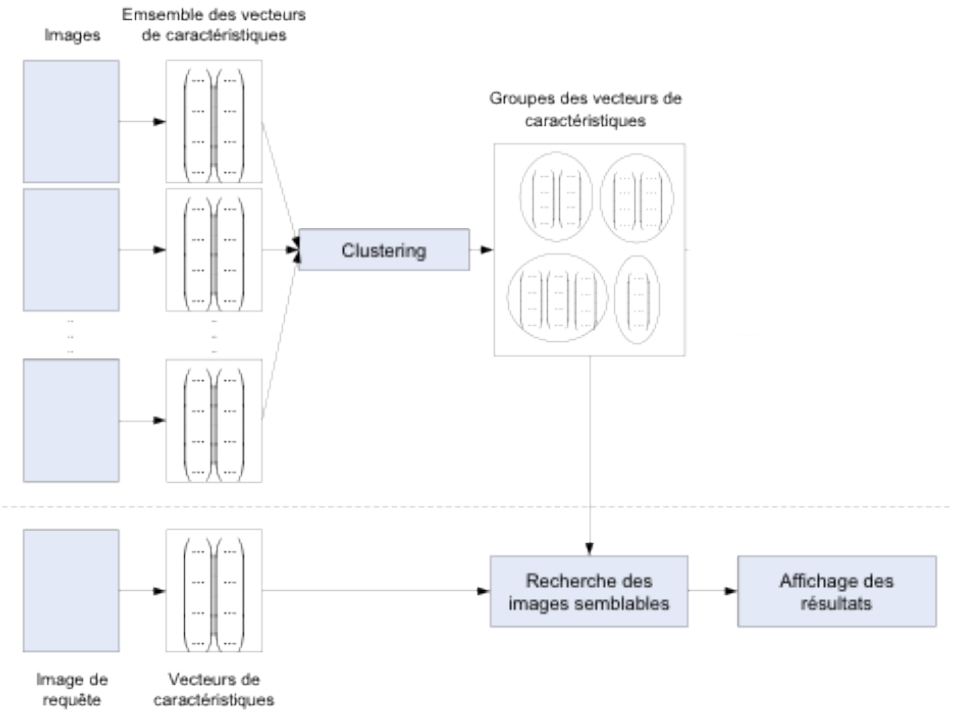


FIGURE 7.1.4: Schéma général de construction du dictionnaire des mots visuels de traits dans les lettrines (partie supérieure), et de la comparaison avec une image requête (partie inférieure)

Une fois les histogrammes définis pour chaque image de la base, comparer des lettrines revient à calculer des distances entre histogrammes. Quand un utilisateur fournit une image requête au système, celui-ci extrait les traits selon la même procédure que celle présentée dans la section précédente. Une fois les traits extraits, ils sont décrits à l'aide des caractéristiques utilisées sur les images de la base, puis chaque trait est associé à un mots-visuel de trait. Il devient aisé de calculer l'histogramme de l'image requête et de le comparer à ceux présents dans la base, et ainsi d'obtenir la liste des lettrines (ou des images de documents anciens) les plus ressemblantes à l'image requête.

7.1.4 Extraction de zones de traits et construction de graphes

Le modèle complexe d'analyse d'images de documents anciens que nous proposons depuis le début de ce manuscrit repose sur sa capacité à simplifier l'image en couche, et à caractériser l'information contenue dans chacune d'elles. Concernant la couche texture, nous venons de voir une méthode pour décrire les traits des images de documents anciens à partir de sacs de traits. Les images sont donc décrites statistiquement, à l'aide d'une signature spécifique du point de vue des traits qui les composent. Une simple connaissance statistique ne serait être suffisante pou décrire pleinement l'information contenue dans cette couche. C'est pourquoi nous proposons, en accord avec notre modèle, d'extraire des régions d'intérêts à partir de la couche texture. Ces régions doivent permettre, en complément de la première approche proposée, de décrire cette couche d'un point de vue structurelle. De plus, ces régions sont à la base de la connaissance que nous souhaitons intégrer dans l'ontologie qui doit permettre de réduire le fossé sémantique entre les utilisateurs et les résultats de traitements des images.

Ainsi, dans la suite de cette section, nous présentons comment nous venons extraire des zones de traits dans les images, et comment nous analysons leur organisation spatiale. Cette description est composée de trois étapes : regrouper les traits pour construire des régions, sélectionner les régions à retenir, et enfin générer le graphe topologique de ces régions.

7.1.4.1 Regroupement des traits pour la segmentation

Partant du clustering intra-image, effectué lors de la caractérisation des traits, chaque trait de l'image est attribué à une classe. Si l'on observe des exemples de lettrines hachurées dans la figure 7.1.5, on peut remarquer que des traits similaires sont généralement les uns à côté des autres. Si l'on regroupe tous ces traits d'une même classe au sein d'une même région spatiale, il nous est possible de déterminer une région de traits. Le but est donc de définir comment rassembler au sein d'un même groupe les traits spatialement proches appartenant à la même classe.

Pour y parvenir, nous proposons de définir le voisinage d'un trait. Dans une image I, le voisinage d'un trait correspond à la partie du fond adjacente à celui-ci. Afin de définir clairement le voisinage d'un trait, nous utilisons un algorithme de ligne de partage des eaux [Meyer 94], qui permet de trouver les frontières entre les voisinages des différentes traits. Si deux traits partagent le même voisinage (c'est à dire qu'ils ont une frontière en commun), et qu'ils appartiennent à la même classe, alors ils sont groupés au sein d'une même région qui contient les deux traits et leurs voisinages respectifs. Un exemple de résultats obtenus en utilisant cette approche est présenté dans la Figure 7.1.5.

Pour faciliter la visualisation des résultats, nous présentons dans la figure 7.1.6 quelques classes de traits. Ces images présentent uniquement les zones contenant des traits d'une classe. Les caractéristiques moyennes des traits contenus dans ces zones sont présentées en dessous.

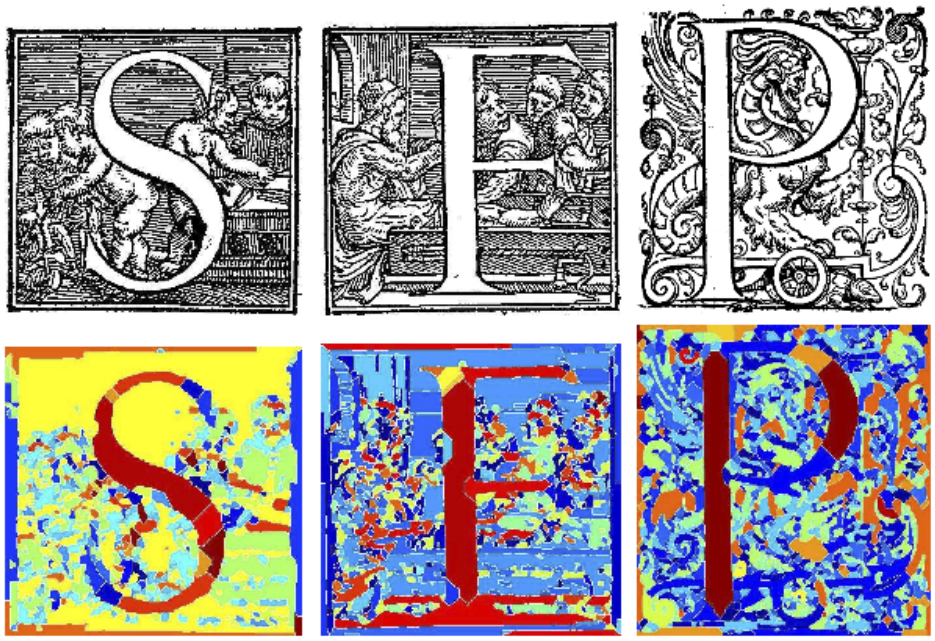


FIGURE 7.1.5: La première ligne présente l'image originale et la seconde présente les images segmentées où chaque couleur représente un rassemblement de traits appartenant à la même classe et proches spatialement (la même couleur n'est pas associée à la même classe dans toutes les images)

7.1.4.2 Sélection des zones

Une fois que les traits ont été regroupés au sein de zones, nous obtenons une image composées de régions. Nous rappelons que nous considérons qu'une image issue de la couche texture est composée entièrement de traits. Tous les pixels sont donc annotés en tant que traits, et comme appartenant à une région. Cependant, si l'on observe les exemples présentés dans la figure 7.1.6, on peut constater que certaines régions de traits ne sont composées que d'un trait, ou que des traits parasites peuvent apparaître (comme tout est considéré comme trait, une partie de lettre ou du fond peut apparaître).

Pour chaque région, nous calculons donc sa taille, le nombre de traits qui la composent, et un paramètre de solidité. Ce dernier paramètre correspond à la surface des traits qui compose une région par rapport à l'aire de la région. La combinaison de ces deux paramètres nous permet d'ignorer les régions qui ne comportent qu'un trait épais qui pourrait correspondre à un élément du fond, d'un personnage, ou d'une lettre. Une fois ces critères appliqués, nous obtenons les résultats présentés dans le tableau 7.1.

On peut donc constater que les traits de l'image sur la première ligne sont extraits correctement, tandis que pour l'image de la deuxième ligne, qui n'est composée d'aucun

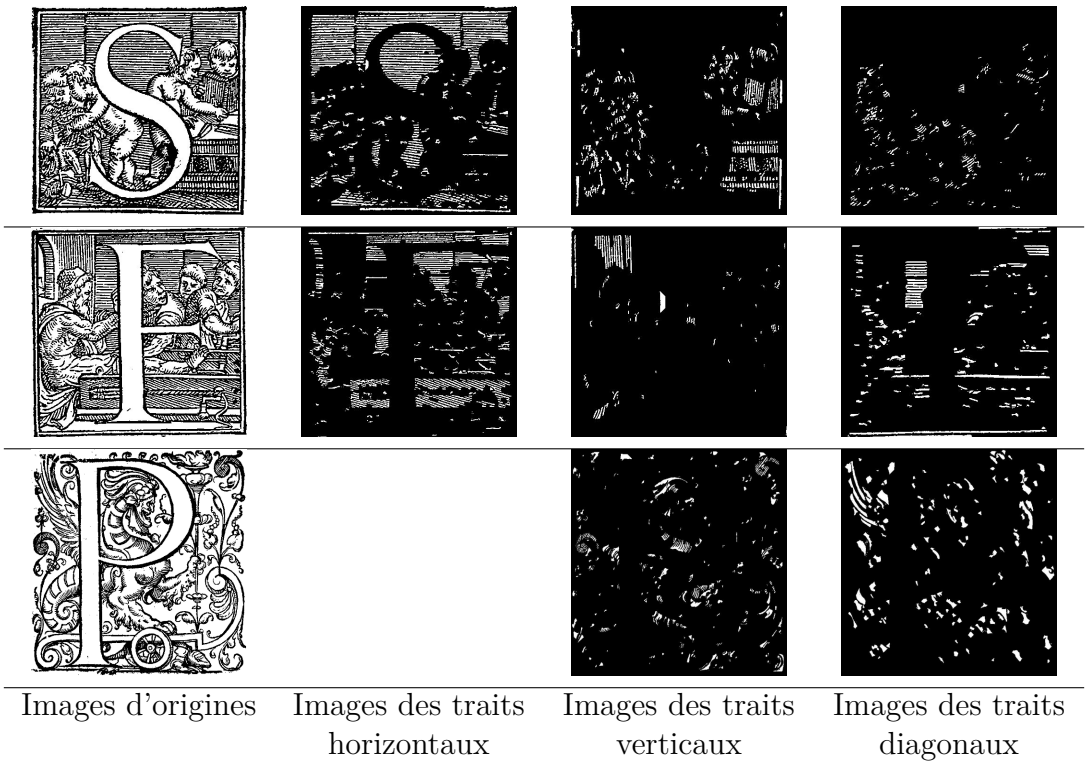


FIGURE 7.1.6: Exemple d'extraction et de regroupement de traits dans des lettrines

trait caractéristique, aucune zone n'a été détectée.

7.1.4.3 Génération du graphe

Une fois les zones sélectionnées, nous obtenons un ensemble de régions qui composent l'image. Tout comme nous l'avons réalisé pour la couche forme, nous cherchons à décrire l'organisation topologique des régions de nos images à l'aide d'un graphe. Le choix des graphes a été explicité dans la section 6.1.3.1.

À partir des régions qui ont été extraites, il nous est possible de construire un graphe de voisinage où les noeuds correspondent aux régions d'intérêts, et un arc relie une région à la région la plus proche. Chaque noeud est alors décrit par l'aire et les coordonnées du centre de gravité de la région, et chaque arc indique la distance entre les noeuds qu'il relie. Nous nous sommes appuyés sur cette démarche pour décrire les régions extraites par notre approche.

7.2 Validation de l'approche sur la couche traits

Afin de valider l'extraction des zones de traits dans les lettrines, nous avons décidé de mener deux séries d'expérimentations. La première série permet d'évaluer l'extraction

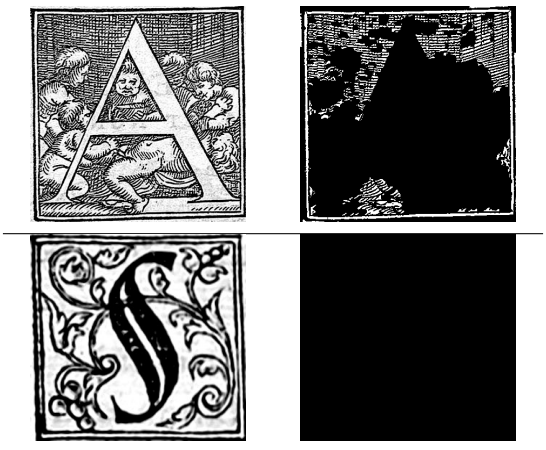


Image originale	Sélection basée sur l'ensemble des critères

TABLE 7.1: Exemples de résultats de sélection des zones de traits en utilisant différents critères. Les régions sélectionnées correspondent au sous-ensemble de région minimaliste qui contiennent des traits

des traits, en utilisant des experts pour valider la segmentation, et un critère de rappel. La seconde série d'expérimentations, quant à elle, permet de comparer notre système à une approche de la littérature, en utilisant des critères qualitatifs. Ces deux séries de tests sont présentées ci-après, et ont été réalisées sur une base de 228 images de lettrines composées de traits.

7.2.1 Validation de l'extraction des traits

Tout comme nous avons pu le faire avec la couche forme, nous présentons trois séries d'expérimentations qui visent à évaluer la pertinence de notre chaîne d'extraction de traits, et de leur caractérisation. La validation porte également sur trois grands thèmes : 1) validation de l'extraction des traits ; 2) Validation de la description des images par sacs de traits ; 3) Validation de la description par des graphes topologiques.

La première validation consiste à comparer notre approche avec une vérité terrain d'une part pour vérifier que notre approche est capable d'extraire les traits comparativement aux attentes des utilisateurs. D'autre part, nous avons cherché à comparer nos résultats avec une méthode très répandue d'extraction de textures dans les images, pour valider la volonté de travailler avec des éléments de plus haut niveau que le pixel.

La seconde série d'expérimentation consiste à évaluer le pouvoir discriminant des traits et des caractéristiques qui leurs sont associées. Ainsi, nous observons les taux de reconnaissances que nous obtenons en utilisant l'approche basée sur des sacs de traits. Enfin, la dernière série d'expérimentations est là pour nous permettre d'évaluer la qualité

de la discrimination de nos images en utilisant un graphe des régions de traits.

Évaluation de l'approche à l'aide d'une vérité terrain

Dans cette première série d'expérimentations, nous avons commencé par une évaluation subjective. Nous avons demandé à deux personnes d'observer les images et d'indiquer si elles étaient bien segmentées ou non, c'est à dire sous ou sur-segmentées. Les résultats obtenus sont présentés dans le tableau 7.2. Afin de permettre une compréhension plus intuitive de ces résultats, nous présentons dans la figure 7.2.1 des exemples de lettrines sous ou sur-segmentées. On peut ainsi remarquer que les erreurs de sur-segmentation peuvent être expliquées par le fait que certaines parties de visages ou le fond de la page qui entoure la lettrine sont considérés comme des traits. A l'opposé, on peut remarquer que certains traits en bas de l'image de la seconde ligne sont ignorés puisque nombre d'entre eux sont reliés et n'apparaissent pas comme des traits après le processus de classification.

	Images sous-segmentées	Images bien segmentées	Images sur-segmentées
Nombre d'image	16	193	19
% par rapport à la base	11,4%	84,6%	8,3%

TABLE 7.2: Nombre d'images bien, sous, ou sur segmentées sur un sous-ensemble de 228 images

Une seconde évaluation a été menée pour valider plus précisément les résultats de la segmentation en utilisant un critère objectif. Le but de cette évaluation est d'obtenir une information quantitative sur les résultats en utilisation un critère semblable à ceux utilisés dans l'évaluation des moteurs de recherche (rappel-précision). Nous avons donc généré une vérité terrain manuelle sur un sous-ensemble de 45 images de lettrines, où l'utilisateur devait dessiner des rectangles ou des ellipses pour sélectionner des zones de traits.

Un exemple de masques obtenus à partir de cette vérité terrain peut-être observé dans la figure 7.2.2. Dans ces masques, un pixel noir correspond à une région composée de traits pendant qu'un pixel blanc correspond à une zone qui ne contient pas de traits.

Dans notre cas, les images issues de la vérité terrain sont composées de rectangles ou d'ellipses. Nous nous sommes donc naturellement basés sur les régions de traits extraites dans la section 7.1.4 pour comparer la qualité d'extraction de nos régions à celle de la vérité terrain. Afin d'évaluer cette extraction, nous nous sommes basés sur les critères que nous avons définis au cours de l'évaluation de la couche forme.

Pour rappel, les critères que nous avons utilisés sont le **rappel** et la **précision**. Le premier critère permet d'indiquer le pourcentage de régions issues de la vérité terrain qui



Image d'origine | Résultat sous-segmenté

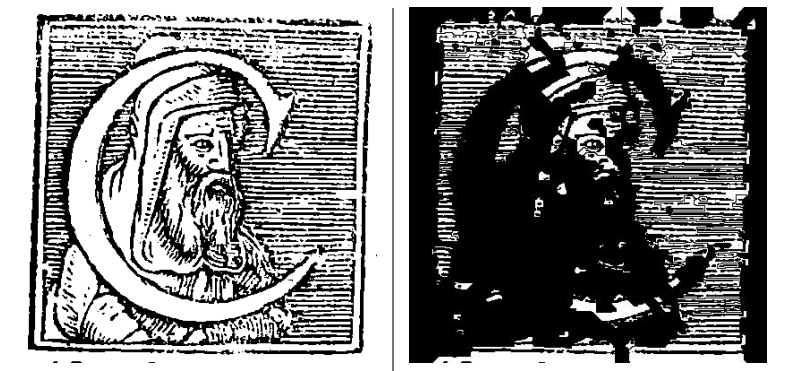


Image d'origine | Résultat sur-segmenté

FIGURE 7.2.1: Exemples de lettrines sous et sur-segmentées

contiennent des régions de traits extraites automatiquement. Une valeur égale à 100% signifie que pour toutes les régions de la vérité terrain, des régions de traits ont été extraites par l'approche décrite ci-dessus. A contrario, une valeur de rappel égale à 0 signifie qu'aucun trait n'a été extrait dans les régions marquées par la vérité terrain. La précision, quant à elle, permet de connaître le pourcentage de surface de la vérité terrain qui est recouverte par des régions de traits extraites automatiquement.

Le tableau 7.3 présente les valeurs de rappel et de précision obtenues sur un sous-ensemble annoté de 45 images. On peut remarquer que notre système obtient un très bon résultat en rappel, puisque l'on peut constater que pour quasiment toutes les zones marquées dans la vérité terrain, des traits ont été extraits. La seconde valeur, la précision, est plutôt bonne puisque l'on peut l'interpréter comme suit : un quart de la surface des zones marquées dans la vérité terrain est recouverte par les traits extraits. Finalement, à partir de cette première série d'évaluation, nous pouvons déduire que notre système est capable d'extraire les traits dans les lettrines. Cependant, afin d'être exhaustif dans l'évaluation de notre approche, nous proposons une comparaison de celle-ci avec une approche connue de la littérature d'extraction de textures : les filtres de Gabor.

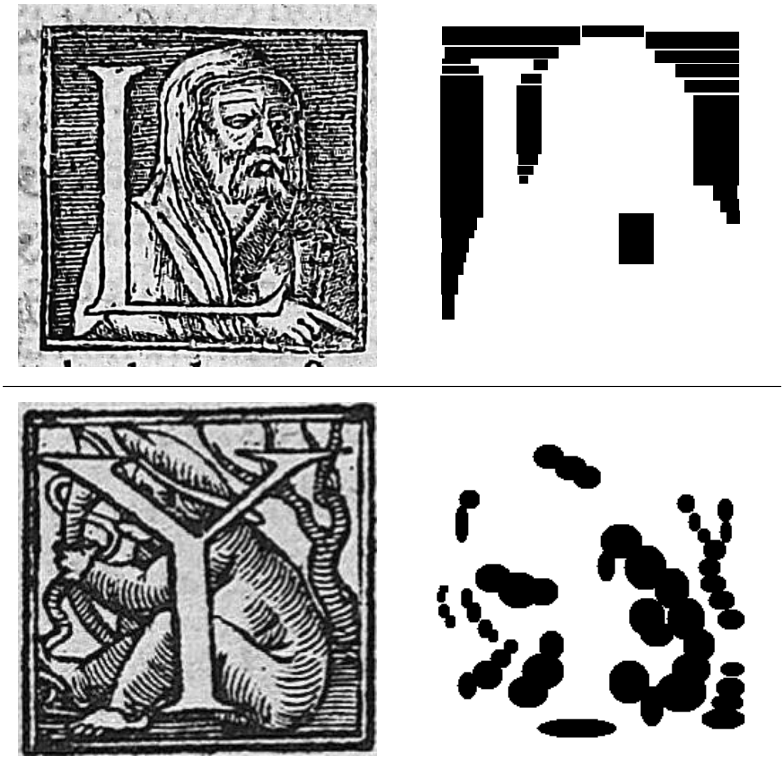


FIGURE 7.2.2: Exemple de lettrines avec les masques obtenus après saisie de la vérité terrain. Les masques (tâches noires) correspondent aux zones de traits dans l'image

Comparaison avec une approche de la littérature : les filtres de Gabor

La seconde série d'expérimentations a été menée dans le but de comparer notre approche avec une méthode célèbre en segmentation de textures, les filtres de Gabor [Jain 91, Teuner 95, Tanaka 04]. Un filtre de Gabor est un filtre linéaire dont la réponse impulsionnelle est une sinusoïde modulée par une fonction gaussienne. Dans le domaine spatial, un filtre de Gabor est le produit d'une sinusoïde complexe et d'une enveloppe Gaussienne :

$$g(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} \times \exp\left(-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right)\right) \times \exp\left(2j\pi f_0(x\cos\theta + y\sin\theta)\right) \quad (7.2.1)$$

où σ_x et σ_y sont les écart-types respectivement le long des axes x et y de la Gaussienne, f_0 la fréquence de modulation des fonctions sinus et cosinus, et θ l'angle entre l'axe des abscisses et l'orientation du filtre.

Pendant le processus de segmentation, les caractéristiques extraites résultent de la convolution de l'image avec un ensemble de filtres de Gabor. À chaque filtre sont associées

	Rappel	Précision
Approche proposée	100 %	71,3 %

TABLE 7.3: Rappel et précision obtenus avec notre méthode d'extraction de traits sur un sous-ensemble de 45 lettrines accompagnées de leur vérité terrain. Le critère de rappel permet de mettre en avant la capacité de notre approche à extraire les traits marqués dans la vérité terrain, pendant que la précision permet de connaître le pourcentage de surface des régions de la vérité terrain qui est couverte par les traits extraits automatiquement

une fréquence et une orientation propres qui permettent la détection de textures (des traits dans notre cas). Pour comparer notre approche avec les filtres de Gabor, nous avons choisi d'orienter les filtres à 0°, 45°, 90°, et 135° puisqu'ils correspondent aux orientations de traits que l'on peut trouver dans les images. À partir de ces filtres, une étape de segmentation basée sur un K-means [Lloyd 82] en utilisant les caractéristiques de fréquences et d'orientations permet la segmentation de l'image. Dans cette évaluation, nous avons fixé $K = 5$ car cela correspond aux régions des quatre orientations citées ci-dessus, plus une cinquième région qui correspond aux zones sans traits.

Les images présentées dans la figure 7.2.3 correspondent aux résultats de classification avec l'approche proposée et l'approche reposant sur des filtres de Gabor. On peut voir que l'approche proposée permet d'obtenir des zones de traits plus homogènes que celles obtenues avec les filtres de Gabor. Dans cette seconde approche, le fond est séparé en plusieurs parties et associé à trois types de textures (qui apparaissent en bleue, blanc et rouge dans les images ci-dessous). Avec l'approche proposée, nous obtenons visuellement de meilleurs résultats puisqu'une seule texture est associée au fond. A contrario, la segmentation réalisée à partir des filtres de Gabor permet une meilleure segmentation des zones lisses de l'image (comme c'est le cas avec la lettre F ou T), mais ne différencie pas de manière convaincante les différents éléments lisses (la lettre et les parties des corps sont identifiés comme appartenant à la même classe). Ceci peut s'expliquer par le fait que les filtres de Gabor nécessiteraient d'être paramétrés de manière fine pour pouvoir correspondre à chaque type de trait que l'on rencontre dans les lettrines. Cependant, rien ne garantit que ce paramétrage serait générique, et il faudrait certainement adapter ces paramètres pour chaque image. Enfin, dans le cas ou les traits se sont détériorés dans le temps, l'espace entre les traits n'est plus aussi régulier et plusieurs traits peuvent se toucher. Par exemple dans la figure 7.2.4, on peut voir au milieu de la lettrine que des traits sont composés de deux orientations. Là ou notre approche sera tolérante et arrive à identifier des traits complexes, une approche par filtre de Gabor ne permettra pas une description aussi fine et robuste (les filtres ne sont pas conçus pour répondre à plusieurs orientations en même temps).

De plus, comme notre approche commence par squelettiser l'image d'origine, elle n'omet quasiment aucun trait contrairement aux filtres de Gabor. Un exemple illus-

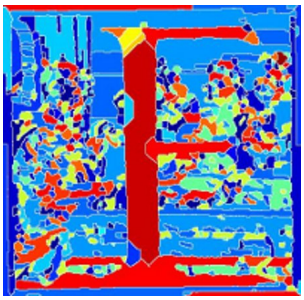
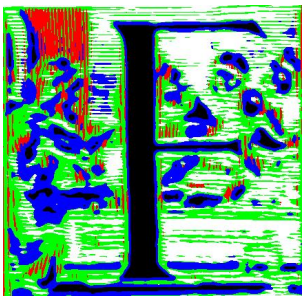
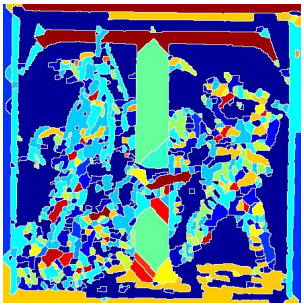
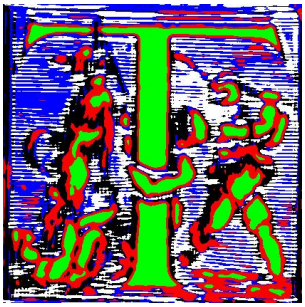
Image originale	Résultat de la segmentation avec l'approche proposée	Résultat de la segmentation avec l'approche utilisant les filtres de Gabor
		
		

FIGURE 7.2.3: Comparaison entre les segmentations obtenues avec l'approche proposée et une segmentation avec des filtres de Gabor. Chaque couleur correspond à une classe de trait dans l'image (une même couleur dans deux images différentes ne représente pas forcément la même classe)

trant ces propos peut être trouvé dans la figure 7.2.4, puisque les zones correspondant aux traits horizontaux (traits décrivant le fond de la lettrine) ne sont pas catégorisés dans la même classe.

Enfin, afin d'obtenir une comparaison quantitative entre les deux approches, nous avons comparé les images segmentées à la vérité terrain définie précédemment. Nous avons donc obtenu des images de régions composées des traits issus de notre approche et des filtres de Gabor. Des exemples peuvent être observés dans la figure 7.2.5.

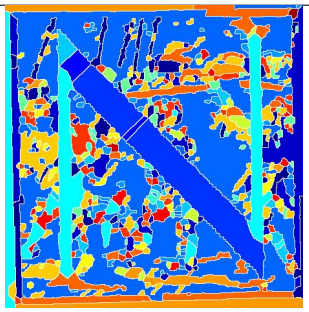
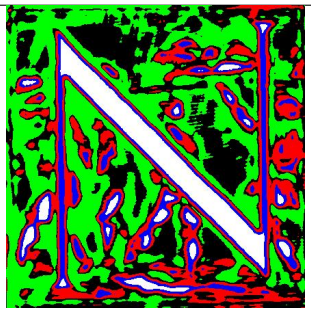
Image originale	Résultat de la segmentation avec l'approche proposée	Résultat de la segmentation avec l'approche utilisant les filtres de Gabor
		

FIGURE 7.2.4: Exemple d'image où la segmentation à l'aide de filtres de Gabor donne des résultats inférieurs. On peut voir clairement qu'une partie des régions de traits est omise dans la phase d'extraction

Image d'origine	Traits extraits avec notre approche	Traits extraits avec les filtres de Gabor
		

FIGURE 7.2.5: Exemple d'images de régions de traits obtenus à l'aide de l'approche proposée et de la segmentation basée sur des filtres de Gabor. À gauche on peut trouver l'image d'origine, au centre les traits extraits avec notre Cette étape approche, et enfin à droite les traits extraits avec les filtres de Gabor

Nous avons alors utilisé les deux critères définis page **??** pour évaluer le rappel et la précision des deux méthodes. Les résultats sont présentés dans le tableau 7.3.

	Rappel	Précision
Segmentation de traits proposée	100 %	71,3 %
Segmentation basée sur les filtres de Gabor	82,3 %	14,6 %

TABLE 7.4: Comparaison quantitative globale entre notre approche et celle basée sur les filtres de Gabor utilisant les critères de rappel et de précision définis page **??**

Nous pouvons constater que l'approche proposée obtient de meilleurs résultats que les filtres de Gabor. Premièrement, on peut voir que le rappel est supérieur de 14% à celui obtenu avec les filtres de Gabor, ceci permet donc de conclure que notre méthode extrait des traits dans un plus grand nombre de zones des images. Cela peut également être vu sous la forme : 1 zone sur 6 qui est couverte par notre approche ne l'est pas par celle basée sur les filtres de Gabor. Concernant la précision, on peut remarquer que notre méthode obtient également un meilleur score que celle basée sur les filtres de Gabor. Les traits qui sont extraits sont donc mieux ciblés et couvrent une plus grande surface des zones de traits. En conclusion, on peut dire que les régions extraites par notre méthode sont plus proches de la vérité terrain.

7.2.2 Évaluation des sacs de traits

Afin d'évaluer notre description statistique des traits dans les lettrines, nous avons menés une série d'expérimentations sur la base de 358 lettrines utilisée dans la couche forme (voir section 6.2.3 pour plus de détails). Ainsi, nous pourrons comparer les résultats obtenus sur la couche forme et la couche texture.

Pour chaque image de la base, nous venons extraire l'ensemble des traits qui la composent, nous calculons les caractéristiques associées à chaque trait, puis nous venons construire le dictionnaire de mots (nos sacs de traits). À partir de cette dernière étape, un histogramme des fréquences d'apparition de chaque type de trait est calculé, et nous réduisons le problème de comparaison de lettrines à une distance entre histogrammes présenté dans la section 152. Il nous est donc possible de connaître la distance entre une lettrine et toutes les autres de la base. Nous appliquons ensuite un algorithme des K-Plus Proches Voisins [Cover 67] pour retrouver les images les plus similaires à l'image requête. Les résultats obtenus sont présentés dans le tableau 7.5, et présentés pour trois valeurs de K différentes. Tout comme nous l'avons fait pour la couche forme, pour $K = 3$ et $K = 5$, avons appliqué un processus de vote à la majorité, et sélectionné la classe à attribuer à l'image requête.

Les résultats montrent que l'approche proposée permet d'obtenir des résultats très encourageant, puisque qu'en moyenne 4 images sur 5 seront bien annotées. Par contre, contrairement aux approches basées sur les formes, le fait d'augmenter le nombre de réponses renvoyées par notre système ne permet pas d'améliorer les résultats. Cela s'ex-

K	Taux reconnaissance
1	78,8 %
3	78,2 %
5	78,5 %

TABLE 7.5: Taux de reconnaissances obtenus avec l'approche statistique proposée pour décrire des images de documents anciens à partir des traits qui les composent. Pour chaque valeur de K utilisé dans l'algorithme des K plus proches voisins, nous présentons les taux de reconnaissances obtenus. Dans le cas ou K est supérieur à 1, la classe associée à chaque image correspond à la classe majoritaire parmi l'ensemble des résultats

plique par le fait qu'une partie de la base (les images des styles 2 et *autres styles*) ne comportent pas de traits. Ces images sont donc difficile à caractériser en utilisant une approche par sac de traits. Cela se répercute sur les taux de reconnaissances qui ne sont pas exceptionnels, et qui stagnent même si l'on va chercher plus de réponses de la part du système.

Cette première approche statistique apporte des résultats intéressants. Nous allons maintenant présenter les résultats obtenus avec l'approche structurelle, pour pouvoir les comparer sur le cas de la couche texture.

7.2.3 Évaluation de la description par graphes topologiques

Pour cette dernière évaluation, nous avons réutilisé le protocole défini pour l'évaluation des graphes topologiques issus de la couche forme (voir section 6.2.3). Nous avons donc réutilisé la signature de noeud présentée dans l'équation 6.2.1, et couplé cette signature à la distance entre graphes définie par [Jouili 11].

Tout comme nous l'avons fait pour la couche forme, nous avons limité les attributs des noeuds à la position du centre de gravité de chaque région de traits, et l'aire de la région. Cette limitation a pour but de mettre en avant la pertinence des régions extraites et les qualités topologiques des graphes générés. Enfin, les arcs des graphes sont pondérés par la distance euclidienne qui sépare les centres de gravité des formes.

Nous avons utilisé la même base de 358 lettrines que celle utilisée dans la partie validation de la couche forme (voir section 6.2.2.1 pour les détails), afin de pouvoir comparer les résultats entre les approches basées sur la couche forme, et celle basée sur la couche texture. Nous avons donc extrait les régions de traits pour l'ensemble de la base, décrit ces régions, puis nous avons généré un graphe par lettrine. Pour chaque lettrine, nous calculons la distance qui la sépare des autres. Enfin, nous appliquons l'algorithme des K-Plus Proches Voisins pour vérifier que notre description permet bien de faire la distinction entre les styles de lettrines. Tout comme pour l'évaluation précédente, nous appliquons un mécanisme de vote à la majorité pour les valeurs de K supérieurs à 1. Les

résultats obtenus sont présentés dans le tableau 7.6.

K	Taux de reconnaissance
1	60,6 %
3	62,8 %
5	62,8 %

TABLE 7.6: Taux de reconnaissance obtenus sur la base de lettrines utilisées dans [Pareti 06a] en utilisant une signature structurelle sur les zones de traits, et une distance hybride entre graphes

Contrairement aux approches reposant sur les graphes de formes, on peut constater que les résultats obtenus à partir des graphes de traits ne donnent pas de très bons résultats. Tout comme nous l'avons constaté avec l'approche par sac de traits, cela s'explique facilement par le fait que les styles retenus permettent de mettre en avant certaines caractéristiques recherchées par les historiens. Ainsi, si l'on observe les images du premier style, on peut remarquer que celles-ci correspondent principalement à des images avec un fond hachuré (composé de traits). Dans le cas des autres styles, ces images sont composées d'un fond noir ou blanc, et d'un motif figuratif (style 3), ou décoratif (autres styles).

Ces images ne contiennent donc que peu de traits, et les graphes extraits sont soient vides (cas des images du style numéro puisqu'elles ne contiennent aucun trait), soient ils ne contiennent que peu de noeuds. Ces résultats sont donc à pondérer en fonction de ces paramètres, et cette expérimentations nous permet de vérifier que les résultats sont cohérents avec l'approche proposée (la décomposition permet de décrire certaines caractéristiques des images et d'ignorer les autres). On peut donc déduire qu'une approche de description des images de traits permet de bien discriminer les images contenant des traits des autres, et qu'il paraît intéressant de combiner ces deux descriptions pour voir si les résultats confirment notre modèle.

7.2.4 Complexification des descriptions statistiques et structurelles

Avant de combiner les signatures issues des formes et celles issues de la couche texture, nous avons combiné les signatures statistiques et structurelles de la couche texture. De manière analogue à ce que nous avons réalisé sur la couche forme, nous cherchons à complexifier la description pour combiner une information liée aux statistiques et une liée à la topologie des régions de traits. Les résultats obtenus sont présentés dans le tableau 7.7.

On observe à nouveau que la combinaison d'informations spécifiques et avec des buts différents permet d'accroître les taux de reconnaissances. Là encore, le gain obtenu par la combinaison est de l'ordre de 40% entre une description topologique seule et une des-

K	Taux de reconnaissance
1	85,2 %
3	86,3 %
5	87,4 %

TABLE 7.7: Taux de reconnaissance des lettrines à partir de la couche texture en utilisant une description re-complexifiée des lettrines

cription qui combine les informations statistiques et structurelles.

Ceci nous permet d'observer le problème de la représentation unique qui est insuffisante pour décrire des images complexes de documents anciens. Il parait donc évident de combiner ces différentes descriptions pour mieux caractériser nos images, et vérifier que l'utilisation conjointe d'information provenant de deux couches d'informations nous permet d'obtenir une description plus poussée de nos images.

7.3 Complexification de la description - mise en relation des informations sur les formes et des informations sur les traits

L'approche proposée depuis le début de ce manuscrit propose de simplifier des images complexes de documents anciens pour les étudier sous deux angles de vus différents. Ainsi, nous venons de présenter comme nous séparons les formes des textures qui composent les images. Pour chacune de ces couches, nous venons extraire une information statistique (à l'aide d'une approche par sac de mots visuels) et une information topologique (à l'aide de graphes).

Nous constatons que bien que les résultats obtenus sur chaque couche donnent des résultats corrects, ils ne sont cependant pas adaptés à toutes les images. En effet, certaines sont principalement composées de formes (comme par exemple une lettrine sur fond blanc composée de fleurs) pendant que d'autres sont principalement composées de traits (comme par exemple une lettrine avec un fond hachuré). Ceci perturbe les résultats , et ne paraît pas cohérent [2]. Ainsi, nous proposons de combiner ces résultats pour re-complexifier la description des images.

Ainsi, nous commençons par présenter une première fusion des résultats obtenus avec une signature statistique issue de la couche forme et de la couche texture. Nous rappelons que les signatures statistiques correspondent à des approches par sacs de mots visuels définis pour chaque couche. Concernant la couche forme, les mots correspondent aux motifs les plus fréquents au sens du critère $TF-IDF$ (voir section 6.1.2.1 pour plus de détails). Concernant la couche texture, les mots correspondent à la fréquence d'apparition de classes de traits caractéristiques de l'ensemble de la base (voir section 7.1.3 pour plus de détails). Cette fusion peut être interprétée comme la recherche de motifs visuels issus de la forme et des traits. Nous présentons les résultats dans le tableau 7.8.

K	Couche formes	Couche textures	Total
1	63,1 %	78,8 %	88,3 %
3	67 %	78,2 %	83,2 %
5	70,1 %	78,5 %	82,4 %

TABLE 7.8: Taux de reconnaissance en utilisant une description recomplexifiée des lettrines, qui mélange une description statistique issue de la couche forme, et une de la couche texture

Nous pouvons constater que contrairement à nos attentes, plus nous augmentons le nombre de voisin dans la recherche des images similaires, plus le taux de reconnaissance

2. comme par exemple le mauvais résultat obtenu à partir des graphes de traits sur les images du style 2 dans la section 7.2.3

chute, tandis que les taux des deux couches respectives augmentent ou restent stables. Ceci s'explique par la combinaison de deux raisons principales. Premièrement, les mots visuels issus de la couche forme sont caractéristiques d'une image s'ils correspondent à des motifs relativement rares (s'il sont présent dans toute la base, leur IDF sera faible et ils ne seront pas conservés). Dans le cas d'images composées de grandes zones lisses (comme le style 2), celles-ci sont difficilement identifiables puisque composées du même motif en grand nombre.

La seconde raison provient de la stabilité des taux de reconnaissances provenant de la couche texture. Comme nous l'expliquions précédemment, près d'un quart de la base est composé d'images avec très peu ou pas de traits. Il devient donc difficile d'en extraire des sacs de traits pertinents. Hors, les images qui sont difficilement identifiables par les sacs de traits, le sont également par des sacs de motifs. Par combinaison de ces deux phénomènes, on peut constater que les taux de reconnaissances augmentent mais ne permettent pas de renvoyer une information sure pour toute la base.

La seconde fusion que nous présentons vient mettre ensemble les descriptions structurelles des deux couches pour voir si la combinaison des graphes issus des deux couches présentent le même soucis que pour la fusion des signatures statistiques. Nous rappelons que les graphes sont construit à partir des régions qui sont extraites dans la couche forme (voir section 6.1.3) et la couche texture (voir section 7.1.4). Pour l'ensemble des graphes, un noeud correspond à une région extraite, et les arcs sont pondérés par la distance entre les régions. Les résultats sont présentés dans le tableau 7.9.

K	Couche formes	Couche textures	Total
1	73,7 %	60,6 %	87,7 %
3	92,5 %	62,8 %	95,5 %
5	95,8 %	62,8 %	96,9 %

TABLE 7.9: Taux de reconnaissance en utilisant une description recomplexifiée des lettrines, qui mélange une description structurelle issue de la couche forme, et une de la couche texture

Contrairement aux approches statistiques, nous pouvons constater que combiner les signatures structurelles des deux couches fonctionne bien et permet d'accroître le résultat finale de manière remarquable (au rang 1, le taux de reconnaissance final augmente de près de 20 %). L'explication de cette augmentation repose sur le fait que les zones de formes permettent de décrire de manière plus précise les grandes zones lisses des images de documents anciens, pendant que le graphe des zones de traits permet de décrire de manière pertinente les images composées de traits.

Maintenant que nous avons observé la fusion des signatures de même nature provenant de deux couches d'informations différentes, nous proposons de fusionner les résultats

provenants de signatures différentes et de couches différentes. Nous commençons par fusionner la signature statistique provenant de la couche forme, et la signature structurelle provenant de la couche textures. Les résultats sont présentés dans le tableau 7.10.

K	Couche formes	Couche textures	Total
1	63,1 %	78,8 %	81,3 %
3	67 %	78,2 %	83,5 %
5	70,1 %	78,5 %	84,4 %

TABLE 7.10: Taux de reconnaissance en utilisant une description recomplexifiée des lettrines, qui mélange une signature structurelle provenant de la couche texture, et une signature statistique provenant de la couche forme

Ces deux signatures sont celles qui donnent les moins bons taux de reconnaissances parmi celles présentées, et elles sont toutes les deux efficaces sur un sous-ensemble d'images de la base. Cependant, on peut voir que leur fusion donne des résultats très corrects, et surtout que les résultats obtenus au rang 5 sont meilleurs que ceux obtenus par la combinaison des deux signatures statistiques. Cela nous laisse donc présager que nous pouvons améliorer les taux en combinant de manière plus large nos signatures.

La dernière combinaison partielle de nos signature entre couche s'attarde sur la fusion de l'approche topologique de la couche forme, et de l'approche par sac de traits. Ces signatures permettent individuellement d'obtenir les meilleurs résultats sur leur couche respective. Nous souhaitons pouvoir constater, à travers cette expérimentation, que cette supériorité perdure et que les deux approches se complètent. Les résultats sont présentés dans le tableau .

K	Couche formes	Couche textures	Total
1	73,7 %	78,8 %	93,6 %
3	92,5 %	78,2 %	96,9 %
5	95,8 %	78,5 %	98 %

TABLE 7.11: Taux de reconnaissance obtenus en utilisant une description recomplexifiée des lettrines qui mélange les résultats obtenus avec les graphes issus de la couche forme, et ceux de l'approche par sacs de traits

Cette fusion est la plus intéressante puisque nous pouvons observer que dès le rang 1, le taux de reconnaissance est augmenté de plus de 20%. Cela met clairement en avant l'efficacité et l'importance de combiner des informations issues de sources différentes, et de nature différente.

Enfin, la dernière fusion que nous présentons porte sur la combinaison des différentes signatures qui ont été définies dans ce manuscrit. Elle permet donc de mettre en lien les

signatures statistiques et structurelles issues de la couche forme et de la couche texture.

Couche	Forme		Texture		
K	Statistique	Structurelle	Statistique	Structurelle	Total
1	63,1 %	73,7 %	78,8 %	60,6 %	96,9 %
3	67 %	92,5 %	78,2 %	62,8 %	98 %
5	70,1 %	95,8 %	78,5 %	62,8 %	98,6 %

TABLE 7.12: Taux de reconnaissance utilisant une description recomplexifiée des lettrines suite à un vote à la majorité de toutes les caractéristiques (fusion globale)

A travers ce tableau, on constate clairement que simplifier l'information provenant d'image complexe ne permet pas de décrire pleinement une image. Cependant, en accord avec le paradigme d'Edgar Morin, on peut constater qu'en complexifiant les descriptions, il est possible de retrouver les images contenant le même type d'information complexe.

En conclusion, nous pouvons affirmer que fusionner des informations statistiques et structurelles provenant de couches différentes permet d'obtenir de meilleurs résultats. Cependant, combiner des signatures n'a d'intérêt que si l'on est sur de pouvoir compenser les faiblesses de l'une avec les forces d'une autre. Cette complémentarité est visible dans le cas de l'utilisation conjointe de la signature statistique de la couche forme et de la signature structurelle sur la couche texture.

Dans l'ensemble des résultats que nous venons de démontrer, nous pouvons observer qu'une analyse par mots-visuels pertinents n'a de sens que si l'information contenue dans les images est résumable par des mots qui ne sont pas présent dans toute la base. Dans le cas de la couche forme, on peut remarquer que les images composées de très peu de motifs différents (comme les images associées au style 2) sont difficilement caractérisables. A contrario, on peut constater que représenter les documents anciens par des mots-visuels obtenus sur la couche texture est pertinent (les traits sont caractéristiques d'un type d'image).

Synthèse du chapitre

Le troisième chapitre de cette partie traite de l'analyse de la couche texture de la décomposition de Meyer. Tout comme pour la partie forme, il se décompose en deux parties consacrées à la couche texture : la première présente comment est extraite l'information de cette couche, pendant que la seconde présente les différentes évaluation menées. Une dernière partie a été ajoutée pour complexifier les descriptions entre les deux couches d'informations, et observer l'évolution des résultats face à cette fusion d'information.

L'extraction d'information à partir de la couche texture reste conforme au modèle, puisque des motifs de traits sont utilisés pour représenter les traits des images. À partir de ces motifs, deux signatures sont définies : une première statistique qui repose sur l'utilisation de sacs de traits caractéristiques des images de la base, et une seconde qui se sert des régions extraites pour définir leur agencement topologique. La partie évaluation est également semblable à celle présentée sur la couche forme, puisqu'elle présente tout d'abord une comparaison entre les régions extraites et une vérité terrain. Puis, dans un second temps, l'approche statistique est évaluée la base de lettrines. Enfin, une dernière évaluation permet de mettre en avant les capacités de la signature structurelle. La dernière partie porte sur la complexification de la description des images à l'aide d'une fusion tardive des différents résultats obtenus via les signatures de la couche forme et de la couche texture.

Bien que la combinaison de l'information statistique et structurelle, à partir des formes et des textures des images de documents anciens, permette d'obtenir des résultats très satisfaisants, il reste difficile d'interpréter ces résultats. Les motifs et les régions extraites n'ont pas de sémantique particulière, et ne sont pas directement interprétables par un utilisateur final. Ce type d'outil limite donc l'utilisateur à des requêtes par l'exemple, sans qu'il puisse réellement comprendre le mécanisme sous-jacent, et expliquer les erreurs du système..

Dans le cadre de ce manuscrit, nous souhaitons aller plus loin dans la démarche, et faire un pas vers la réduction du fossé sémantique. L'idée est d'utiliser les résultats obtenus via nos traitements d'images pour extraire une connaissance de plus haut niveau sur les images. Cette connaissance est alors utilisée conjointement avec celle issue de l'annotation des historiens pour trouver une cohérence entre l'information extraite automatiquement, et les éléments recherchés par les historiens. C'est ce que nous présentons dans la dernière partie de ce manuscrit.

Chapitre 8

Ontologie finale et règles d'inférences

Nous rappelons qu'afin de proposer une description des images de documents anciens qui permette de prendre en compte les attentes des historiens, nous avons développé un modèle de description des images du point de vue des historiens et du point de vue informatique. Chacun des modèles permet de structurer la description des images, soit à partir de mots-clés issus du thesaurus, soit à partir des caractéristiques. Il est donc possible d'extraire de l'information de ces deux ontologies. Cependant, le problème de fossé sémantique reste présent, puisque chacun des modèles impose des requêtes semblables à l'information qu'il structure.

Pour résoudre ce problème de fossé sémantique, une solution consiste à lier les modèles pour créer une correspondance entre les différents types d'informations. Cette correspondance permet ainsi d'extraire de l'information à partir de l'ontologie de la connaissance des historiens, et de celle des traitements d'images et relations spatiales de manière conjointe. Cette correspondance aboutit à une nouvelle ontologie, que nous dénommons ontologie finale, et qui rassemble dans une même base de connaissance les méta-données issues du thesaurus et des traitements des images.

L'extraction d'information à partir de cette base de connaissance a lieu à l'aide de requêtes. Une requête formulée selon le formalisme d'un des modèles pourra proposer des réponses issues des deux ontologies présentées précédemment via des règles d'inférences. Ces règles assurent donc la correspondance entre les deux domaines d'expertise, et une connaissance de haut-niveau sémantique est alors déduite à partir des caractéristiques bas-niveau des images. Nous présentons en détails ces étapes dans ce chapitre.

8.1 Ontologie finale

Les deux ontologies - Ontologie de la connaissance des historiens ; Ontologie des traitements d'images et des relatons spatiales - ont été créées à partir des concepts définis précédemment (T-Box) et peuplées à l'aide de la vérité terrain saisie par les historiens (A-Box), d'une part, et des formes et caractéristiques extraites automatiquement, d'autre

part. Le concept *Image* est commun aux deux ontologies. Nous avons donc lié ces deux concepts pour combiner la connaissance issue des deux ontologies. Cette combinaison a été réalisée en utilisant le constructeur OWL : equivalentClass (OWL first version [McGuinness 04, OWL 11]) pour lier les concepts au niveau des T-Box, et en utilisant les constructeurs OWL : sameAs pour lier les A-Box. L'intérêt principal de lier les ontologies au niveau du concept image (concept de plus haut niveau dans la hiérarchie des ontologies) est qu'elle offre une généricité à notre modèle.

8.2 Les outils de modélisation

Afin de valider notre modèle, nous l'avons instancié en utilisant via langage de description. Les langages de description possèdent tous des syntaxes particulières. Afin de faciliter leur utilisation et de simplifier la manipulation des ontologies, des outils de modélisation ont été proposés. Nous présentons les plus utilisés.

8.2.1 Protégé

Protégé [Gennari 03] est un éditeur d'ontologies distribué en open-source par l'université d'informatique médicale de Stanford. C'est un éditeur hautement extensible, capable de manipuler des formats très divers. Cette librairie Java permet de créer de véritables applications à bases de connaissances en utilisant un moteur d'inférence pour raisonner (les moteurs d'inférence JESS [Friedman-Hill 03], ou RACER [Haarslev 01] par exemple) et ainsi déduire de nouveaux faits.

8.2.2 SWOOP

SWOOP [Kalyanpur 05] est un éditeur d'ontologie développé par l'Université du Maryland dans le cadre du projet MINDSWAP. Contrairement à Protégé, il a été développé de façon native sur les standards RDF et OWL, qu'il prend en charge dans leurs différentes syntaxes (pas seulement XML). C'est une application plus légère que Protégé, mais moins évoluée en termes d'interface, qui intègre également des outils de raisonnement.

8.2.3 Jena

Jena [JEN 2005] est un framework open-source écrit en Java dans le cadre du projet *HP Labs Semantique Web Program*, dont l'objectif est de fournir un environnement facilitant le développement d'applications dédiées au web sémantique. Jena permet de manipuler des documents RDF, RDFS et OWL, et fournit en plus un moteur d'inférences permettant des raisonnements sur les ontologies. La contribution principale de Jena réside dans son API pour la manipulation des graphes RDF mais la plate-forme inclue

également des outils divers comme, par exemple, un analyseur RDF/XML, un langage d'interrogation, des modules de lecture/écriture pour N3, etc.

8.2.4 Oracle

Oracle est une plate-forme ouverte, évolutive et sécurisée qui permet de stocker des données RDF et OWL. Elle intègre également des processus d'inférence et est compatible avec la majorité des langages de description d'ontologie (SPARQL, OWL, OWL 2, Pellet OWL DL reasoneur, Jena v2.5, ...).

8.2.5 Jess

Développé au Sandia National Laboratories par le Dr. Ernest J. Friedman-Hill à la fin des années 1990, Jess (Java Expert System Shell) [Friedman-Hill 03] est un moteur d'inférences et un environnement de script écrit en langage Java. Ce moteur permet de faire de l'inférence sur les connaissances fournies à l'aide de règles déclaratives.

8.3 Peuplement des ontologies

Les modèles de connaissance que nous avons créés sous forme ontologie nécessitent d'être peuplés, c'est à dire qu'il est nécessaire d'y insérer l'information liée à nos lettrines si nous souhaitons pouvoir en extraire de la connaissance.

La première ontologie,

l'ontologie qui regroupe les informations issues des historiens, est peuplée avec la vérité terrain saisie par les historiens. Pour chaque image de lettrine, les historiens ont eu à indiquer une série de sept informations (la figure 8.3.1 présente l'interface utilisée par les historiens pour saisir ces informations) :

- Le cadre : entoure la lettrine, et est composé d'aucun, un ou deux filets ;
- Le fond : correspond à l'arrière plan et peut-être plein (noir ou blanc), hachuré ou criblé ;
- Le motif : correspond aux ornements. Il peut être décoratif ou figuratif ;
- L'alphabet : utilisé pour imprimer la lettre ;
- La couleur de la lettre : elle peut être noire ou blanche ;
- La lettre : identifie la lettre à partir de l'alphabet utilisé :
- La Police : représentation de la lettre utilisée au moment de l'impression.

Les informations saisies pour 4288 lettrines ont été utilisées pour peupler cette première ontologie

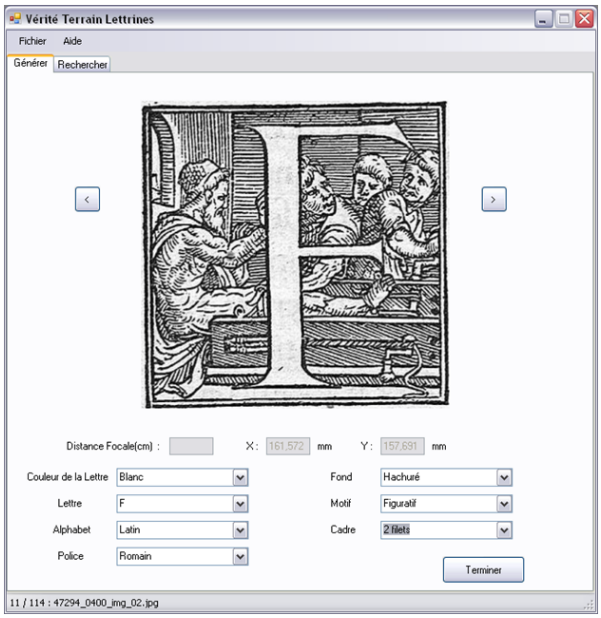


FIGURE 8.3.1: Interface utilisée par les historiens pour saisir leur connaissance sur les images de lettrines

La seconde ontologie,

celle qui rassemble les informations extraites par les algorithmes de traitements des images présentés dans la partie 5. Nous tenons à préciser que des régions de formes et de traits ont été extraites dans la partie caractérisation des images. Cependant, afin de tester notre approche, nous avons pour le moment limité nos tests aux régions issues de la couche forme. Cette limitation peut se justifier par le fait que les taux de reconnaissances obtenus sur cette couche sont meilleurs, et que l'information principalement recherchée par les historiens (la lettre des lettrines) se trouve dans la couche forme. De plus, même si notre approche est uniquement présentée sur la couche forme, des travaux sont en cours actuellement pour la tester avec des régions issues de la couche textures. Cette approche reste donc générique et adaptable à tout type d'image qui peut être décomposée en régions.

Nous avons donc extrait les régions de la couche forme à partir des algorithmes présentés dans la section 6.1.3.1 sur l'ensemble des images. Nous obtenons les résultats présentés dans la figure 8.3.2.

Nous rappelons qu'à partir des résultats obtenus sur la couche forme, nous avons constaté que décrire les images à partir d'une signature purement statistique ou purement structurelle ne permettait pas de profiter des meilleurs taux de reconnaissances. De plus, une signature statistique par sacs de mots visuels permet de décrire une image sans prendre en compte la position de l'information dans l'image. A contrario, une signature structurelle limite la connaissance extraite aux relations entre les zones qui la composent,



(a) (b)

FIGURE 8.3.2: Exemples de régions extraites automatiquement en utilisant les traitements présentés dans la partie 5. A gauche, image d'origine, et à droite les régions obtenues (une couleur par région)

sans connaître l'aspect de chacune d'entre elles. Afin de palier à ces problèmes, et de cumuler les avantages des deux types de signatures, nous proposons de décrire les régions en utilisant les caractéristiques qui permettent de déduire de la connaissance sur leur aspect et leur position :

- Coordonnées du centre de gravité : indiquent la position de la forme dans l'image ;
- Aire : pour avoir le nombre de pixels qui composent la forme ;
- Excentricité : permet de renseigner sur l'allongement de la forme ;
- Niveau de gris moyen : pour connaître la teinte moyenne d'une forme ;
- Ecart-type du niveau de gris : pour savoir si la forme est relativement uniforme du point de vue de la teinte ;
- Nombre d'Euler : informe sur le nombre de trous dans une forme. Comme nous recherchons des lettres, celles-ci ne comportent pas plus de 2 trous ;
- 3 premiers moments de Hu : descripteurs usuels de formes. L'intérêt des trois premiers moments reposent sur leur rapidité de calcul, et sur leur description globale et certaine de la forme. Notre intérêt ici n'est pas d'obtenir une description précise de chaque forme mais plutôt sur l'ensemble des formes et leur agencement ;
- Caractéristiques de partitions : variables booléennes utilisées pour savoir si une forme s'entrecroise avec une des sous-parties de le la lettrine (comme définies dans la section 6.1.3.2). Nous tenons à rappeler que dans le cadre de ce manuscrit, nous avons préféré calculer les relations spatiales entre les régions et le masque composé des sous-parties au niveau traitement des images et non au niveau de l'ontologie par soucis de simplicité. Cependant, cette étape n'enlève rien du point de vue de la généricité de notre approche et du modèle proposé. Cette étape pourrait également être traitée au niveau de l'ontologie en prenant le contour des formes, et en calculant des relations spatiales issues de l'algèbre RCC-8 entre les points du contour et les différentes sous-parties.

À partir de la description des images telle que présentée ci-dessus, nous peuplons l'ontologie avec les 589 régions formes extraites de 100 lettrines. Nous nous sommes limités

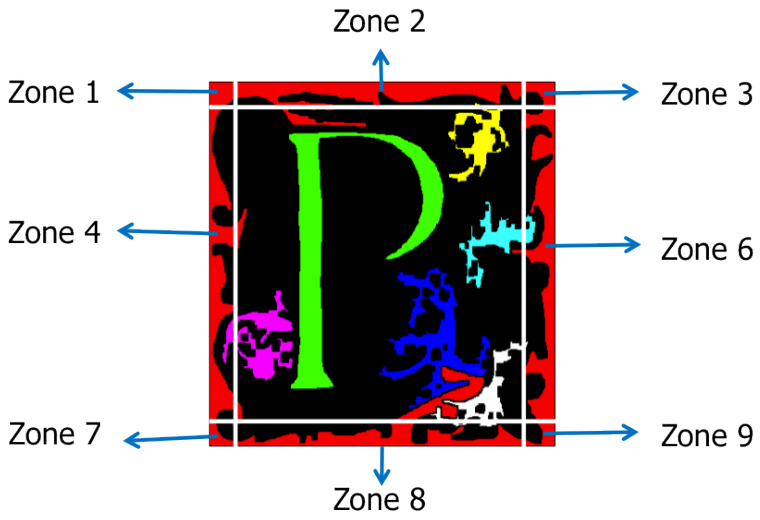


FIGURE 8.3.3: Exemple de découpage d'une image en neuf sous-parties pour le calcul des relations spatiales entre les régions et le masque défini. Chaque couleur correspond à une région extraite automatique par les algorithmes présentés dans la section 6.1

à 100 lettrines pour le moment pour faciliter la tâche de vérification de nos résultats [1].

Enfin, la mise en correspondance des deux ontologies a été réalisée et nous a permis d'obtenir une base de connaissance qui rassemble les informations issues des informations saisies par les historiens, et celle obtenue par les algorithmes de traitement des images. Cette connaissance multiple est effective pour 57 images figuratives et 43 images décoratives.

8.4 Règles d'inférences

Les règles d'inférences permettent de raisonner sur la base de connaissance pour en extraire de nouvelles connaissances (implicites). L'objectif de règles est donc d'enrichir la base de connaissance en utilisant les valeurs des différentes caractéristiques (inférence sémantique à partir de l'information dans les concepts), et les relations spatiales entre les formes (inférence spatiale à partir des concepts eux-mêmes).

Les règles sont appliquées sur les régions extraites pour leur associer automatiquement un sens. Cette connaissance se modélise dans l'ontologie par l'ajout de nouvelles propriétés.

1. La vérification de la connaissance associée à chaque région est pour l'instant manuelle. Des travaux sont en cours, au sein du laboratoire, pour automatiser ce traitement, et ainsi vérifier la pertinence de l'approche sur de plus grandes bases

8.4.1 Règles mises en places

Afin d'extraire de l'information de la base de connaissance, il est indispensable d'utiliser des règles d'inférences. Celles-ci permettent de déduire de la connaissance "cachée" dans la base, puisque non exprimée directement par les informations saisies. Dans le cas de l'ontologie finale, l'idée est de permettre à un utilisateur d'exprimer sa requête à partir des mots-clés du domaine des historiens, et d'utiliser des règles d'inférences pour faire le lien entre l'information recherchée et l'information extraite par les traitements des images.

Afin de valider notre modèle, nous proposons trois exemples de règles qui permettent d'extraire de la connaissance implicite, soit :

- de l'ontologie du domaine des historiens,
- de l'ontologie des traitements d'image,
- de l'ontologie finale.

Interroger une base de connaissance impose également l'utilisation d'un langage adapté. Plusieurs langages de requêtes ont été présentés dans la section 4.2.4.1, et dans le cadre de ce manuscrit, nous avons choisi d'utiliser le langage SPARQL.

8.4.2 Règles pour la validation du modèle du domaine historien

Les règles mises en places dans la validation de l'ontologie des connaissances des historiens ont été choisies avec deux buts principaux :

- Prouver que notre modèle permette d'extraire de la connaissance de cette base de connaissance ;
- Être comparable avec des résultats obtenus avec une approche classique de la littérature.

Ainsi, l'information saisie par les historiens a été insérée dans notre l'ontologie qui traite la connaissance des historiens, et nous l'avons également injectée dans une base de donnée relationnelle au format SQL. Afin de pouvoir comparer les deux approches, nous avons soumis, tant que faire se peut, les mêmes requêtes aux deux systèmes pour observer les différences. Les différents types de requêtes que nous avons soumis sont les suivants :

- Les requêtes complètes : une valeur a été saisie pour tous les champs ;
- Les requêtes incomplètes : aucune valeur n'est précisée pour un ou plusieurs champs ;
- Les requêtes incomplètes avec négation de champ : un critère de négation est utilisé pour décrire le contenu d'un champ.

Nous présentons en détails les requêtes que nous avons proposées dans chaque type ci-dessous.

Les requêtes complètes

Elles correspondent aux requêtes classiques que l'on peut voir dans les bases de données relationnelles. Une requête consiste donc à préciser la valeur attendue pour chaque

champ. Ces requêtes sont directement réalisables sur la base de donnée relationnelle et sur la base de connaissances. Un exemple de requête de lettrine est : *"Tous les V Blancs, Latin, Romain, Hachuré, Figuratif avec un filet"*.

Pour cette requête nous avons obtenu 2 réponses, avec les deux systèmes et les deux langages de requêtes. On peut donc constater que l'ontologie du domaine de connaissance des historiens permet de retrouver une information exhaustive aussi efficacement qu'un système de gestion de base de données.

Les requêtes incomplètes

Une requête partielle consiste à préciser une valeur pour certains champs seulement. Elle peut correspondre à une attente particulière d'un utilisateur qui n'aurait pas connaissance de certains détails, ou tout simplement que ne s'intéresserait qu'à une information spécifique. Voici trois exemples de requêtes incomplètes que nous avons soumis aux deux systèmes :

- *"Toutes les lettres noires de l'alphabet Latin, de Police Romain avec 2 filets"*
- "Tous les L hachurés"
- "Toutes les lettrines dont l'alphabet est grec"

On peut remarquer que dans ce type de requête, l'utilisateur ne s'intéresse pas à l'ensemble des caractéristiques mais plus spécifiquement à un sous-ensemble de celles-ci. Pour chacune des requêtes citées ci-dessus, nous avons obtenu le même nombre de réponses avec les deux systèmes, ce qui nous permet une fois de plus de valider notre modèle.

Requêtes incomplètes avec absence d'un champ

Le dernier type de requêtes, consiste à indiquer au système que l'on souhaite retrouver des images qui n'ont pas un critère particulier. Par exemple, la requête : *"Toutes les Lettrines qui n'ont pas de Cadre"* impose d'être capable d'identifier les lettrines pour lesquelles aucune information n'a été saisie pour leur cadre. Ce type de requête possède une sémantique forte qui n'est pas facilement transposable via les langages de requêtes sur des bases de données relationnelles. A contrario, la recherche de l'absence d'une information via le langage SPARQL est possible via l'utilisation de certains mots-clés. Cela permet donc de montrer que l'utilisation d'une ontologie et de requêtes nous permet d'obtenir des résultats identiques.

Cette première partie de l'évaluation nous a permis de valider l'ontologie du domaine des lettrines que nous avons mis en place, et de vérifier que la base de connaissances de lettrine peut s'adapter à tous les types de requêtes que les historiens sont amenés à proposer au système.

Afin de proposer une évaluation des deux autres ontologies (ontologie associée aux traitements d'images, et ontologie finale), nous avons créé des règles particulières. Celles-

ci permettent de vérifier qu'il est possible d'extraire de l'information des zones obtenues par les traitements d'images, et de les lier aux connaissances obtenues via l'ontologie des connaissances du domaine historien. Dans le cadre de ce manuscrit, nous nous sommes plus particulièrement intéressés à l'extraction de la lettre et des parties du corps dans les lettrines (éléments prégnants pour les utilisateurs). Nous avons donc créé deux propriétés simples des régions :

- *Region isLettre* : Pour indiquer qu'une région décrit une lettre ou non. Cette règle est utilisée pour valider le modèle des traitements d'images ;
- *Region isBody* : Pour indiquer qu'une région est identifiée comme une partie d'un corps dans les images de lettrines. Cette règle utilise la connaissance issue des deux ontologies et permet de valider le modèle global et l'ontologie finale, en proposant de réduire le fossé sémantique.

Nous présentons en détails ces règles dans ce qui suit.

8.4.3 Règles pour la validation de l'ontologie des traitements d'images

Après avoir évalué la pertinence de l'ontologie associée aux connaissances des historiens, nous avons cherché à évaluer la pertinence de celle qui permet d'intégrer la connaissance associée aux traitements des images. Nous tenons à rappeler que la chaîne de traitements des images a été présentée en détails dans la partie II. Celle-ci vise à simplifier une image en décomposant l'information qui la compose. Cette décomposition sépare donc les zones uniformes des zones texturées (zones de traits dans le cas des images de documents anciens). Une fois l'image simplifiée, nous obtenons deux couches d'informations (deux images) qui contiennent chacune soit la texture, soit les formes de l'image. À partir de ces couches, nous venons extraire des régions d'intérêts décrites par des caractéristiques propres au type d'information qu'elles renferment.

Tout comme nous venons de le voir pour l'ontologie associée aux connaissances des historiens, nous avons créé une règle qui permette de vérifier la pertinence et la cohérence du modèle (l'ontologie des traitements des images) que nous proposons. Cette règle se base donc uniquement sur les connaissances issues des traitements d'images pour proposer l'annotation automatique de formes extraites des lettrines.

La lettre contenue dans les lettrines est un élément important pour les historiens. L'intérêt de la lettre est qu'elle ne nécessite aucune connaissance particulière sur les lettrines pour être extraite. En effet, celle-ci peut-être vue comme une grande forme de l'image sans avoir besoin de connaissance de la part des historiens. Partant donc de l'ensemble des régions annotées obtenues par les traitements des images obtenus dans le chapitre 6.1, nous cherchons à identifier la forme qui correspond à la lettre. L'identification de la forme repose sur deux types d'information :

1. La connaissance issue des caractéristiques bas-niveaux calculées sur les formes ;
2. La localisation de la forme dans l'image (à l'aide de relation spatiales).

Nous présentons en détails ces deux types d'information ci-après.

8.4.3.1 Règle pour la propriété *isLetter : détection de la lettre*

La règle associée à la propriété *isLetter* a pour but d'annoter les régions correspondantes aux lettres dans les lettrines. Nous avons remarqué qu'une région peut être considérée comme une lettre si :

- Elle se situe au centre : la lettre est l'élément principal, il se situe donc au milieu ;
- Elle comporte peu de trous : une lettre n'est composée que de deux trous au plus (cas du B par exemple) ;
- C'est la plus grande forme vérifiant les deux premiers critères : étant l'élément principal, il occupe une grande place, et l'on suppose que les deux premiers critères ont éliminés les régions parasites.

La transcription de cette règle en langage ontologique impose d'être capable de faire le lien entre les caractéristiques statistiques extraites sur les formes, et d'identifier la position de la forme dans l'image. Les caractéristiques extraites sur les formes ont été présentées en détail dans la section 6.1.3.2, et nous reviendrons dessus dans la suite de cette partie. En ce qui concerne la localisation des formes dans l'image, nous avons choisi d'utiliser des relations spatiales entre les formes extraites et un quadrillage standard des images. Nous présentons ce choix ci-dessous.

Détection de la position de chaque région

Nous avons obtenu, via nos traitements d'images, un ensemble de régions et leur position dans l'image. Afin d'être capable de localiser les régions dans l'image, nous avons mis en place un quadrillage de l'image qui permette d'indiquer si une région se trouve (ou non) dans une zone du quadrillage (la figure 8.4.1 présentent un exemple de découpage possible, où les sous-parties sont numérotées de 1 à 9). Une liste de booléens indique si une forme s'entrecroise avec une sous-partie de l'image. Les booléens associés à cette partition sont utilisés pour localiser les formes dans l'image. Plus précisément, cette partition a été utilisée pour indiquer si une forme est située au centre de l'image, ou si elle est en contact avec un bord de l'image. Par exemple, la lettre d'une lettrine est située au centre, elle correspondra donc à une forme qui n'aura pas de recouvrement avec une zone proche d'un bord, mais aura un recouvrement avec la zone numéro 5 (dans le cas de l'exemple présenté dans la figure 8.4.1).

Nous avons choisi dans le cadre de ce manuscrit d'indiquer si une région croise une zone du quadrillage à l'aide d'un booléen pour une raison de simplification d'implémentation. Cependant, nous aurions pu également utiliser une approche similaire à celles développées dans les SIG, et rechercher la superposition entre une région et le quadrillage en utilisant un mécanisme d'inférence spatiale associé aux relations énoncées dans l'algèbre RCC-8 (voir page 87).

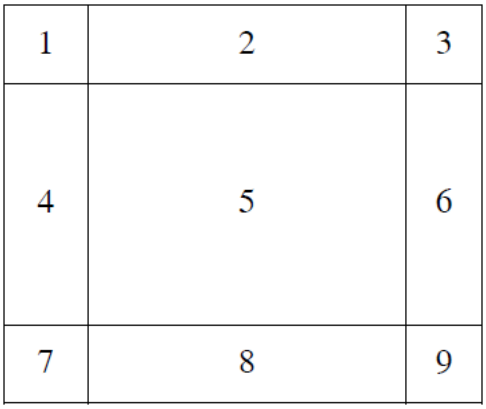


FIGURE 8.4.1: Sous-parties de l'image utilisées pour localiser les formes dans l'image (régions au centre, ou plutôt sur les bords)

Règle d'extraction de la lettre

Partant de la description statistique et de localisation des formes dans l'image, nous avons établi qu'une région forme peut-être annotée comme lettre si :

- Elle s'entrecroise avec la sous-partie numéro 5, et qu'elle ne touche pas les autres (connaissance déduite des relations spatiales entre les formes et les sous-parties)
- La valeur de son nombre d'Euler est comprise entre -2 et 2 (le signe du nombre d'Euler varie en fonction de la couleur de la lettre et des trous)
- Parmi l'ensemble des régions qui vérifient les deux premiers critères, celle avec la plus grande aire est finalement annotée avec la propriété isLetter (la forme qui valide tous ces critères est annotée comme étant une lettre)

8.4.3.2 Validation de la règle d'extraction de la lettre dans l'ontologie

L'ontologie des traitements d'images est créée en utilisant Protégé 3.4.4 (T-Box) et peuplée à l'aide de Jena [2] (A-Box). En utilisant le moteur d'inférence Jess, la règle est appliquée sur les régions extraites et annotées automatiquement.

Afin de vérifier la pertinence de l'ontologie, nous avons appliqué la règle "isLetter" sur un sous-ensemble de 100 images de lettrines. Ces images sont composées de 584 formes, et 112 sont identifiées comme étant au centre. Sur ces 112 régions, 100 ont été annotées comme étant une lettre (une seule forme par image peut-être annotée comme correspondant à la lettre). Une vérification manuelle nous a permis de constater que 91 régions étaient correctement identifiées, ce qui correspond à un taux de bonne extraction de 91%.

2. Jena - Librairie java pour le Web Semantic - http ://jena.sourceforge.net/

8.4.3.3 Validation de la méthode d'extraction de la lettre

Nous venons de constater qu'à partir de la séquence de traitements d'images qui repose sur la décomposition de Meyer et une analyse par une Loi de Zipf, et de la règle d'inférence définie ci-dessus, il est possible d'annoter une région de l'image en tant que lettre.

L'évaluation de cette série de traitements n'est pas triviale, mais est cependant indispensable, pour garantir son utilisabilité par les utilisateurs, et apporter un point de vue objectif sur le système. Dans le contexte du projet NaviDoMass, et en partenariat avec des chercheurs issus des sciences humaines, nous avons cherché à évaluer la pertinence de la forme extraite. L'évaluation du système porte donc sur l'application d'un système de reconnaissance de caractères (OCR) sur la forme extraite pour valider la segmentation de celle-ci. En conséquence, le taux de reconnaissance obtenu est donc le critère d'évaluation que nous retiendrons pour cette évaluation.

Nous avons effectué différentes séries de tests afin de valider nos choix. La première série de tests consistent à mettre en évidence les difficultés d'extraction de la lettre en comparant nos résultats à un processus de binarisation usuel reposant sur le critère de Otsu [Otsu 79]. Les deux approches sont suivies d'une extraction et de sélection de composantes connexes. Puis nous comparons les résultats obtenus à l'aide de deux OCR sur un ensemble de lettrines. Pour l'évaluation, nous avons utilisé un système de reconnaissance de caractères commercial (FineReader), et un système open-source (Tesseract). Nous avons appliqué cette démarche sur une base de données contenant 4288 images, et le dictionnaire usuel français a été utilisé dans les logiciels de reconnaissance de caractères. Les résultats obtenus sont résumés dans le tableau 8.1.

Extraction de la lettre à l'aide du critère d'Otsu

	FineReader	Tesseract
Taux de reconnaissance de la lettre	20,76 %	22 %

Extraction de la lettre à partir de l'image d'origine et de la Loi de Zipf

	FineReader	Tesseract
Taux de reconnaissance de la lettre	48,2%	39,6%

TABLE 8.1: Taux de reconnaissance des lettres extraites dans les lettrines en utilisant deux approches d'extraction de la lettre différentes et deux OCR

On peut observer que les taux de reconnaissances obtenus avec l'algorithme reposant sur la loi de Zipf sont deux fois supérieurs à ceux obtenus en utilisant une binarisation de l'image à l'aide du critère défini par Otsu. Les images présentées dans la figure 8.4.2 et la figure 8.4.3 présentent des exemples de lettres extraites en utilisant les deux approches. On peut remarquer dans la figure 8.4.2 que l'extraction des composantes connexes nécessite une connaissance à priori sur la couleur des formes à extraire. Dans le cas de la binarisation à partir du critère de Otsu, les lettres noires ne peuvent pas être

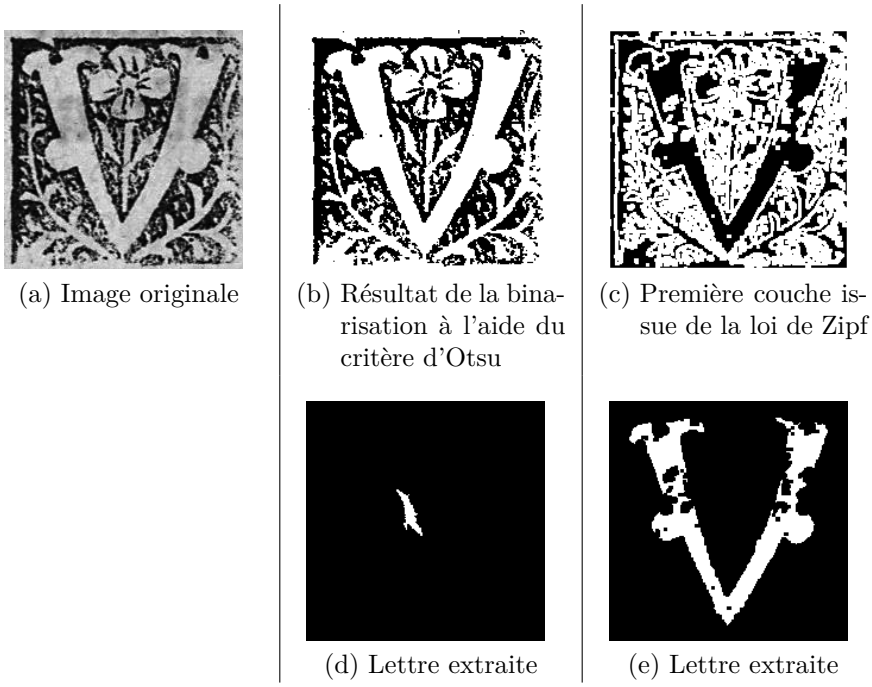
extraites. A contrario, comme la lettre peut-être vue comme une grande zone uniforme dans l'image, elle est extraite par la méthode reposant sur la loi de Zipf, indépendamment de sa couleur.



(a) Image originale (b) Résultat en utilisant le critère d'Otsu (c) Résultat en utilisant la loi de Zipf

FIGURE 8.4.2: Comparaison entre l'extraction de la lettre à l'aide de la binarisation à partir du critère de Otsu [Otsu 79], et de la segmentation obtenue à partir de la loi de Zipf. On peut remarquer que la principale différence repose sur le fait que la méthode développée par Otsu suppose que les formes sont blanches sur fond noir

La deuxième différence majeur, l'extraction de la plus grande partir d'une composante connexe, est présentée dans la figure 8.4.3. Dans cette figure, la même image a été segmentée en utilisant les deux critères utilisés dans la comparaison, et la lettre extraite est présentée. On peut voir dans l'image originale que le fond est dégradé et que la lettre est reliée à celui-ci (des points d'encre noire qui délimitaient la lettre ont disparus avec le temps). Le critère d'Otsu laisse donc la lettre connectée au fond puisqu'il consiste à binariser l'image à partir de l'histogramme des niveaux de gris de l'image. Dans le cas de la binarisation reposant sur la loi de Zipf, la segmentation repose sur les niveaux de gris présents dans l'image mais également sur la fréquence d'apparition de motifs de taille 3x3. Cette méthode vient donc privilégier les grandes zones uniformes (comme l'intérieur de la lettre) mais ignore les petites ramifications qui peuvent exister entre la lettre et le fond.



(a) Image originale

(b) Résultat de la binarisation à l'aide du critère d'Otsu

(c) Première couche issue de la loi de Zipf

(d) Lettre extraite

(e) Lettre extraite

FIGURE 8.4.3: Comparaison entre les résultats de binarisation obtenus avec le critère d'Otsu et la segmentation basée sur la Loi de Zipf. La binarisation obtenue avec le critère définit par Otsu classe chaque pixel comme appartenant au fond ou à la forme à partir de sa valeur. La segmentation obtenue par une loi de Zipf, quant à elle, repose sur l'analyse de la fréquence des motifs qui composent l'image. Les grandes zones comme la lettre peuvent ainsi être extraites même si elle est connectée au fond (les liens correspondent à des motifs moins fréquents)

À partir de cette première série d'expérimentation, nous pouvons déduire que l'approche basée sur la loi de Zipf donne de meilleurs résultats pour l'extraction de composantes connexes. Maintenant que nous avons pu évaluer de manière qualitative et comparative l'approche proposée pour extraire les formes dans les lettrines, nous proposons une seconde série d'expérimentations qui doit permettre de valider le choix de la décomposition de Meyer. Cette série d'expérimentations consiste à évaluer la qualité de la simplification des images, et à la comparer à une approche classique de la littérature, les ondelettes, pour mettre en avant son adéquation avec l'extraction de composantes connexes dans les documents anciens. Le tableau 8.2 présente les résultats obtenus avec notre base de lettrine et les OCR utilisés dans la première série d'expérimentations.

Comme nous pouvions l'espérer, les taux de reconnaissance obtenus à partir de la décomposition de Meyer et de la segmentation à partir de la loi de Zipf donne de meilleurs résultats que ceux obtenus en utilisant les ondelettes de Haar. La figure 8.4.4 présente la même lettre extraite en utilisant les deux approches évaluées. A chaque niveau de dé-

Résultats issus de l'ondelette de Haar et de la loi de Zipf		
	FineReader	Tesseract
Taux de reconnaissance	29,1%	27,8%

Résultats issus de la couche forme de Meyer et de la loi de Zipf		
	FineReader	Tesseract
Taux de reconnaissance	72,8%	67,9%

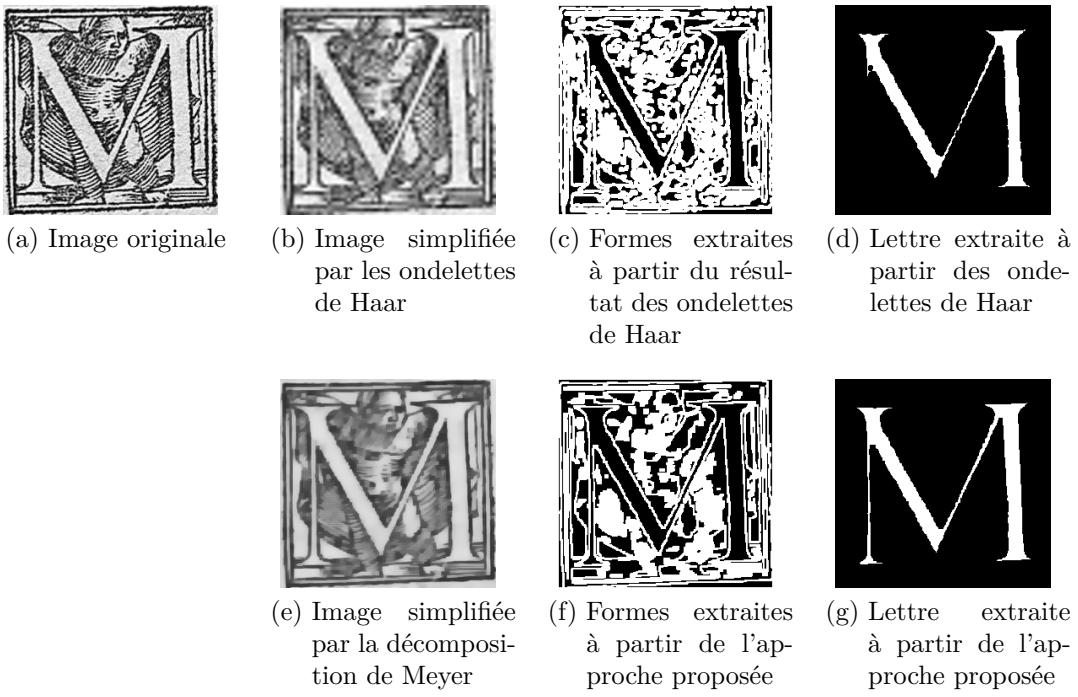
TABLE 8.2: Taux de reconnaissance de la lettre dans les lettrines en utilisant deux approches de simplification d'images différentes et un OCR

composition de l'ondelette, la fonction de lissage appliquée réduit les contours et ouvre les régions, dans le sens que celles-ci se retrouvent connectées entre elles. Dans l'exemple présenté sur la première ligne, la lettre M se retrouve séparée en deux parties et ne peut-être correctement extraite. Dans le cas de la décomposition de Meyer, la couche forme est obtenue par projection de l'image d'origine dans un espace qui préserve les contours, pendant que les zones uniformes sont lissées. La lettre est alors complètement extraite et reconnue par l'OCR. De plus, la décomposition de Meyer permet de débruiter l'image d'origine, et les variations de niveaux de gris sont réduites. La lettre est donc mieux identifiée dans le processus d'extraction des zones par une loi de Zipf (extraction des zones les plus fréquentes). Ceci se traduit finalement par un meilleur taux de reconnaissance des OCRs pour les lettres extraites.

Enfin, les cas pour lesquels notre système échoue correspondent à des images très difficiles à traiter. Par exemple, elles correspondent aux lettres composées de plusieurs composantes connexes, aux images très dégradées par le temps, ou encore celles qui ont été numérisées à base résolution. La séparation en plusieurs composantes connexes peut s'expliquer par deux phénomènes principaux : la dégradation du papier à travers les âges et les effets stylistiques. Des exemples sont présentés dans la figure 8.4.5. Le premier cas présente un problème stylistique, puisque l'intérieur de la lettre a été muni d'un liseré noir, tandis que le deuxième présente une lettre composée de plusieurs composantes connexes. Enfin, les deux dernières images présentées correspondent à des lettrines dégradées par le temps.

8.4.4 Règles pour la validation de l'ontologie finale - réduction du fossé sémantique

Nous venons de valider la pertinence des étapes de notre système, de valider l'ontologie issue des connaissances des historiens, et celle issue des traitements des images. La dernière validation que nous proposons a pour but de tester l'idée globale de ce manuscrit, c'est à dire la réduction du fossé sémantique qui existe entre les connaissances des historiens et celles issues des traitements des images. Nous rappelons que pour réduire ce



(a) Image originale

(b) Image simplifiée par les ondelettes de Haar

(c) Formes extraites à partir du résultat des ondelettes de Haar

(d) Lettre extraite à partir des ondelettes de Haar

(e) Image simplifiée par la décomposition de Meyer

(f) Formes extraites à partir de l'approche proposée

(g) Lettre extraite à partir de l'approche proposée

FIGURE 8.4.4: Exemples de résultats d'extraction de la lettre. La première colonne présente les images simplifiées, la seconde correspond aux images obtenues avec l'algorithme associé à la Loi de Zipf. Enfin, la troisième colonne correspond à la lettre extraite. Chaque ligne présente les résultats obtenus pour chaque approche testée et les différences entre les deux résultats justifient les différences dans les taux de reconnaissance

fossé, nous avons construit une ontologie finale. Celle-ci est une mise en correspondance des deux premières ontologies, en liant les concepts images. Ce lien nous permet donc de créer une base de connaissance hétérogène, dans laquelle une même image est décrite par les caractéristiques liées aux traitements des images, et par l'annotation des historiens.

La réduction du fossé sémantique proposée repose donc sur la mise en correspondance de connaissances hétérogènes associée à une même image. Celui-ci est réalisé à l'aide d'une règle d'inférence qui porte sur les deux types de connaissances. Afin de tester cette démarche, nous avons cherché à identifier les zones de l'image qui sont associées à des parties de corps humaines (des zones qui correspondent à des bras ou des jambes par exemple). Cette règle impose donc de savoir si une image contient des personnages (images avec un motif figuratif - information issue de l'ontologie associée à la connaissance des historiens), d'identifier les zones de l'image (information issue de l'ontologie des traitements des images), et d'être sur que cette zone ne correspond pas à la lettre (information déduite de la règle précédente).

Cette règle s'appuie donc sur la première règle, et utilise les deux ontologies qui com-

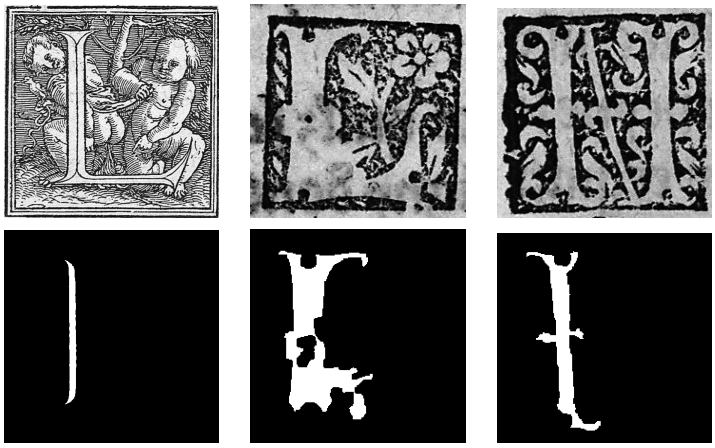


FIGURE 8.4.5: Exemples d'images très difficiles à traiter pour l'extraction de la lettre. La première ligne présente l'image d'origineT, tandis que la seconde présente la forme qui a été considérée comme la lettre par notre système

posent l'ontologie finale. Le but de cette règle est de montrer que le système développé est capable de se servir des connaissances qu'il a déduit de lui-même, et de proposer des requêtes en appui sur les annotations automatiques et manuelles des images. Enfin, cette requête permet également de montrer comment l'utilisation conjointe des connaissances issues des deux domaines, et d'une règle d'inférence adaptée permet de réduire le fossé sémantique qui peut exister entre les deux domaines d'expertise.

Règle pour la propriété *isBody* : détection des parties des corps

La règle associée à la propriété isBody a pour but d'annoter les régions qui correspondent à des parties des corps des personnages qui se situent en arrière plan. Cette règle est spécifique aux lettrines avec un motif figuratif. Nous avons remarqué qu'une forme peut être associée à un corps si :

- Le motif de la lettrine est de type figuratif
- Cette forme est située au centre de l'image
- Elle ne comporte que peu de trous
- Elle est de "couleur" claire (entre gris clair et blanc)
- Elle ne correspond pas à la lettre

À partir de cette description, et en utilisant les critères suivant, une forme est annotée comme faisant partie du corps si :

- La lettrine a un motif, de type figuratif (information extraite de l'ontologie sur la connaissance des historiens)
- Elle intersecte avec la sous-partie numéro 5, et qu'elle ne touche pas les autres (connaissance déduite des relations spatiales entre les formes et les sous-parties)
- La valeur de son nombre d'Euler est comprise entre -2 et 2 (le choix de la valeur 2

a été choisi expérimentalement dans ce cas)
- Le niveau de gris moyen de la forme est supérieur à 128 (valeur médiane de l'espace des niveaux de gris [0 ;255])
- Sa propriété *isLetter* est à faux

Cette règle a été testée sur un sous-ensemble de 45 images. Parmi ces 45 images, 18 ont un motif figuratif, et ces 18 images sont composées de 112 régions de formes. La règle "isBody" a parfaitement annotée les régions de 17 images, comme le présente la figure 8.4.6. Par conséquent, le taux d'erreur est inférieure à 2% pour cette expérimentation.

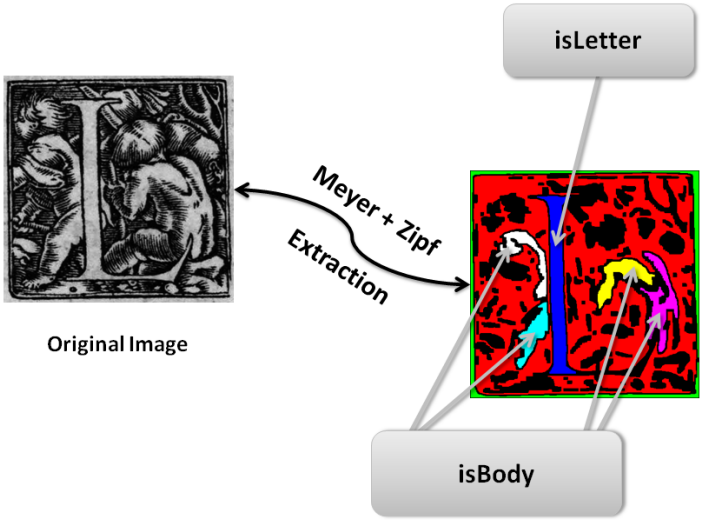
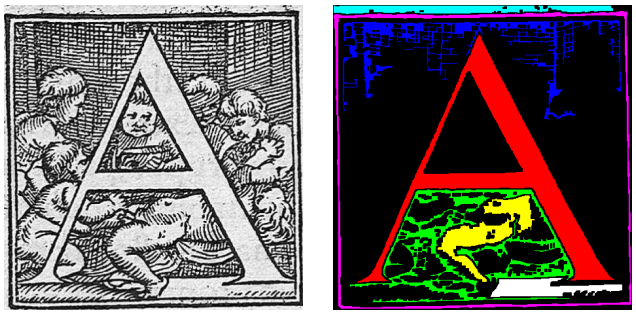


FIGURE 8.4.6: Un exemple d'image ou les zones de formes ont été annotées avec les règles *isLetter* et *isBody*

La lettrine pour laquelle nous n'arrivons pas à extraire correctement la forme qui correspond à une partie du corps du personnage est présentée dans la figure 8.4.7. La région jaune n'est pas annotée comme une région du corps puisque son nombre d'Euler est trop grand, et c'est la région blanche qui est annotée comme étant une partie du corps. Ces deux erreurs pourraient être traitées en prenant en compte plus d'information sur les formes, comme par exemple en ajoutant des caractéristiques plus élaborées.

Synthèse du chapitre

Le dernier chapitre du manuscrit présente l'ontologie finale, qui rassemble les connaissances des historiens et du traitement des images. Nous présentons donc comment nous lions les deux ontologies des domaines pour créer une base de connaissances. Après une brève présentation des outils existant pour insérer des connaissances dans les ontologies,



(a) Image originale (b) Régions extraites automatiquement de l'image

FIGURE 8.4.7: Exemple d'image pour laquelle la partie du corps n'a pas été correctement extraite

nous présentons les données insérées dans les ontologies. Des règles d'inférences sont utilisées pour déduire de nouvelles connaissances à partir des informations insérées dans les ontologies. Ainsi, nous présentons tout d'abord des règles qui permettent de valider le modèle des connaissances des historiens. Puis dans un second temps, nous présentons une règle qui ne s'appuie que sur l'ontologie des traitements des images. Cette règle permet d'extraire la lettre des lettrines, et une validation des résultats est présenté. Enfin, la dernière partie de ce chapitre présente une règle d'inférence qui mêle à la fois des connaissances issues du domaine d'expertise des historiens, et du domaine d'expertise du traitement des images. Cette règle permet de réduire le fossé sémantique existant entre les deux domaines en associant automatiquement un concept de l'ontologie des connaissances des historien à des régions extraites automatiquement.

Conclusion

Les documents de la renaissance représentent un héritage culturel et scientifique inestimable, mémoire de nos sociétés. Afin d'étudier le contenu de ces documents, de nombreux historiens les analysent, s'intéressant aux textes et aux images qui les composent. Malheureusement, tous ces documents se dégradent avec le temps.

Afin de les préserver, de plus en plus de campagnes de numérisation de documents anciens sont initiées en France et à travers l'Europe pour numériser et sauvegarder le contenu de ces documents. Toutes ces campagnes sauvegardent ces centaines de milliers de documents au format image, sans réellement proposer de services d'indexation et de recherche pour les historiens.

Le projet Navidomass (ANR-06-MDCA-012) propose d'étudier ces images de documents anciens pour les décrire et les organiser selon des critères définis par les historiens. Organiser automatiquement des images impose d'analyser leurs contenus pour les décrire et les classer par critères de ressemblance. Ainsi, dans un premier temps de ce manuscrit, nous avons proposé une synthèse des approches de recherche d'information dans de grandes bases. Nous avons étudié la recherche d'images par mots-clés, ou bien interactive. Puis, dans le cas particulier d'outils d'analyse d'images par le contenu, nous avons passé en revue les grandes méthodes de descriptions d'images (description globale des images, description spatiale ou locale). Nous avons terminé cette première partie par un état de l'art des méthodes dédiées aux documents anciens, à leurs caractéristiques, et aux outils qui leur ont été dédiés.

Cette thèse s'intéresse plus particulièrement à l'analyse et la description d'images graphiques de documents anciens. Plus spécifiquement, nous nous sommes intéressés aux lettrines, qui ont été très largement utilisées au cours du temps, et permettent de retracer la vie des ouvrages. Ces images sont complexes puisque composées d'une lettre et d'un fond illustratif d'un point de vue sémantique, et de zones lisses et de zones texturées d'un point de vue radiométrique. Le but de cette thèse est de proposer des outils de recherche d'images graphiques de documents anciens à partir de leurs caractéristiques sémantiques (mots clés des historiens), et à partir d'exemples d'images (CBIR). Cela pose le problème difficile du fossé sémantique qui réside entre des mots-clés et une description du contenu radiométrique utilisés pour décrire les images.

Partant du constat qu'aucune méthode de la littérature ne permet de réduire le fossé sémantique pour des documents anciens, nous avons présenté dans la seconde partie de manuscrit, le modèle proposé et développé au cours de la thèse. Celui-ci propose de rassembler au sein d'une même base des connaissances issues des historiens et des traitements des images. Cette base repose sur l'utilisation de deux ontologies (une liée aux connaissances des historiens, et une pour les traitements d'images), qui permettent de représenter de manière formelle et hiérarchisée les connaissances de deux domaines d'expertises.

Après un rappel sur les logiques de description et les ontologies, nous avons défini pour chaque domaine les concepts utilisés pour modéliser les connaissances. Dans le cas du domaine d'expertise des historiens, nous décrivons les images en utilisant les termes issus du domaine expert historien (thesaurus), et nous avons hiérarchisé cette connaissance pour proposer un premier type de recherche uniquement basé sur les connaissances du domaine historien. Une seconde ontologie, associée au domaine du traitement des images de documents anciens, a été définie suivant un modèle d'analyse de la complexité de ces documents. Ce modèle est présenté dans la troisième partie de ce manuscrit.

Le modèle d'analyse de la complexité des documents anciens que nous proposons dans ce manuscrit repose sur le paradigme de la complexité définit par Edgar Morin [Morin 96], et propose d'étudier les images graphiques de documents anciens en les caractérisants de manière adaptée. Ainsi, avant même de décrire une image, nous proposons de la simplifier en séparant les différentes informations qui la composent (formes et traits) au sein de couches spécifiques. Pour chaque couche, nous avons défini des signatures statistiques et structurelles spécifiques au type d'information qu'elle renferme. La combinaison de ces signatures permet d'extraire d'une part les motifs caractéristiques de la couche, et de ne retenir que les plus pertinents, et d'autre part d'extraire des régions d'intérêts de ces images et de décrire leur organisation spatiale.

Partant de ces caractérisations adaptées, nous avons démontré que la description d'une image à partir d'un seul type d'information n'est pas pertinent. Nous les combinons donc pour obtenir une description complexe des images, et nous prouvons que cette fusion permet d'accroître considérablement les résultats de reconnaissances des images.

Enfin, la dernière partie de ce manuscrit a été consacrée à la définition d'une ontologie globale permettant de lier les différentes connaissances du domaine historien et du domaine du traitement d'images. À partir de ces ontologies, des règles d'inférences nous permettent de combiner des connaissances issues des deux domaines, de déduire une connaissance de plus haut niveau sur les images et ainsi de réduire le fossé sémantique. C'est donc la combinaison de ces deux connaissances diverses qui nous a permis de définir un moteur de recherche qui s'appuie à la fois sur les connaissances des historiens, et sur celle issue des traitements automatiques des images. Cette combinaison de connaissance propose ainsi aux utilisateurs de rechercher des images, soit uniquement à partir des connaissances issues du domaine des historiens, soit uniquement en comparant des

images en utilisant une approche type CBIR, soit finalement en combinant annotations manuelles et descriptions automatiques.

Discussions et perspectives

Caractérisation des images de documents anciens

Les signatures développées sur la couche forme et la couche textures, dans la partie 5 de ce manuscrit, visent pour le moment à décrire les images de manière statistique ou structurelle. Les signatures structurelles utilisées se limitent à la description des relations topologiques entre les régions extraites. Afin d'améliorer la pertinence, il pourrait être envisagé d'enrichir ces graphes pour décrire le contenu des régions, à l'aide des sacs de motifs que nous avons défini, ou de signatures statistiques adaptées.

De plus, nous avons pour le moment présenté une fusion tardive de ces signatures, car celle-ci permet de sélectionner la classe majoritaire par rapport aux résultats des différentes signatures. Il serait intéressant d'envisager une fusion précoce de ces résultats, en plongeant les graphes dans espaces vectoriels, via des approches de *graphes embedding* [Bunke 11], et de rechercher les caractéristiques les plus pertinentes (à l'aide d'algorithme de sélection de caractéristiques).

Nous avons pour le moment limité la simplification des images à une couche forme et une couche textures, afin de répondre aux spécificités des documents anciens. Afin d'étendre notre modèle d'analyse d'images en couches à une plus grande variété d'images, il serait intéressant d'envisager l'intégration d'autres couches d'information, comme une couche couleur par exemple. L'extraction de motif pourrait alors reposer sur une approche similaire à celle de la couche forme (quantification couleur, et extraction des motifs les plus fréquents).

Concernant l'extraction de régions et leur sélection, nous avons présenté dans la fin de manuscrit, une méthode pour extraire la lettre des lettrines. Cette approche souffre de la recherche de composantes connexes dans l'extraction des régions. Ainsi, toutes les lettres qui sont coupées par des ornements ne sont pas directement extractibles. Une perspective intéressante serait de ne plus considérer une seule composante connexe, mais plutôt un ensemble de régions voisines de celle sélectionnée qui partageraient des caractéristiques similaires (par exemple la même orientation ou la même épaisseur).

Modélisation ontologique des documents anciens

Nous avons présenté, au travers de ce manuscrit, les modèles que nous proposons pour modéliser les connaissances des historiens et les connaissances issues de traitements d'images appliqués aux images graphiques de documents anciens. L'ontologie des

traitements d'images a été peuplée et validée à partir des régions de formes des images. Cette ontologie nous a permis d'identifier des éléments sémantiques des images (comme la lettre ou des parties de personnages) mais nous nous sommes limités à une connaissance issue des formes. Nous somme actuellement en train de travailler à l'intégration de connaissances à partir des régions de traits. Cette information devrait nous permettre d'identifier les styles de fond dans les lettrines (en particulier les images composées d'un fond hachuré), et de valider ou d'invalider des régions identifiées comme « isBody » dans le dernier chapitre de cette thèse (une région forme associée à une partie de corps est souvent liée à une partie de traits qui correspond à son ombre). De plus, l'intégration de régions possédant des propriétés différentes nous permettra d'enrichir notre modèle et de vérifier sa pertinence quand à l'intégration de connaissances hétérogènes.

Enfin, bien que nous ayons conçu ce modèle de la manière la plus générique possible, il est pour le moment limité aux images de documents anciens. Afin de valider sa pertinence sur un plus grand type d'images, nous envisageons de l'étendre à d'autres images de documents. En ce sens, des travaux sont actuellement en cours au sein du laboratoire L3i, pour proposer d'extraire automatiquement une connaissance de plus haut niveau sur des bandes dessinées.

Il faut également noter que notre modèle s'appuie sur des connaissances d'un domaine d'expertise. Extraire une connaissance sémantique sur les images, à partir uniquement de traitements des images, est complexe et difficilement généralisable (le même algorithme appliqué à différents types d'images pourra aboutir à l'extraction d'éléments sémantiques très différents). Notre modèle reste donc orienté sur la modélisation de connaissances spécifiques à un domaine, et il permet de répondre à des requêtes ciblées pour un domaine.

Règles d'inférences et réduction du fossé sémantique

Le modèle présenté vise à réduire le fossé sémantique entre deux domaines de compétences à l'aide de règles d'inférences. Passer de résultats de traitements des images à une connaissance sémantiquement plus riche repose donc sur la définition de ces règles. Définir une règle nécessite ainsi la présence d'experts du domaine d'application (les historiens dans notre cas) et d'experts du domaine du traitement des images. Notre approche globale n'est donc pas, pour le moment, évolutive puisque les règles sont définies de manière figée dans l'ontologie.

Une des perspectives que nous pensons à plus long terme, de ce modèle, serait d'intégrer des mécanismes similaires à ceux utilisés en bouclage de pertinences pour définir automatiquement de nouvelles règles. Ainsi, un utilisateur pourrait proposer des images qui partagent une propriété commune au système. Le système pourrait rechercher les connaissances partagées par ces images (connaissances issues des deux domaines d'expertise) pour les associer à une règle. Cette perspective peut-être assimilée à un système de sélection de caractéristiques hétérogènes, guidées par l'utilisateur.

Annexes

Annexe A

Physique du papier

A.1 Le papier

Le papier est apparu en Chine il y a environ 2000 ans et était initialement composé d'éléments composés de fibres tels que le chanvre, le bambou, la paille de riz ou de blé comme le précise [Remazeilles 01]. Le papier a ensuite été introduit en europe par l'intermédiaire des italiens au XIème siècle et la France commence à en produire au cours du XIVème siècle. Avec l'apparition de l'imprimerie et l'utilisation de plus en plus répandue du papier, le bois est utilisé à partir des années 1870 dans la babrication du papier. On distingue deux méthodes principales pour extraire les fibres qui composent le bois :

- Un procédé mécanique dans lequel un défibreur est utilisé pour séparer les fibres. Le rendement de cette méthode est élevé mais au détriment de la qualité. En effet, le bois est composé de lignine et d'hémicelluloses qui se retrouvent dans le papier. De ce fait, le papier obtenu par cette méthode est généralement utilisé pour des journaux ou des emballages à faibles prix car ils se dégradent rapidement dans le temps.
- Un procédé chimique qui vise à éliminer la lignine et les hémicelluloses du bois pour obtenir un papier plus résistant, de meilleur qualité et plus blanc. En contrepartie, ce procédé est plus polluant, plus onéreux et de rendement moindre.

Le papier est composé de trois éléments : **les fibres, la charge** et **l'encollage**. Les fibres comme nous venons de le voir ci-dessus, sont principalement issues de matériaux organiques, comme le lin, le coton ou le bois depuis les années 1870. Afin d'améliorer les qualités et les propriétés du papier, des produits minéraux sont ajoutés lors de la fabrication du papier et composent la charge. Leur but est de favoriser sa blancheur, sa stabilité et sa tenue à l'impression. Enfin, une dernière phase d'encollage sert à réduire l'aspect hydrophile du papier pour conserver l'encre à la surface de celui-ci.

A.2 L'encre

L'encre est une préparation composée de **pigments colorés** et d'un **liant**. Les pigments donnent la couleur tandis que le liant permet d'amalgamer et de faciliter l'application sur le support. L'origine des encres est très ancienne et difficile à situer dans le temps puisque l'on sait d'après [Remazeilles 01], qu'elles étaient employées au IVème millénaire avant notre ère. Parmi les encres anciennes, on peut citer les deux catégories suivantes :

- Les encres au carbone qui sont composées d'un pigment noir issu de produits carbonisés. Plus cette pourdre sera fine et légère, plus la qualité de l'impression sera bonne. A ces pigments sont associés un liant (d'origines très variées) qui doit présenter l'avantage de pouvoir sécher à l'air lorsqu'il est déposé en fine couche.
- Les encres Mettalogalliques qui sont quant à elles composées de trois éléments :

    1. Un tannin composé d'une substance végétale contenant plusieurs acides phénoliques dont l'acide gallique.
    2. Un sel métallique se présentant généralement sous forme de sulfate (de fer ou de cuivre) qui donnaient la couleur souhaitée à l'encre. Ces sels étaient nommés vitriol dans les recettes anciennes.
    3. Un liant qui a le même but que dans les encres carbones.

Le pigment est obtenu par réaction chimique entre le tannin et les ions métalliques des sels. Ce mélange, à dominante noir, vire au bleu, violet ou vert en fonction de l'ion métallique employé. Les tannins les plus couramment utilisés étaient sensibles à l'hydrolyse et libéraient de l'acide gallique. C'est ainsi que l'on parle, dans le cas de l'ion fer, d'**encre ferrogallique**. Les encres ferrogalliques ont la propriété d'être très acide et de contenir du fer libre qui attaque le papier, le dégrade et provoque des taches rougeâtres autour du texte. Dans notre cas, les documents que nous traiterons dans ce rapport sont issus d'imprimeries du XVème et XVIème siècle et sont composés d'encres carbones. La dégradation du papier n'est du qu'aux effets de la dégradation du papier lui-même et de la cellulose qui le compose.

Annexe B

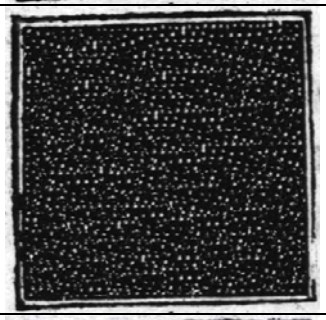
Caractéristiques des lettrines

Cette deuxième annexe présente les résultats de la décomposition d'une lettrine telle qu'elle est réalisée par les historiens. Suite à une rencontre le 2 février 2009 entre les historiens et les informaticiens du projet, a été définit une taxonomie des différents éléments que l'on peut trouver dans une lettrine, et leurs différentes variantes. Ces éléments sont repris dans cette annexe pour décrire en détails les éléments constitutifs des lettrines.

BIBLIOTHEQUES VIRTUELLES HUMANISTES

INDEXATION DES LETTRINES

Rédaction : Alice Nué ; Rémi Jimenes.

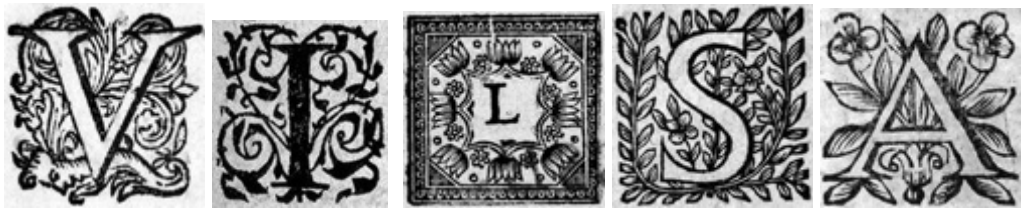
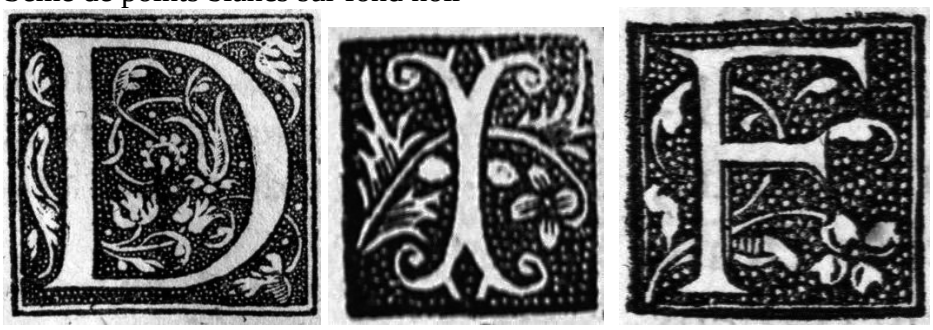
	LETTRINE
	**FOND**
	**MOTIF**
	**LETTRE**
	**CADRE**

Remarques préliminaires :

– Comme remarqué lors de la réunion Navidomass du 02 février 2009, nous considérons que la lettrine est constituée de trois couches superposées : 1) le fond, 2) le motif, 3) la lettre, l'ensemble étant inséré dans un cadre. Le schéma ci-dessus illustre ces quatre « zones ».

– Le « style » des lettrines constitue une notion trop floue pour être utilisée comme critère d'indexation.

I. LE FOND

Type de fond	Définition et exemples
Noir	Fond noir uniforme, sans criblage ni hachures. 
Blanc	Fond blanc uniforme 
Hachuré	Stries parallèles 
Criblé	Semé de points blancs sur fond noir 
Indéfini	Fond indéfini ou inclassable (voir liste des « cas problématiques »)

II. LE MOTIF

Le « motif » est constitué de tous les objets concrets représentés à fin d'ornementation. Il se superpose au fond.

Notre projet serait d'effectuer une indexation sous *Iconclass* sur un lot de lettrines présentant un motif significatif (indexation manuelle). Nous aimerions ensuite automatiser l'indexation des autres lettrines grâce au logiciel de recherche de similarité de motifs. Par similarité, les lettrines pourraient être indexées automatiquement.
Afin de faciliter l'indexation manuelle des lettrines, il nous serait utile de distinguer d'une part les motifs décoratifs (qui ne pourront pas être indexés sous *Iconclass*) et d'autre part les motifs figuratifs.

Motif	Exemples
Décoratif	Lettrines proposant des motifs récurrents, souvent symétriques.   
Figuratif	Lettrines illustrées de scènes historiées (personnages, animaux, architecture, paysage…)   

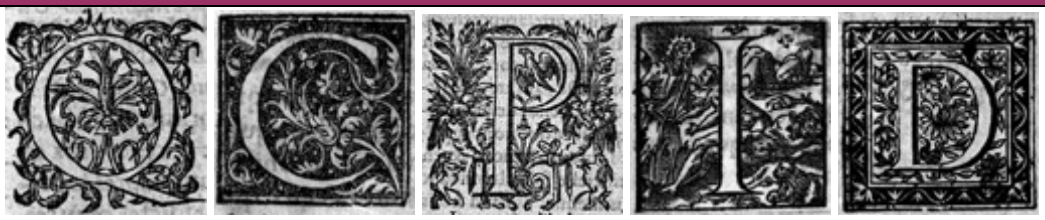
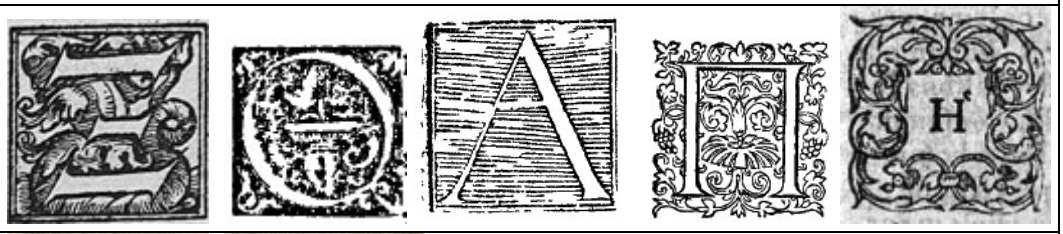
III. LA LETTRE

L'identification de la lettre devrait pouvoir se faire, dans la mesure du possible, à plusieurs niveaux.

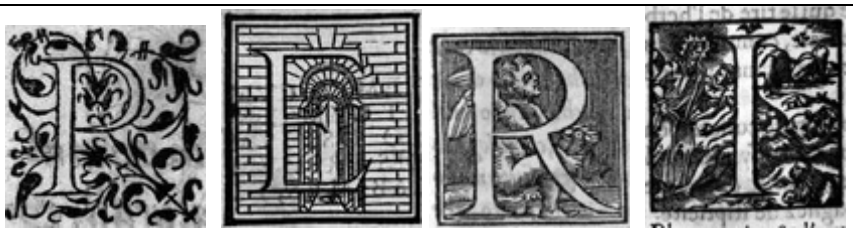
1. Identification de la lettre

Lettrine	Lettre
	B
	L

2. Alphabet

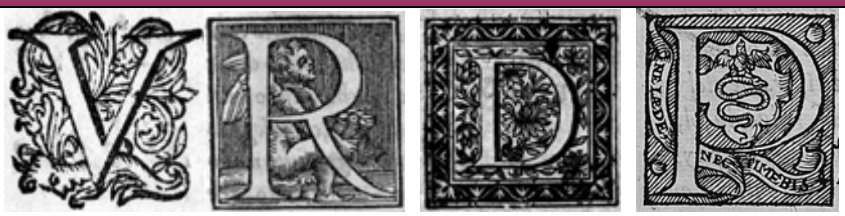
Type d'alphabet	Exemples
Latin	
Grec	
Hébreu	

3. Couleur de la lettre

Couleur de la lettre	Exemples
Noir	
Blanc	

4. Police (famille d'écriture)

Cette notion n'est valable que pour les lettres latines.

Type d'alphabet	Exemples
Romain	
Gothique	
Indéfini	tous autres alphabets, tous autres styles (civilité, italiques,…)

IV. LE CADRE

Le cadre définit le contour (silhouette) de la lettrine (bloc typographique). Il peut ne pas apparaître, être constitué d'un simple trait noir, de traits multiples voire d'un encadrement.

Type de cadre	Exemples
Aucun	
1 filet	
2 filets	
3 filets	[rare, aucun exemple dans la base]
Encadrement	
Passe-partout	Un passe-partout est un cadre de bois ou de métal, à l'intérieur duquel on insère un caractère typographique de plomb. 
Autre	

CAS PROBLEMATIQUES ET PARTIS PRIS

(Propositions de Rémi pour la vérification de la base de lettrines et l'indexation de nouvelles lettrines).

Lettrine	Indexation	Justification
	Fond hachuré. Motif figuratif. Lettre E, alphabet latin, blanche, romain. Cadre de deux filets Dimensions inconnues	L'ambiguïté concerne le fond : les hachures font ici clairement partie du « motif » (elles dessinent le relief du mur et sont donc figuratives). Au sens strict, cette lettrine ne montre aucun fond (puisque le motif le recouvre entièrement). On choisit cependant de l'indexer en fond hachuré parce l'effet visuel (« couleur » de la lettrine) est identique à celui des hachures, et parce que cette lettrine s'inscrit dans une série comportant par ailleurs des lettres à fond hachuré.
	Fond criblé Motif [décoratif] Lettre d, alphabet latin, blanche, gothique. Sans filet Dimensions inconnues.	Nouvelle ambiguïté sur le fond : la lettre est usée, ou bien l'encre trop grasse, et l'ornement semble avoir un fond « noir » plein. Seuls quelques points blancs apparaissent, qui montre que le fond est criblé.
	Fond criblé. Motif [décoratif]. Lettre V, alphabet latin, blanche, gothique. Cadre d'un filet simple. Dimensions inconnues	Cette lettrine, employée comme un A par Henri Estienne au feuillet 241 de l'édition de Porphyre (1511) est en réalité un V retourné, comme en témoigne l'orientation des tiges de la fleur. Ce bois est d'ailleurs utilisé comme V au feuillet 233, et doit être indexé comme tel.

Publications

Les travaux présentés dans ce manuscrit ont donné lieu à plusieurs publications que nous présentons ci-dessous

Article dans des revues avec comité de lecture

Towards historical document indexing : extraction of drop cap letters, **Coustaty, Mickael** and Pareti, Rudolf and Vincent, Nicole and Ogier, Jean-Marc, (2011). In International Journal on Document Analysis and Recognition, Springer Berlin / Heidelberg, Issn : 1433-2833, [lien]

Chapitres de livres avec comité de lecture

Segmenting and Indexing Old Documents Using a Letter Extraction, **Mickaël Coustaty**, Sloven Dubois, Jean-Marc Ogier, and Michel Menard (2010), pages 142-149, [lien]

Analyzing Old Documents Using a Complex Approach : application to lettrines indexing, **Mickael Coustaty**, Vincent Courboulay, and Jean-Marc Ogier, to appear in Advances in Knowledge Discovery and Management vol. 2, Springer

Conférences invité

Historical Document Analysis : A Review of French Projects and Open Issues, **Mickael Coustaty**, Romain Raveaux and Jean-Marc Ogier, (2011), to appear in 19th EUropean SIgnal Processing COnference

Conférences internationales avec comité de lecture

Bags of Strokes Based Approach for Classification and Indexing of Drop Caps - Thi Thuong Huyen Nguyen, Jean-Marc Ogier, **Mickael Coustaty** (2011), p349-353 in ICDAR 2011

Using ontologies to reduce the semantic gap between historians and image processing algorithms - **Mickael Coustaty**, Alain Bouju, Karell Bertet, Georges Louis (2011),

p156-160 in ICDAR 2011

Discrimination of old document images using their style - **Mickael Coustaty**, Jean-Marc Ogier (2011), p389-393 in ICDAR 2011

Ancient Documents Denoising and Decomposition for Indexing - **Mickael Coustaty**, Sloven Dubois, Michel Menard, Jean-Marc Ogier - In Ninth IAPR International Workshop on Graphics Recognition, to appear (2011)

NAVIDOMASS : Structural-based approaches towards handling historical documents - Salim Jouili, **Mickael Coustaty**, Salvatore Tabbone, and Jean-Marc Ogier (2010), 1-4. In ICPR

Stroke feature extraction for lettrine indexing - Giap Nguyen, **Mickael Coustaty**, and Jean-marc Ogier (2010), pages 355-360. In 2nd IEEE International Conference on Image Processing Theory, Tools and Applications

Content-Based Old Documents Indexing - **Mickael Coustaty**, et al. In Eighth IAPR International Workshop on Graphics Recognition, pages 217-223 (2009)

Drop Caps Decomposition for Indexing a New Letter Extraction Method - **Mickael Coustaty**, Jean-Marc Ogier, Rudolf Pareti, and Nicole Vincent In 10th International Conference on Document Analysis and Recognition, 476-480 (2009) [lien]

Conférences nationales avec comité de lecture

Approche complexe de l'analyse de documents anciens - **Mickael Coustaty**, Giap Nguyen, Vincent Courboulay, and Jean-Marc Ogier, In RNTI E-19 EGC'2010 - Revue des Nouvelles Technologies de l'Information, 597—608 (2010)

Extraction et caractérisation de traits dans les lettrines - **Mickael Coustaty**, Giap Nguyen, Jean-Marc Ogier, In Actes du colloque international francophone sur l'écrit et le document, 1-15 (2010)

Extraction d'informations d'images de documents anciens - **Mickaël Coustaty,** Jean-marc Ogier, Rudolf Pareti, et Nicole Vincent, In Actes de la douzième conférence ORASIS (2009) [lien]

Bibliographie

[Adam 00a] S. Adam, J.M. Ogier, C. Cariou, R. Mullot, J. Labiche & J. Gardes. *Symbol and character recognition : application to engineering drawings*. International Journal on Document Analysis and Recognition, vol. 3, no. 2, pages 89–101, 2000.

[Adam 00b] Sébastien Adam, Jean-Marc Ogier, Claude Cariou, Rémy Mullot, Joel Gardes & Yves Lecourtier. Fourier-mellin based invariants for the recognition of multi-oriented and multi-scaled shapes - application to engineering drawings analysis, volume 42 of *Machine Perception and Artificial Intelligence*, pages 125–146. World Scientific Publishing Co., 2000.

[Alboody 10] Ahed Alboody, Florence Sèdes & Jordi Inglada. *Enriching The Qualitative Spatial Reasoning System RCC8*. In Shyamanta M. Hazarika, editeur, Qualitative Spatio-Temporal Representation and Reasoning : Trends and Future Directions. Information Science Reference Publishing, août 2010.

[Alhichri 02] Haikel Salem Alhichri & Mohamed Kamel. *Multi-resolution image registration using multi-class Hausdorff fraction*. Pattern Recogn. Lett., vol. 23, pages 279–286, January 2002.

[Allen 83] James F. Allen. *Maintaining knowledge about temporal intervals*. Commun. ACM, vol. 26, pages 832–843, November 1983.

[Andrews 71] H. C Andrews. *Multidimensional Rotations in Feature Selection*. IEEE Transactions on Computers, vol. 20, page 1045–1051, September 1971. ACM ID : 1311482.

[Aujol 05] Jean-François Aujol & Antonin Chambolle. *Dual Norms and Image Decomposition Models*. International Journal of Computer Vision, vol. 63, no. 1, pages 85–104, 2005.

[Aujol 06a] Jean-François Aujol, Guy Gilboa, Tony Chan & Stanley Osher. *Structure-Texture Image Decomposition - Modeling, Algorithms, and Parameter Selection*. International Journal of Computer Vision, vol. 67, no. 1, pages 111–136, 2006.

[Aujol 06b] Jean-Francois Aujol & Sung Ha Kang. *Color image decomposition and restoration.* J. Visual Communication and Image Representation, vol. 17, no. 4, pages 916–928, 2006.

[Aurenhammer 91] Franz Aurenhammer. *Voronoi diagrams a survey of a fundamental geometric data structure.* ACM Comput. Surv., vol. 23, pages 345–405, September 1991.

[Baader 03] Franz Baader & Werner Nutt. *Basic description logics.* pages 43–95, 2003.

[Baeza Yates 99] Ricardo A. Baeza Yates & Berthier R. Neto. Modern Information Retrieval. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1999.

[Bannour 11] Hichem Bannour & Celine Hudelot. *Towards ontologies for image interpretation and annotation.* In Proceedings of the CBMI'2011, page to appear. IEEE computer society, 2011.

[Barnard 03] Kobus Barnard, Pinar Duygulu, David Forsyth, Nando de Freitas, David M. Blei & Michael I. Jordan. *Matching words and pictures.* J. Mach. Learn. Res., vol. 3, pages 1107–1135, March 2003.

[Barrat 09] Sabine Barrat. *Modeles graphiques probabilistes pour la reconnaissance de formes.* PhD thesis, Universite de Nancy 2, 2009.

[Baudrier 07a] Etienne Baudrier, Nathalie Girard & Jean-Marc Ogier. *A Non-symmetrical Method of Image Local-Difference Comparison for Ancient Impressions Dating.* In Seventh IAPR International Workshop on Graphics Recognition (GREC'07), Curitiba Brésil, 2007.

[Baudrier 07b] Étienne Baudrier, Gilles Millon, Frédéric Nicolier, Ralph Seulin & Su Ruan. *Hausdorff distance-based multiresolution maps applied to an image similarity measure.* Imaging Science Journal, vol. 55, no. 3, pages 164–174, 2007.

[Baudrier 08] Étienne Baudrier, Frédéric Nicolier, Gilles Millon & Su Ruan. *Binary-image comparison with local-dissimilarity quantification.* Pattern Recogn., vol. 41, pages 1461–1478, May 2008.

[Bay 06] Herbert Bay, Tinne Tuytelaars & Luc Van Gool. *Surf : Speeded up robust features.* In European Conference on Computer Vision, pages 404–417, 2006.

[Belkasim 91] S.O. Belkasim, M. Shridhar & M. Ahmadi. *Pattern Recognition With Moment Invariants : A Comparative Study and New Results.* Pattern Recognition, vol. 24, no. 12, pages 1117–1138, 1991.

[Belongie 02] S. Belongie, J. Malik & J. Puzicha. *Shape Matching and Object Recognition Using Shape Contexts*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, page 509–522, April 2002. ACM ID : 628792.

[Besag 74] Julian Besag. *Spatial Interaction and the Statistical Analysis of Lattice Systems*. Journal of the Royal Statistical Society. Series B (Methodological), vol. 36, no. 2, pages pp. 192–236, 1974.

[Bibliotheek 11] Koninklijke Bibliotheek. *Digitisation by Proquest of early printed books in KB collection*, 2011. http ://www.kb.nl/nieuws/2011/proquest-en.html.

[Bigun 96] J. Bigun, S. K Bhattacharjee & S. Michel. *Orientation Radiograms for Image Retrieval : An Alternative to Segmentation*. 1996.

[BLSB] BLSB. *British Library Search Beta*. http ://search-beta.bl.uk/primo_library/libweb/action/search.do.

[BNF] BNF. *Bibliothèque nationale de France*. http ://www.bnf.fr/.

[BNF 80] BNF. *RAMEAU - Répertoire d'autorité-matière encyclopédique et alphabétique unifié - Bibliothèque nationale de France*. http ://rameau.bnf.fr/, 1980.

[BNS] BNS. *Bibliothèque Nationale Suisse*. http ://www.nb.admin.ch/slb/index.html ?lang=fr.

[Bodic 09] Pierre Le Bodic, Hervé Locteau, Sébastien Adam, Pierre Héroux, Yves Lecourtier & Arnaud Knippel. *Symbol Detection Using Region Adjacency Graphs and Integer Linear Programming*. In 10th International Conference on Document Analysis and Recognition, pages 1320–1324. IEEE Computer Society, 2009.

[Borgwardt 07] Karsten M. Borgwardt. *Graph Kernels*. PhD thesis, Ludwig-Maximilians-University Munich, 2007.

[Breu 95] Heinz Breu, Joseph Gil, David Kirkpatrick & Michael Werman. *Linear Time Euclidean Distance Transform Algorithms*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 17, pages 529–533, 1995.

[Brut 09] Mihaela Brut, Dana Al Kukhun & Florence Sèdes. *Towards Service-Oriented Semantic Annotation and Retrieval within Pervasive E-Learning Systems*. International Journal of Business Intelligence and Data Mining, Advances in Intelligent Information Management Systems and Applications, vol. 4, no. 2, pages 122–140, juillet 2009.

[Bui 99] T. D Bui & G. Chen. *Invariant Fourier-Wavelet Descriptor For Pattern Recognition.* Pattern Recognition, vol. 32, pages 1083—1088, 1999.

[Bunke 01] Horst Bunke, Simon Günter & Xiaoyi Jiang. *Towards Bridging the Gap between Statistical and Structural Pattern Recognition : Two New Concepts in Graph Matching.* In Proceedings of the Second International Conference on Advances in Pattern Recognition, ICAPR '01, pages 1–11, London, UK, 2001. Springer-Verlag.

[Bunke 11] Horst Bunke & Kaspar Riesen. *Recent advances in graph-based pattern recognition with applications in document analysis.* Pattern Recognition, vol. 44, no. 5, pages 1057–1067, 2011.

[BVH] BVH. *Les Bibliothèques Virtuelles Humanistes - Centre d'Etude Supérieur de la Renaissance - http ://www.bvh.univ-tours.fr/.*

[C. Rosenberger 01] K. Chehdi C. Rosenberger. *Nouveaux attributs de texture invariants par rotation.* Traitement du signal, vol. 18, no. 5-6, pages 447–454, 2001.

[Canny 86] J Canny. *A computational approach to edge detection.* IEEE Trans. Pattern Anal. Mach. Intell., vol. 8, no. 6, pages 679–698, 1986.

[Cao 95] Jun Cao, M. Ahmadi & M. Shridhar. *Recognition of handwritten numerals with multiple feature and multistage classifier.* Pattern Recognition, vol. 28, no. 2, pages 153–160, February 1995.

[Caron 03] Yves Caron, Harold Charpentier, Pascal Makris & Nicole Vincent. *Power Law Dependencies to Detect Regions of Interest.* In DGCI, pages 495–503, 2003.

[Chambolle 05] A. Chambolle. *Total variation minimization and a class of binary MRF models.* EMMCVPR, vol. 3757 of Lecture Notes in Computer Sciences, pages 136–152, 2005.

[Chen 03] V. Chen, A. Szabo & M. Roussel. *Recherche d'images iconiques utilisant les moments de Zernike.* In Actes de CORESA'03, Lyon, 2003.

[Choksuriwong 05] A. Choksuriwong, H. Laurent & B. Emile. *Etude Comparative de Descripteurs Invariants d'Objets.* In ORASIS'05 - Congrès des jeunes chercheurs en vision par ordinateur, Mai 2005.

[Choksuriwong 07] Anant Choksuriwong. *Reconnaissance d'objets dans une image - Application à la biométrie et à la robotique mobile.* PhD thesis, Université d'Orléans, 2007.

[Chouaib 09] Hassan Chouaib, Florence Clopet & Nicole Vincent. *Graphical Drop Caps Indexing.* In Eighth IAPR International Workshop on Graphics Recognition, pages 179–185, La Rochelle, 2009.

[Chowdhury 10] Pinaki. Chowdhury, Sukhendu. Das, Suranjana. Samanta & Utthara. Mangai. *A Survey of Decision Fusion and Feature Fusion Strategies for Pattern Classification.* IETE Technical Review, vol. 27, no. 4, pages 293–307, 2010.

[Ciocca 99] G. Ciocca, I. Gagliardi & R. Schettini. *Quicklook : A Content-Based Image Retrieval System with Learning Capabilities.* International Conference on Multimedia Computing and Systems, vol. 2, page 1028, 1999.

[Comscore, Inc] Comscore, Inc. *U.S. Search Engine Rankings.*

[Conte 04] D. Conte, P. Foggia, C. Sansone & M. Vento. *Thirty Years Of Graph Matching In Pattern Recognition.* International Journal of Pattern Recognition and Artificial Intelligence, 2004.

[Coquin 01] D. Coquin & Ph. Bolon. *Application of Baddeley's distance to dissimilarity measurement between gray scale images.* Pattern Recognition Letters, vol. 22, no. 14, pages 1483 – 1502, 2001.

[Coüasnon 03] Bertrand Coüasnon & Ivan Leplumey. *A Generic Recognition System for Making Archives Documents accessible to Public.* In Proceedings of the Seventh International Conference on Document Analysis and Recognition - Volume 1, pages 228–232, Washington, DC, USA, 2003. IEEE Computer Society.

[Coustaty 08] Mickael Coustaty, Stephanie Guillas, Muriel Visani, Karell Bertet & Jean-Marc Ogier. *On the Joint Use of a Structural Signature and a Galois Lattice Classifier for Symbol Recognition.* In Graphics Recognition. Recent Advances and New Opportunities, volume 5046 of *LNCS*, pages 61–70. 2008.

[Coustaty 11] M. Coustaty, K. Bertet, M. Visani & J.-M. Ogier. *A New Adaptive Structural Signature for Symbol Recognition by Using a Galois Lattice as a Classifier.* IEEE Transactions on Systems, Man, and Cybernetics, Part B : Cybernetics, vol. 41, no. 99, pages 1–13, 2011.

[Cover 67] T. Cover & P. Hart. *Nearest neighbor pattern classification.* IEEE Transactions on Information Theory, vol. 13, no. 1, pages 21 – 27, jan 1967.

[Croft 88] W. B. Croft & D. J. Harper. *Using probabilistic models of document retrieval without relevance information.* pages 161–171, 1988.

[Cross 83] George R. Cross & Anil K. Jain. *Markov Random Field Texture Models.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 5, no. 1, pages 25 –39, jan. 1983.

[Datta 08] Ritendra Datta, Dhiraj Joshi, Jia Li & James Z. Wang. *Image retrieval : Ideas, influences, and trends of the new age.* ACM Comput. Surv., vol. 40, pages 5 :1–5 :60, May 2008.

[Davies 04] M. Davies. *Identifiability issues in noisy ICA.* IEEE Signal Processing Letters, vol. 11, no. 5, pages 470–473, 2004.

[Delalandre 09] Mathieu Delalandre. *Retrieval of the ornaments from the Hand-Press Period : an overview.* In Internation Conference on Document Analysis and Recognition, volume 2, pages 496–500, Barcelona, Spain, 2009.

[Delalandre 10] Mathieu Delalandre, Ernest Valveny, Tony Pridmore & Dimosthenis Karatzas. *Generation of synthetic documents for performance evaluation of symbol recognition and spotting systems.* International Journal on Document Analysis and Recognition, vol. 13, pages 187–207, 2010.

[Dempster 77] A. P. Dempster, N. M. Laird & D. B. Rubin. *Maximum likelihood from incomplete data via the EM algorithm.* Journal of the Royal Statistical Society. Series B. Methodological, vol. 39, no. 1, pages 1–38, 1977. With discussion.

[Deriche 87] Rachid Deriche. *Using Canny's criteria to derive a recursively implemented optimal edge detector.* International Journal of Computer Vision, vol. 1, no. 2, pages 167–187, 1987.

[Derin 87] H. Derin & H. Elliott. *Modeling and segmentation of noisy and textured images using Gibbs random fields.* IEEE Trans. Pattern Anal. Mach. Intell., vol. 9, pages 39–55, January 1987.

[Deselaers 08] Thomas Deselaers, Daniel Keysers & Hermann Ney. *Features for image retrieval : an experimental comparison.* Information Retrieval, vol. 11, pages 77–107, 2008.

[Dewey 76] Melvil Dewey. *Classification décimale de Dewey.* http ://www.oclc.org/fr/fr/default.htm, 1876.

[DMOZ] DMOZ. *Open Directory Project - http ://www.dmoz.org/.*

[Doermann 98] David Doermann. *The indexing and retrieval of document images : a survey.* Comput. Vis. Image Underst., vol. 70, pages 287–298, June 1998.

[Dosch 04] P. Dosch & J. Lladós. *Vectorial signatures for symbol discrimination.* In J. Lladós & Y.B. Kwon, editeurs, Graphics Recognition : Recent Advances and Perspectives - Selected papers from GREC'03, volume 3088, pages 154–165. LNCS, 2004.

[Dubois 08] S. Dubois, M. Lugiez, R. Péteri & M. Ménard. *Adding a Noise Component To A Color Decomposition Model For Improving Color Texture Extraction.* CGIV 2008 and MCS'08 Final Program and Proceedings, pages 394–398, 2008.

[Dubois 10] Sloven Dubois. *Décompositions spatio-temporelles pour l'etude des textures dynamiques : contribution a l'indexation video.* PhD thesis, Université de La Rochelle, 2010.

[Duda 72] Richard O. Duda & Peter E. Hart. *Use of the Hough transformation to detect lines and curves in pictures.* Commun. ACM, vol. 15, pages 11–15, January 1972.

[Duda 00] Richard O. Duda, Peter E. Hart & David G. Stork. Pattern classification. Wiley-Interscience, 2 edition, October 2000.

[Duygulu 06] P. Duygulu, K. Barnard, J. de Freitas & D. Forsyth. *Object Recognition as Machine Translation : Learning a Lexicon for a Fixed Image Vocabulary*, 2006.

[Eglin 99] Véronique Eglin, Stéphane Bres & Hubert Emptoz. *Structuration de documents par repérage de zones d'intérêt.* Traitement du Signal, vol. 16, pages 217–239, 1999.

[Escalera 11] S. Escalera, A. Forné ands, O. Pujol, J. Lladós ands & P. Radeva. *Circular Blurred Shape Model for Multiclass Symbol Recognition.* IEEE Transactions on Systems, Man, and Cybernetics, Part B : Cybernetics, vol. 41, no. 2, pages 497 –506, april 2011.

[Etemadi 91] A. Etemadi, J.-P. Schmidt, G. Matas, J. Illingworth & J. Kittler. *Low-Level Grouping of Straight Line Segments.* British Machine Vision Conference, pages 119–126, 1991.

[Europeana] Europeana. *Europeana.* http ://www.europeana.eu/portal/.

[EXI] *EXIF.* http ://www.exif.org/.

[Fadili 10] M.J. Fadili, J.-L. Starck, J. Bobin & Y. Moudden. *Image Decomposition and Separation Using Sparse Representations : An Overview.* Proceedings of the IEEE, Special Issue : Applications of Sparse Representation, vol. 98, no. 6, pages 983–994, 2010.

[Fei-Fei 09] Li Fei-Fei, Rob Fergus & Antonio Torralba. *Recognizing and Learning Object Categories.* In Course of ICCV, 2009.

[Felzenszwalb 04] Pedro F. Felzenszwalb & Daniel P. Huttenlocher. *Efficient Graph-Based Image Segmentation.* Int. J. Comput. Vision, vol. 59, pages 167–181, September 2004.

[Ferman 02] A.M. Ferman, A.M. Tekalp & R. Mehrotra. *Robust color histogram descriptors for video segment retrieval and identification.* IEEE Transactions on Image Processing, vol. 11, no. 5, pages 497–508, 2002.

[for 11] *Illustration de l'algorithme de Fortune.* http ://www.diku.dk/hjemmesider/studerende/duff/Fortune/, 20 mai 2011.

[Fortune 86] S Fortune. *A sweepline algorithm for Voronoi diagrams.* In SCG '86 : Proceedings of the second annual symposium on Computational geometry, pages 313–322, New York, NY, USA, 1986. ACM.

[Fraenkel 69] Aviezri S. Fraenkel, Franz L. Alt & Morris Rubinoff. *Legal Information Retrieval.* volume 9, pages 113–178. Elsevier, 1969.

[Francos 93] J. M. Francos, A. Z. Meiri & B. Porat. *A unified texture model based on a 2-D Wold-like decomposition.* IEEE Trans. Signal Processing, vol. 41, no. 8, pages 2665–2678, 1993.

[Freeman 61] Herbert Freeman. *On the Encoding of Arbitrary Geometric Configurations.* IRE Transactions on Electronic Computers, vol. EC-10, no. 2, pages 260 –268, 1961.

[Friedman-Hill 03] Ernest Friedman-Hill. Jess in action : Java Rule-Based systems. Manning Publications, July 2003.

[Gagnon 07] Michel Gagnon. Logique descriptive et OWL. Cours dispensé à l'école polytechnique de Montréal, Canada, 2007.

[Gallica] Gallica. *Bibliothèque numérique Gallica.* http ://gallica.bnf.fr/.

[Gennari 03] John H. Gennari, Mark A. Musen, Ray W. Fergerson, William E. Grosso, Monica Crubézy, Henrik Eriksson, Natalya F. Noy & Samson W. Tu. *The evolution of Protégé : an environment for knowledge-based systems development.* International Journal of Human-Computer Studies, vol. 58, no. 1, pages 89–123, January 2003.

[Ghorbel 94] Faouzi Ghorbel. *A complete invariant description for gray-level images by the harmonic analysis approach.* Pattern Recognition Letters, vol. 15, no. 10, pages 1043 – 1051, 1994.

[Girard 10] Nathalie Girard, Jean-Marc Ogier & Étienne Baudrier. *A new image quality measure considering perceptual information and local spatial feature*. In Proceedings of the 8th international conference on Graphics recognition : achievements, challenges, and evolution, GREC'09, pages 242–250, Berlin, Heidelberg, 2010. Springer-Verlag.

[Glauberman 56] M. H. Glauberman. *Character recognition for business machines*. Electronics, pages 132–136, 1956.

[Gonzalez 09] Rafael C. Gonzalez, Richard E. Woods & Steven L. Eddins. Digital image processing using MATLAB, 2nd ed. Gatesmark Publishing, 2nd edition, 2009.

[Goodchild 99] Michael F. Goodchild. Interoperating geographic information systems. Springer, February 1999.

[Google a] Google. *Google Images*. http ://images.google.fr/.

[Google b] Google. *Google Livres*. http ://books.google.fr/.

[Graham 98] Norma Graham & Anne Sutter. *Spatial summation in simple (fourier) and complex (non-fourier) texture channels*. Vision Research, vol. 38, no. 2, pages 231–257, 1998.

[Grosicki 09] Emmanuele Grosicki & Haikal El Abed. *ICDAR 2009 Handwriting Recognition Competition*. In 10th International Conference on Document Analysis and Recognition, Barcelona, Spain, 26-29 July 2009, pages 1398–1402. IEEE Computer Society, 2009.

[Gruber 93] Thomas R. Gruber. *Toward Principles for the Design of Ontologies Used for Knowledge Sharing*. In Formal Ontology in Conceptual Analysis and Knowledge Representation, Substantial Revision of Paper Presented at the International Workshop on Formal Ontology. Kluwer Academic Publishers, 1993.

[Gruber 95] Thomas R. Gruber. *Toward principles for the design of ontologies used for knowledge sharing*. Int. J. Hum.-Comput. Stud., vol. 43, pages 907–928, 1995.

[Guillas 07] Stéphanie Guillas. *Reconnaissance d'Objets Graphiques Détériorés : Approche Fondée sur un Treillis de Galois*. PhD thesis, Université de La Rochelle, decembre 2007.

[Guillaumin 10] Matthieu Guillaumin, Jakob J. Verbeek & Cordelia Schmid. *Multiple Instance Metric Learning from Automatically Labeled Bags of Faces*. In 11th European Conference on Computer Vision Heraklion, Crete, Greece, September 5-11, Lecture Notes in Computer Science, pages 634–647, 2010.

[Guo 89] Zicheng Guo & Richard W. Hall. *Parallel thinning with two-subiteration algorithms*. Commun. ACM, vol. 32, pages 359–373, March 1989.

[Haarslev 01] Volker Haarslev & Ralf Möller. *Description of the RACER System and its Applications*. In Description Logics 2001 International Description Logics Workshop. Stanford, CA, USA, August 1-3 2001.

[Haarslev 03] V. Haarslev & R. Möller. *Racer : An OWL Reasoning Agent for the Semantic Web*. In Proceedings of the International Workshop on Applications, Products and Services of Web-based Support Systems, in conjunction with the 2003 IEEE/WIC International Conference on Web Intelligence, Halifax, Canada, October 13, pages 91–95, 2003.

[Haese-Coat 93] V. Haese-Coat, K. KPALMA & J. RONSIN. *Apport comparatif de la texture en segmentation d'image SPOT : application à la forêt de Paimpont (I. & V.)*. Photo interpretation, vol. 31-32, no. 4-1-2, pages 49–64, 1993.

[Hamidi 10] A. El Hamidi, M. MÈnard, M. Lugiez & C. Ghannam. *Weighted and extended total variation for image restoration and decomposition*. Pattern Recognition, vol. 43, no. 4, pages 1564 – 1576, 2010.

[Han 02] Ju Han & Kai-Kuang Ma. *Fuzzy color histogram and its use in color image retrieval*. IEEE Transactions on Image Processing, vol. 11, no. 8, pages 944–952, 2002.

[Haralick 73] R. Haralick. *Textural features for image classification*. IEEE Transactions on Systems, Man, and Cybertinetics, vol. 3, no. 6, pages 610–621, 1973.

[Haralick 92] Robert M. Haralick & Linda G. Shapiro. Computer and robot vision. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1st edition, 1992.

[Harris 88a] C. Harris & M. Stephens. In Proceedings of the 4th Alvey Vision Conference, 1988.

[Harris 88b] C. Harris & M. Stephens. *A Combined Corner and Edge Detection*. In Proceedings of The Fourth Alvey Vision Conference, pages 147–151, 1988.

[Hechenbichler 04] K. Hechenbichler & K. Schliep. *Weighted k-Nearest-Neighbor Techniques and Ordinal Classification*, 2004.

[Heck 99] Rachel M. Heck, Sarah M. Luebke & Chad H. Obermark. *A Survey of Web Annotation Systems.*

http ://www.math.grin.edu/∼rebelsky/Blazers/ Annotations/Summer1999/Papers/survey_ paper.html, 1999.

[Heckbert 82] Paul Heckbert. *Color image quantization for frame buffer display.* SIGGRAPH Comput. Graph., vol. 16, pages 297–307, July 1982.

[Hero 99] III. Hero A.O. & O.J.J. Michel. *Asymptotic theory of greedy approximations to minimal k-point random graphs.* IEEE Transactions on Information Theory, vol. 45, no. 6, pages 1921 –1938, sep 1999.

[Heutte 98] Laurent Heutte, Thierry Paquet, Jean-Vincent Moreau, Yves Lecourtier & Christian Olivier. *A structural/statistical feature based vector for handwritten character recognition.* Pattern Recognition Letters, vol. 19, no. 7, pages 629–641, 1998.

[Hsu 93] Tao Hsu, A. D. Calway & R. Wilson. *Texture Analysis Using the Multiresolution Fourier Transform.* Rapport technique, Bristol, UK, UK, 1993.

[Hu 62] Ming-Kuei Hu. *Visual pattern recognition by moment invariants.* IEEE Transactions on Information Theory, vol. 8, no. 2, pages 179–187, February 1962.

[Hudelot 08] Céline Hudelot, Jamal Atif & Isabelle Bloch. *Fuzzy spatial relation ontology for image interpretation.* Fuzzy Sets Syst., vol. 159, pages 1929–1951, August 2008.

[Iconclass] Iconclass. *Iconclass - Alphanumeric calssification codes.* http ://www.iconclass.nl/texts/usingic.htm.

[Iften 04] Tahar Iften & Abdelmounaine Safia. *Comparaison entre la Matrice de Cooccurrence et la Transformation en Ondelettes pour la Classification Texturale des Images HRV (XS) SPOT.* Teledetection, vol. 4, no. 1, pages 39–49, 2004.

[IMPACT] IMPACT. *Improving Access to Text.* http ://www.impact-project.eu/.

[IPT] *International Press Telecommunications Council.*

[Ishikawa 98] Yoshiharu Ishikawa, Ravishankar Subramanya & Christos Faloutsos. *MindReader : Querying Databases Through Multiple Examples.* In Proceedings of the 24rd International Conference on Very Large Data Bases, VLDB '98, pages 218–227, San Francisco, CA, USA, 1998. Morgan Kaufmann Publishers Inc.

[Itti 98] L. Itti, C. Koch & E. Niebur. *A model of saliency-based visual attention for rapid scene analysis.* IEEE Transactions on Pat-

tern Analysis and Machine Intelligence, vol. 20, no. 11, pages 1254–1259, 1998.

[Jain 91] Anil K. Jain & Farshid Farrokhnia. *Unsupervised texture segmentation using Gabor filters*. Pattern Recognition, vol. 24, pages 1167–1186, December 1991.

[Jau-Ling 02] Shih Jau-Ling & Chen Ling-Hwei. *Color Image Retrieval Based on Primitives of Color Moments*. In Shi-Kuo Chang, Zen Chen & Suh-Yin Lee, editeurs, Recent Advances in Visual Information Systems, volume 2314 of *Lecture Notes in Computer Science*, pages 19–27. Springer Berlin / Heidelberg, 2002. 10.1007/3-540-45925-1_8.

[Jeon 03] J. Jeon, V. Lavrenko & R. Manmatha. *Automatic image annotation and retrieval using cross-media relevance models*. In Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval, SIGIR '03, pages 119–126, New York, NY, USA, 2003. ACM.

[Jimenes 06] Rémi Jimenes. L'imprimerie nationale. Master's thesis, Université de Tours, 2006.

[Jimenes 08] Rémi Jimenes. *Les Bibliothèques Virtuelles Humanistes et l'étude du matériel typographique*. Rapport technique, Centre d'Etude Superieur de la Renaissance, 2008.

[Joint 04] Magali Joint. *PIRIA : a general tool for indexing, search, and retrieval of multimedia content*. In Proceedings of SPIE, pages 116–125, San Jose, CA, USA, 2004.

[Jones 72] Karen Sparck Jones. *A Statistical Interpretatio of Term Specificity and its Application in Retrieval*. Journal of Documentation, vol. 28, no. 1, pages 11–21, 1972.

[Jones 79] K. Sparck Jones. *Experiments in Relevance Weighting of Search Terms*. Information Processing and Management, vol. 15, no. 3, pages 133–44, 1979.

[Jouili 08] Salim Jouili & Salvatore Tabbone. *Applications des graphes en traitement d'images*. In ROGICS'08, pages 434–442, Mahdia Tunisia, 2008. University of Ottawa, Canada and University of Sfax, Tunisia.

[Jouili 09a] Salim Jouili, Ines Mili & Salvatore Tabbone. *Attributed Graph Matching Using Local Descriptions*. In Jacques Blanc-Talon, Wilfried Philips, Dan C. Popescu & Paul Scheunders, editeurs,

Advanced Concepts for Intelligent Vision Systems, 11th International Conference, volume 5807 of *Lecture Notes in Computer Science*, pages 89–99. Springer, 2009.

[Jouili 09b] Salim Jouili & Salvatore Tabbone. *Graph Matching Based on Node Signatures*. In Andrea Torsello, Francisco Escolano & Luc Brun, editeurs, Graph-Based Representations in Pattern Recognition, volume 5534 of *Lecture Notes in Computer Science*, pages 154–163. Springer Berlin / Heidelberg, 2009.

[Jouili 10] Salim Jouili, Mickaël Coustaty, Salvatore Tabbone & Jean-Marc Ogier. *NAVIDOMASS : Structural-based Approaches Towards Handling Historical Documents*. In 20th International Conference on Pattern Recognition, ICPR 2010, Istanbul, Turkey, 23-26 August 2010, pages 946–949, 2010.

[Jouili 11] Salim Jouili. *Indexation de masses de documents graphiques : approches structurelles*. PhD thesis, Université Nancy 2, Mars 2011.

[Journet 06] Nicholas Journet. *Analyse d'images de documents anciens : une approche texture*. PhD thesis, Université de La Rochelle, decembre 2006.

[Journet 08a] N. Journet, J.-Y. Ramel & C. Cureau. *Vers une plate-forme d'indexation interactive de documents*. In Dixième Colloque International Francophone sur l'Ecrit et le Document Colloque International Francophone sur l'Ecrit et le Document, pages 197–198, France, 2008. Groupe de Recherche en Communication Ecrite.

[Journet 08b] Nicholas Journet, Rémy Mullot, Veronique Eglin & Jean-Yves Ramel. *Analyse d'Images de Documents Anciens : une Approche Texture*. Traitement du Signal, vol. 24, no. 6, pages 461–479, 09 2008.

[Journet 08c] Nicholas Journet, Jean-Yves Ramel, Rémy Mullot & Véronique Eglin. *Document image characterization using a multiresolution analysis of the texture : application to old documents*. IJDAR, vol. 11, no. 1, pages 9–18, 2008.

[Kalyanpur 05] Aditya Kalyanpur, Bijan Parsia, Evren Sirin, Bernardo Cuenca Grau & James Hendler. *Swoop : A Web Ontology Editing Browser*. journal of Web Semantics, vol. 4, page 2005, 2005.

[Kammersell 07] William Kammersell & Mike Dean. *Conceptual Search : Incorporating Geospatial Data into Semantic Queries*. In Arno

Scharl & Klaus Tochtermann, editeurs, The Geospatial Web, Advanced Information and Knowledge Processing, pages 47–54. Springer London, 2007.

[Karaf Falah 93] R. Karaf Falah & Ph. BOLON. *Mesure de dissimilarité entre deux segmentations*. In GRETSI - Actes de Colloques, 1993.

[Karathanassi 00] V. Karathanassi, C. Iossifidis & D. Rokos. *A texture-based classification method for classifying built areas according to their density*. In International Journal of Remote Sensing, Volume 21, Number 9, pages 1807–1823. Taylor and Francis Ltd, 2000.

[Kasturi 90] Rangachar Kasturi & Mohan M. Trivedi. Image analysis applications. M. Dekker, May 1990.

[Kauniskangas 99] Hannu Kauniskangas. *Document Image Retrieval with Improvements in Database Quality*. PhD thesis, Oulu, 1999.

[Kermad 95] Chafik Kermad, Kacem Chehdi & Claude Cariou. *Segmentation d'images par multi-seuillage et fusion de régions labellisées minimisant un critère de similarité*. In GRETSI, pages 641–644, 1995.

[Kim 99] Kim Y.-S. Kim W.-Y. *A new region-based shape descriptor*. In Mpeg Meeting, TR 15-01, 1999.

[Kim 05] Deok-Hwan Kim, Chin-Wan Chung & Kobus Barnard. *Relevance feedback using adaptive clustering for image similarity retrieval*. Journal of Systems and Software, vol. 78, no. 1, pages 9 – 23, 2005.

[Kimura 91] F. Kimura & M. Shridhar. *Handwritten numerical recognition based on multiple algorithms*. Pattern Recognition, vol. 24, no. 10, pages 969–983, 1991.

[Kisku 10] Dakshina Ranjan Kisku, Ajita Rattani, Enrico Grosso & Massimo Tistarelli. *Face Identification by SIFT-based Complete Graph Topology*. Computing Research Repository, 2010.

[Koldovsky 06] Z. Koldovsky & P. Tichavsky. *Methods of Fair Comparison of Performance of Linear ICA Techniques in Presence of Additive Noise*. In Acoustics, Speech and Signal Processing, IEEE International Conference on, volume 5, page V, 2006.

[Komodakis 11] N. Komodakis, N. Paragios & G. Tziritas. *MRF Energy Minimization and Beyond via Dual Decomposition*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, no. 3, pages 531 –552, march 2011.

[Kompatsiaris 08] Paola Kompatsiaris Yiannis ; Hobson, editeur. Semantic multimedia and ontologies, volume 1 of *Theory and Applications*. Springer, 2008.

[Kothuri 07] Ravikanth V. Kothuri, Albert Godfrind & Euro Beinat. Pro oracle spatial for oracle database 11g (expert's voice in oracle). Berkeley, CA, USA, 2007.

[Kuhn 10] Harold W. Kuhn. *The Hungarian Method for the Assignment Problem*. In Michael Jünger, Thomas M. Liebling, Denis Naddef, George L. Nemhauser, William R. Pulleyblank, Gerhard Reinelt, Giovanni Rinaldi & Laurence A. Wolsey, editeurs, 50 Years of Integer Programming 1958-2008, pages 29–47. Springer Berlin Heidelberg, 2010.

[Lacot 05] Xavier Lacot. *Introduction à OWL, un langage XML d'ontologies Web*, 2005.

[Lavrenko 03] Victor Lavrenko, R. Manmatha & Jiwoon Jeon. *A Model for Learning the Semantics of Pictures.* In NIPS'03, 2003.

[Lazebnik 06] S. Lazebnik, C. Schmid & J. Ponce. *Beyond Bags of Features : Spatial Pyramid Matching for Recognizing Natural Scene Categories*. In Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on, volume 2, pages 2169 – 2178, 2006.

[Li 08] Jia Li, Senior Member, James Z Wang & Senior Member. *Real-Time Computerized Annotation of Pictures*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no. 6, pages 985–1002, 2008.

[Liao 94] Q. Liao. *Détection de contours et segmentation d'images. Applications à la télédétection et à la biologie marine = Edge Detection and Image Segmentation. Applications in Remote Sensing and Marine Biology*. PhD thesis, Universite de Rennes 1, 1994.

[Lin 97] Hsin-Chih Lin, Ling-Ling Wang & Shi-Nine Yang. *Extracting periodicity of a regular texture based on autocorrelation functions*. Pattern Recognition Letters, vol. 18, no. 5, pages 433 – 443, 1997.

[Lindeberg 94] Tony Lindeberg. Scale-space theory in computer vision. Kluwer Academic Publishers, Norwell, MA, USA, 1994.

[Lindeberg 98] Tony Lindeberg. *Feature Detection with Automatic Scale Selection*. Int. J. Comput. Vision, vol. 30, no. 2, pages 79–116, 1998.

[Liu 07] Ying Liu, Dengsheng Zhang, Guojun Lu & Wei-Ying Ma. *A survey of content-based image retrieval with high-level semantics.* Pattern Recogn., vol. 40, pages 262–282, 2007.

[Livingstone 88] M. Livingstone & D. Hubel. *Segregation of form, color, movement, and depth : anatomy, physiology, and perception.* Science, vol. 240, no. 4853, pages 740–749, 1988.

[Lloyd 82] Stuart P. Lloyd. *Least squares quantization in PCM.* IEEE Transactions on Information Theory, vol. 28, pages 129–136, 1982.

[Lowe 99] David G. Lowe. *Object recognition from local scale-invariant features.* In International Conference on Computer Vision, pages 1150–1157, Corfu, Greece, September 1999.

[L.Rudin 92] L.Rudin, S.Osher & E.Fatemi. *Nonlinear total variation based noise removal.* Physica D, vol. 60, pages 259–269, 1992.

[Mag] *Base de données d'ornements typographiques utilisés par les imprimeurs de l'ancien régime.* http ://maguelone.enssib.fr/.

[Malki 02] Jamal Malki, Ei-Hadi Zahzah & Laurent Mascarilla. *Indexation et recherche d'image fondée sur les relations spatiales entre objets.* Traitement du signal, vol. 19, 2002.

[Mallat 99] Stephane Mallat. A wavelet tour of signal processing, second edition. Academic Press, 2 edition, September 1999.

[Marr 80] D. Marr & E. Hildreth. *Theory of Edge Detection.* Proceedings of the Royal Society of London. Series B, Biological Sciences, vol. 207, no. 1167, pages 187–217, February 1980.

[Mas 06] Joan Mas, Gemma Sanchez & Josep Lladós. *An Incremental Parser to Recognize Diagram Symbols and Gestures Represented by Adjacency Grammars.* In Liu Wenyin & Josep Lladós, editeurs, Selected papers from GREC'05, volume 3926, pages 243–254. LNCS, 2006.

[Mas 10] Joan Mas, Josep Lladós, @ Sánchez & Joaquim Armando Pires Jorge. *A syntactic approach based on distortion-tolerant Adjacency Grammars and a spatial-directed parser to interpret sketched diagrams.* Pattern Recognition, vol. 43, no. 12, pages 4148–4164, 2010.

[Matas 04] J. Matas, O. Chum, M. Urban & T. Pajdla. *Robust wide-baseline stereo from maximally stable extremal regions.* Image and Vision Computing, vol. 22, no. 10, pages 761 – 767, 2004. British Machine Vision Computing 2002.

[Matsakis 99] Pascal Matsakis & Laurent Wendling. *A New Way to Represent the Relative Position between Areal Objects*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 21, page 634–643, July 1999. ACM ID : 310626.

[McAndrew 04] A. McAndrew. An introduction to digital image processing with matlab. Thomson/Course Technology, 2004.

[McGuinness 04] Deborah L. McGuinness & Frank van Harmelen. *OWL Web Ontology Language*. Rapport technique, W3C, February 2004.

[McQueen 67] J. B. McQueen. *Some Methods for classification and Analysis of Multivariate Observations*. In Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability, 1967.

[Mehnert 97] Andrew Mehnert & Paul Jackway. *An improved seeded region growing algorithm*. Pattern Recognition Letters, vol. 18, no. 10, pages 1065 – 1071, 1997.

[Meyer 94] Fernand Meyer. *Topographic distance and watershed lines*. Signal Process., vol. 38, pages 113–125, 1994.

[Meyer 01] Y. Meyer. Oscillating patterns in image processing and nonlinear evolution equations. The fifteenth dean jacqueline B. Lewis Memorial Lectures, 2001.

[Mohand 10] Kamel Ait Mohand, Thierry Paquet, Nicolas Ragot & Laurent Heutte. *Structure Adaptation of HMM Applied to OCR*. In Proceedings of the 2010 20th International Conference on Pattern Recognition, ICPR '10, pages 2877–2880, Washington, DC, USA, 2010. IEEE Computer Society.

[Mokhtarian 92] Farzin Mokhtarian & Alan K Mackworth. *A Theory of Multiscale, Curvature-Based Shape Representation for Planar Curves*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 14, page 789–805, August 1992.

[Mokhtarian 96] Farzin Mokhtarian, Sadegh Abbasi & Josef Kittler. *Efficient and Robust Retrieval by Shape Content through Curvature Scale Space*. 1996.

[Monay 03] Florent Monay & Daniel Gatica-Perez. *On image auto-annotation with latent space models*. In Proceedings of the eleventh ACM international conference on Multimedia, MULTIMEDIA '03, pages 275–278, New York, NY, USA, 2003. ACM.

[Montanari 68] U. Montanari. *A Method for Obtaining Skeletons Using a Quasi-Euclidean Distance*. Journal of the ACM (JACM), vol. 15, page 600–624, October 1968.

[Montanari 69] Ugo Montanari. *Continuous Skeletons from Digitized Images.* Journal of the ACM (JACM), vol. 16, page 534–549, October 1969.

[Moravec 79] Hans P. Moravec. *Visual mapping by a robot rover.* In IJCAI'79 : Proceedings of the 6th international joint conference on Artificial intelligence, pages 598–600, San Francisco, CA, USA, 1979. Morgan Kaufmann Publishers Inc.

[Morin 96] Edgar Morin. *Pour une réforme de la pensée.* www, vol. 49, no. 2, pages 10–14, 1996.

[Motik 05] Boris Motik, Ulrike Sattler & Rudi Studer. *Query Answering for OWL-DL with rules.* Web Semantics, vol. 3, pages 41–60, July 2005.

[Naegel 09] Benoît Naegel & Laurent Wendling. *Combining shape descriptors and component-tree for recognition of ancient graphical drop caps.* In VISAPP'09 : Fourth International Conference on Computer Vision Theory and Applications, pages 297–302, Lisboa Portugal, 2009.

[Nakazato 03] Munehiro Nakazato, Charlie Dagli & Thomas S. Huang. *Evaluating groupbased relevance feedback for content-based image retrieval.* In Proc. IEEE Int. Conf. on Image Processing, pages 599–602, 2003.

[NaviDoMass 10] NaviDoMass. *http ://navidomass.univ-lr.fr/*, 2007-2010. Projet ANR NAVIDOMASS - NAVIgation into DOcuments MASSes - de sauvegarde et d'indexation du patrimoine historique francais.

[Neumann 08] Bernd Neumann & Ralf Möller. *On scene interpretation with description logics.* Image Vision Comput., vol. 26, pages 82–101, January 2008.

[Newell 82] Allen Newell. *The Knowledge Level.* Artif. Intell., vol. 18, no. 1, pages 87–127, 1982.

[Nguyen 09] Nhu Van Nguyen, Alain Boucher, Jean-Marc Ogier & Salvatore Tabbone. *Region-Based Semi-automatic Annotation Using the Bag of Words Representation of the Keywords.* In Proceedings of the 2009 Fifth International Conference on Image and Graphics, ICIG '09, pages 422–427, Washington, DC, USA, 2009. IEEE Computer Society.

[Nguyen 11] Nhu Van Nguyen. *Représentations visuelles de concepts textuels pour la recherche et l'annotation interactives d'images.* PhD thesis, Université de La Rochelle, 2011.

[Niebles 08] Juan Niebles, Hongcheng Wang & Li Fei-Fei. *Unsupervised Learning of Human Action Categories Using Spatial-Temporal Words*. International Journal of Computer Vision, vol. 79, pages 299–318, 2008.

[Nies 96] Michael Lew Nies & Michael S. Lew. *Content Based Image Retrieval : KLT, Projections, or Templates*. In Image Databases and Multi-Media Search, pages 27–34. Amsterdam University Press, 1996.

[OCA] OCA. *Open Content Alliance (OCA)*. http ://www.opencontentalliance.org/.

[Ogier 00] Jean-Marc Ogier. *De l'image au document technique, Problèmes d'interpretation*. PhD thesis, Univeristé de Rouen, 2000. Habilitation a diriger des recherches.

[Ogier 06] Jean-Marc Ogier & Karl Tombre. *Madonne : Document Image Analysis Techniques for Cultural Heritage Documents*. In International Conference on Digital Cultural Heritage, Vienna, Austria, 08 2006.

[Ogle 95] Virginia E. Ogle & Michael Stonebraker. *Chabot : Retrieval from a Relational Database of Images*. Computer, vol. 28, no. 9, pages 40–48, 1995.

[Ortega-Binderberger 04] Michael Ortega-Binderberger & Sharad Mehrotra. *Relevance feedback techniques in the MARS image retrieval system*. Multimedia Systems, vol. 9, pages 535–547, 2004.

[Otsu 79] Nobuyuki Otsu. *A Threshold Selection Method from Gray-Level Histograms*. IEEE Transactions on Systems, Man and Cybernetics, vol. 9, no. 1, pages 62–66, 1979.

[OWL 11] *OWL Web Ontology Language*. `http://www.w3.org/TR/owl-features/`, may, 6 2011.

[Pal 10] Shyamosree Pal, Partha Bhowmick, Arindam Biswas & Bhargab Bhattacharya. *GOAL : Towards Understanding of Graphic Objects from Architectural to Line Drawings*. In Jean-Marc Ogier, Wenyin Liu & Josep Lladós, editeurs, Graphics Recognition. Achievements, Challenges, and Evolution, volume 6020 of *Lecture Notes in Computer Science*, pages 81–92. Springer Berlin / Heidelberg, 2010.

[Pareti 06a] Rudolf Pareti & Nicole Vincent. *Ancient Initial Letters Indexing*. In ICPR '06, pages 756–759, Washington, DC, USA, 2006. IEEE Computer Society.

[Pareti 06b] Rudolf Pareti & Nicole Vincent. *Ancient Initial Letters Indexing*. In ICPR '06 : Proceedings of the 18th International Conference on Pattern Recognition, pages 756–759, 2006.

[Pareti 06c] Rudolf Pareti, Nicole Vincent, Surapong Uttama, Jean-Marc Ogier, Jean-Pierre Salmon, Salvatore Tabbone, Laurent Wendling & Sebastien Adam. *On defining signatures for the retrieval and the classification of graphical drop caps*. DIAL, vol. 1, pages 220–231, 2006.

[Pareti 08] Rudolf Pareti. *Indexation d'images par une loi puissance*. PhD thesis, Université de Paris 5 René Descartes, 2008.

[Park 00] Dong Kwon Park, Yoon Seok Jeon & Chee Sun Won. *Efficient use of local edge histogram descriptor*. In Proceedings of the 2000 ACM workshops on Multimedia, MULTIMEDIA '00, pages 51–54, New York, NY, USA, 2000. ACM.

[pas 96] *Recherche d'ornement Passe-Partout*, 1996. http ://www2.unil.ch/BCUTodai/app/Todai.do.

[Pfaltz 67] John L Pfaltz & Azriel Rosenfeld. *Computer representation of planar regions by their skeletons*. Communications of the ACM, vol. 10, page 119–122, February 1967.

[Pham 07] M. Pham, R. Susomboon, T. Disney, D. Raicu & J. Furst. *A Comparison of Texture Models for Automatic Liver Segmentation*. SPIE Medical Imaging Conf., January 2007.

[Picard 08] D. Picard, M. Cord & A. Revel. *Image Retrieval Over Networks : Active Learning Using Ant Algorithm*. IEEE Transactions on Multimedia, vol. 10, no. 7, pages 1356 –1365, nov. 2008.

[Picard 10] David Picard, Arnaud Revel & Matthieu Cord. *An application of swarm intelligence to distributed image retrieval*. Information Sciences, vol. In Press, Corrected Proof, 2010.

[Pratt 07] W. Pratt. Digital image processing. Wiley, 2007.

[Prum 10] Sophea Prum, Muriel Visani & Jean-Marc Ogier. *Cursive On-line Handwriting Word Recognition Using a Bi-character Model for Large Lexicon Applications*. In International Conference on Frontiers in Handwriting Recognition, Kolkata, India, pages 194–199, 16-18 November 2010.

[Purves 05] Dale Purves. Neurosciences. De Boeck Université, 2005.

[Puzicha 00] J. Puzicha, M. Held, J. Ketterer, J.M. Buhmann & D.W. Fellner. *On spatial quantization of color images*. IEEE Transactions on Image Processing, vol. 9, no. 4, pages 666 –682, 2000.

[Ramel 06] J.Y. Ramel, S. Busson & M.L. Demonet. *AGORA : the Interactive Document Image Analysis Tool of the BVH Project.* Document Image Analysis for Libraries, International Workshop on, pages 145–155, 2006.

[Randell 92] David A. Randell, Zhan Cui & Anthony G. Cohn. *A Spatial Logic based on Regions and Connection.* In Proceedinfs 3rd International Conference On Knowledge Representation and Reasoning, 1992.

[Raveaux 10] Romain Raveaux, Jean-Christophe Burie & Jean-Marc Ogier. *A graph matching method and a graph matching distance based on subgraph assignments.* Pattern Recognition Letters, vol. 31, no. 5, pages 394 – 406, 2010.

[Raveaux 11] Romain Raveaux, Sebastien Adam, Pierre Heroux & Eric Trupin. *Learning graph prototypes for shape recognition.* Computer Vision and Image Understanding, vol. In Press, Corrected Proof, pages –, 2011.

[Remazeilles 01] Céline Remazeilles. *Etude des processus de degradation des manuscrits anciens ecrit a l'encre ferrogallique.* PhD thesis, La Rochelle, 2001.

[Renz 02] Jochen Renz, editeur. Qualitative spatial reasoning with topological information, volume 2293. Springer Berlin Heidelberg, Berlin, Heidelberg, 2002.

[René Cori 03] Daniel Lascar René Cori. Logique mathématique, tome 1 - calcul propositionnel, algèbre de Boole, calcul des prédicats. Collection : Sciences Sup. Dunod, 2003.

[Retornaz 07] Thomas Retornaz & Beatriz Marcotegui. *Scene text localization based on the ultimate opening.* In Gerald Jean Francis Banon, Junior Barrera, Ulisses de Mendonça Braga-Neto & Nina Sumiko Tomita Hirata, editeurs, International Symposium on Mathematical Morphology, pages 177–188, São José dos Campos, 2007. Instituto Nacional de Pesquisas Espaciais (INPE).

[Ricard 05] Julien Ricard, David Coeurjolly & Atilla Baskurt. *Generalizations of angular radial transform for 2D and 3D shape retrieval.* Pattern Recognition Letters, vol. 26, no. 14, pages 2174–2186, October 2005.

[Riesen 09] Kaspar Riesen. *Classification and Clustering of Vector Space Embedded Graphs.* PhD thesis, 2009.

[Rosenberger 99] Christophe Rosenberger. *Mise en oeuvre d'un système adaptatif de segmentation d'images*, December 1999.

[Rosenfeld 66] Azriel Rosenfeld & John L. Pfaltz. *Sequential Operations in Digital Picture Processing.* J. ACM, vol. 13, pages 471–494, October 1966.

[Rosten 10] Edward Rosten, Reid Porter & Tom Drummond. *FASTER and better : A machine learning approach to corner detection.* IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 32, pages 105–119, 2010.

[Rui 99] Yong Rui, Thomas S. Huang & Shih-Fu Chang. *Image Retrieval : Current Techniques, Promising Directions, and Open Issues.* Journal of Visual Communication and Image Representation, vol. 10, no. 1, pages 39 – 62, 1999.

[Rusiñol 06] Marçal Rusiñol & Josep Lladós. *Symbol Spotting in Technical Drawings Using Vectorial Signatures.* In Liu Wenyin & Josep Lladós, editeurs, Selected papers from GREC'05, volume 3926, pages 35–45. LNCS, 2006.

[Rusinol 07] Marcal Rusinol & Josep Llados. *A Region-Based Hashing Approach for Symbol Spotting in Technical Documents.* In Seventh IAPR International Workshop on Graphics Recognition, pages 41–42, septembre 2007.

[Rusiñol 08] M. Rusiñol & J. Lladós. *A Region-Based Hashing Approach for Symbol Spotting in Technical Documents.* In Graphics Recognition. Recent Advances and New Oportunities, volume 5046 of *Lecture Notes on Computer Science*, pages 321–328. 2008.

[Rusiñol 09a] M. Rusiñol. *Geometric and Structural-based Symbol Spotting. Application to Focused Retrieval in Graphic Document Collections.* PhD thesis, Universitat Autonoma de Barcelona, 2009.

[Rusiñol 09b] M. Rusiñol & J. Lladós. *A Performance Evaluation Protocol for Symbol Spotting Systems in Terms of Recognition and Location Indices.* International Journal on Document Analysis and Recognition, vol. 12, no. 2, pages 83–96, 2009.

[Rusiñol 10] Marçal Rusiñol & Josep Lladós. *Efficient logo retrieval through hashing shape context descriptors.* In Ninth IAPR Workshop on Document Analysis Systems, pages 215–222. ACM, 2010.

[Sergey 06] Brin Sergey & Page Lawrence. *The Anatomy of a Large-Scale Hypertextual Web Search Engine.* 2006.

[Shechtman 07] Eli Shechtman & Michal Irani. *Space-time behavior-based correlation-Or-how to tell if two underlying motion fields are similar without computing them ?* IEEE transactions on pattern analysis and machine intelligence, vol. 29, no. 11, pages 2045–56, November 2007.

[Shen 99] D.G. Shen & H.H.S. Ip. *Discriminative wavelet shape descriptors for recognition of 2-D patterns*. Pattern Recognition, vol. 32, no. 2, pages 151–165, February 1999.

[Shi 94] Jianbo Shi & Carlo Tomasi. *Good Features to Track*. In 1994 IEEE Conference on Computer Vision and Pattern Recognition (CVPR'94), pages 593 – 600, 1994.

[Shridhar 84] M. Shridhar & A. Badreldin. *High accuracy character recognition algorithm using fourier and topological descriptors*. Pattern Recognition, vol. 17, no. 5, pages 515–524, 1984.

[Sidere 09] Nicolas Sidere, Pierre Héroux & Jean-Yves Ramel. *Vector Representation of Graphs : Application to the Classification of Symbols and Letters*. In 10th International Conference on Document Analysis and Recognition, pages 681–685. IEEE Computer Society, 2009.

[Silva 10] Matthieu Perreira Da Silva, Vincent Courboulay, Armelle Prigent & Pascal Estraillier. *Evaluation of preys / predators systems for visual attention simulation*, May 2010.

[Sirin 07] Evren Sirin, Bijan Parsia, Bernardo Cuenca Grau, Aditya Kalyanpur & Yarden Katz. *Pellet : A practical OWL-DL reasoner*. Web Semantics : Science, Services and Agents on the World Wide Web, vol. 5, no. 2, pages 51 – 53, 2007. Software Engineering and the Semantic Web.

[Sivic 09] J. Sivic & A. Zisserman. *Efficient Visual Search of Videos Cast as Text Retrieval*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 4, pages 591 –606, april 2009.

[Smeulders 00] Arnold W. M. Smeulders, Marcel Worring, Simone Santini, Amarnath Gupta & Ramesh Jain. *Content-Based Image Retrieval at the End of the Early Years*. IEEE Trans. Pattern Anal. Mach. Intell., vol. 22, pages 1349–1380, 2000.

[Smith 97] Stephen M. Smith & J. Michael Brady. *SUSAN—A New Approach to Low Level Image Processing*. International Journal of Computer Vision, vol. 23, pages 45–78, 1997.

[Soegaard 10] Mads Soegaard. *Gestalt principles of form perception*, 2010. This is an electronic document. Date of publication : March 22, 2010. Date retrieved : April 26, 2011. Date last modified : March 22, 2010.

[SPA 11] *Langage d'interrogation SPARQL pour RDF*. `http://www.yoyodesign.org/doc/w3c/rdf-sparql-query/`, may, 6 2011.

[Starck 04] Jean-Luc Starck, Michael Elad & David Donoho. *Redundant Multiscale Transforms and Their Application for Morphological Component Separation*. volume 132 of *Advances in Imaging and Electron Physics*, pages 287 – 348. Elsevier, 2004.

[Starck 05] J. L. Starck, M. Elad & D. L. Donoho. *Image decomposition via the combination of sparse representations and a variational approach*. IEEE Trans. Image Processing, vol. 14, no. 10, pages 1570–1582, October 2005.

[Stocker] Markus Stocker & Evren Sirin. *PelletSpatial : A Hybrid RCC-8 and RDF/OWL Reasoning and Query Engine*. In International Semantic Web Conference.

[Stricker 95] Markus A Stricker & Markus Orengo. *Similarity of color images*. http ://adsabs.harvard.edu/abs/1995SPIE.2420..381S, March 1995.

[Surapong 08] Uttama Surapong. *Indexation d'images graphiques : Application a la numerisation et a la valorisation du patrimoine*. PhD thesis, Université de La Rochelle, France, 2008.

[Swain 90] M.J. Swain & D.H. Ballard. *Indexing via color histograms*. In Computer Vision, 1990. Proceedings, Third International Conference on, pages 390–393, 1990.

[Tabbone 03] Salvatore Tabbone & Laurent Wendling. *Binary Shape Normalization Using the Radon Transform*. In Ingela Nyström, Gabriella Sanniti di Baja & Stina Svensson, editeurs, Discrete Geometry for Computer Imagery, volume 2886 of *Lecture Notes in Computer Science*, pages 184–193. Springer Berlin / Heidelberg, 2003.

[Tabbone 06] S. Tabbone, L. Wendling & J.-P. Salmon. *A new shape descriptor defined on the radon transform*. Comput. Vis. Image Underst., vol. 102, pages 42–51, April 2006.

[Takahashi 91] H. Takahashi. *A Neural Net OCR using geometrical and zonal pattern features*. In Proc. 1th. Conf. Document Analysis and Recognition, pages pp. 821–828, 1991.

[Tanaka 04] Hideo Tanaka, Yasuo Yoshida, Kimihiko Fukami & Hiroki Nakano. *Texture segmentation using amplitude and phase information of Gabor filters*. Electronics and Communications in Japan (Part III : Fundamental Electronic Science), vol. 87, no. 4, pages 66–79, 2004.

[Tangelder 04] Johan W.H. Tangelder & Remco C. Veltkamp. *A Survey of Content Based 3D Shape Retrieval Methods*. Shape Modeling

and Applications, International Conference on, pages 145–156, 2004.

[Tao 06] Dacheng Tao, Xiaoou Tang, Xuelong Li & Xindong Wu. *Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 7, pages 1088 –1099, july 2006.

[Taxt 90] Torfinn Taxt, Jórunn B. Ólafsdóttir & Morten Dæhlen. *Recognition of handwritten symbols.* Pattern Recognition, vol. 23, no. 11, pages 1155–1166, 1990.

[Teague 80] Michael Reed Teague. *Image analysis via the general theory of moments.* Journal of the Optical Society of America, vol. 70, no. 8, page 920, 1980.

[Teuner 95] A. Teuner, O. Pichler & B.J. Hosticka. *Unsupervised texture segmentation of images using tuned matched Gabor filters.* IEEE Transactions on Image Processing, vol. 4, no. 6, pages 863 –870, jun 1995.

[Theodoridis 08] Sergios Theodoridis & Konstantinos Koutroumbas. Pattern recognition, fourth edition. Academic Press, 4th edition, 2008.

[Torralba 08] A. Torralba, R. Fergus & W.T. Freeman. *80 Million Tiny Images : A Large Data Set for Nonparametric Object and Scene Recognition.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no. 11, pages 1958–1970, 2008.

[Treuenfels 94] Anton Treuenfels. *An efficient flood visit algorithm.* C/C++ Users J., vol. 12, pages 39–62, August 1994.

[Tsarkov 06] Dmitry Tsarkov & Ian Horrocks. *FaCT++ Description Logic Reasoner : System Description.* In Ulrich Furbach & Natarajan Shankar, editeurs, Automated Reasoning, volume 4130 of *Lecture Notes in Computer Science*, pages 292–297. Springer Berlin / Heidelberg, 2006.

[Tuceryan 90] M. Tuceryan & A.K. Jain. *Texture Segmentation Using Voronoi Polygons.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 12, pages 211–216, 1990.

[Uschold 96] M. Uschold & M. King. *Ontologies : principles, methods and applications.* Knowledge Engineering Review, vol. 11, no. 2, pages 93–155, 1996.

[Uttama 05] Surapong Uttama, Pierre Loonis, Mathieu Delalandre & Jean-Marc Ogier. *Segmentation and Retrieval of Ancient Graphic Documents.* In GREC, pages 88–98, 2005.

[Uttama 06] Surapong Uttama, Pierre Loonis, Mathieu Delalandre & Jean-Marc Ogier. *Segmentation and Retrieval of Ancient Graphic Documents*. In Graphics Recognition. Ten Years Review and Future Perspectives, pages 88–98. Springer, 2006.

[Vese 04] L. A. Vese & S. J. Osher. *Image Denoising and Decomposition with Total Variation Minimization and Oscillatory Functions*. Journal of Mathematical Imaging and Vision, vol. 20, no. 1-2, pages 7–18, January 2004.

[Wang 11] Xiang-Yang Wang, Yong-Jian Yu & Hong-Ying Yang. *An effective image retrieval scheme using color, texture and shape features*. Computer Standards & Interfaces, vol. 33, no. 1, pages 59–68, January 2011.

[Wendling 02] L. Wendling, S. Tabbone & P. Matsakis. *Fast and robust recognition of orbit and sinus drawings using histograms of forces*. Pattern Recognition Letters, vol. 23, no. 14, pages 1687–1694, 2002.

[Wong 82] K. Y. Wong, R. G. Casey & F. M. Wahl. *Document Analysis System*. ibm journal of research and development, 1982.

[Yang 07] Hong-Taek Yang & Doowon Paik. *A Discriminative Color Quantization Depending on the Degree of Focus*. In Julie Jacko, editeur, Human-Computer Interaction. Interaction Platforms and Techniques, volume 4551 of *Lecture Notes in Computer Science*, pages 191–196. Springer Berlin / Heidelberg, 2007.

[Yang 08] Mingqiang Yang, Kidiyo Kpalma & Joseph Ronsin. *A Survey of Shape Feature Extraction Techniques*. In Peng-Yeng Yin, editeur, Pattern Recognition, pages 43–90. 2008.

[Yavlinsky 07] Alexei Yavlinsky. *Image indexing and retrieval using automated annotation*. PhD thesis, University of London and the Diploma of Imperial College, 2007.

[Yue 11] Jun Yue, Zhenbo Li, Lu Liu & Zetian Fu. *Content-based image retrieval using color and texture fused features*. Mathematical and Computer Modelling, vol. 54, no. 3-4, pages 1121 – 1127, 2011.

[Zernike 38] F. Zernike. *The concept of degree of coherence and its application to optical problems*. Physica, vol. 5, no. 8, pages 785–795, August 1938.

[Zhang 84] T. Y. Zhang & C. Y. Suen. *A fast parallel algorithm for thinning digital patterns*. Commun. ACM, vol. 27, pages 236–239, March 1984.

[Zhang 02] Dengsheng Zhang & Guojun Lu. *Shape-based image retrieval using generic Fourier descriptor*. Signal Processing : Image Communication, vol. 17, no. 10, pages 825 – 848, 2002.

[Zipf 49] G.K. Zipf. Human behavior and the principle of least effort. Hafner Pub. Co, 1949.

[ÿivind Due Trier 96] ÿivind Due Trier, Anil K. Jain & Torfinn Taxt. *Feature extraction methods for character recognition-A survey*. Pattern Recognition, vol. 29, no. 4, pages 641 – 662, 1996.

Contribution à l'analyse complexe de documents anciens Application aux lettrines

De nombreux projets de numérisation sont actuellement menés en France et en Europe pour sauvegarder le contenu de dizaines de milliers de documents anciens. Les images de ces documents sont utilisées par les historiens pour identifier l'historique des livres. Cette thèse s'inscrit dans le projet Navidomass (ANR-06-MDCA-012) qui a pour but de valoriser le patrimoine écrit français de la Renaissance, en proposant d'identifier ses images pour les indexer. Dans le cadre de cette thèse, nous nous sommes particulièrement intéressés aux images graphiques. Ces documents, qui sont apparus avec le début de l'imprimerie, sont composées d'images complexes puisque composées de différentes couches d'informations (images de traits).
Afin de répondre à ce problème, nous proposons un modèle ontologique d'analyse complexe d'images de documents anciens. Ce modèle permet d'intégrer dans une même base les connaissances propres aux historiens, et les connaissances extraites par des traitements d'images. De par la nature complexe de ces images, les méthodes habituelles d'analyse d'images et d'extraction automatique de connaissances sont inopérantes. Nous proposons donc une nouvelle approche d'analyse des images de documents anciens qui permet de les caractériser à partir de leurs spécificités. Cette approche commence par simplifier les images en les séparant en différentes couches d'informations (formes et traits). Puis, pour chaque couche, nous venons extraire des motifs utilisés pour décrire les images. Ainsi, les images sont caractérisées à l'aide de sacs de motifs fréquents, et de sacs de traits. Pour ces deux couches d'informations, nous venons également extraire des graphes de régions qui permettent d'extraire une connaissance structurelle des images. La complexification de ces deux descriptions est insérée dans la base de connaissances, pour permettre des requêtes complexes. Le but de cette base est de proposer à l'utilisateur de rechercher une image en indiquant soit un exemple d'images recherchées, soit des éléments caractéristiques des images.
Mots clés : Documents anciens, Analyse complexe d'images, Indexation, Ontologies

Complex analysis of historical documents Application to lettrines

In the general context of cultural heritage preservation campaigns, many digitization projects are being conducted in France and Europe to save the contents of thousands of ancient documents. Images of these documents are used by historians to identify the history of books. This thesis was led into the Navidomass project (ANR-06-MDCA-012) which aims at promoting the written heritage of the documents from the Renaissance, by proposing to identify its images. As part of this thesis, we are particularly interested in graphical images, and more specifically to dropcaps. These graphical images, which emerged with the beginning of printing, are complex images which can be seen as composed of different layers of information (images composed of strokes).

To address this problem, we propose an ontological model of complex analysis of images of old documents. This model allows to integrate the knowledge specific to historians, and the knowledge extracted by image processing, into a single database. Due to the complex nature of these images, the usual methods of image analysis and automatic extraction of knowledge are inefficient. We therefore propose a new approach for analyzing images of old documents that can be characterized on their features basis. This approach begins by simplifying the images, separated in different layers of information (shapes and lines). Then, for each layer, we extract patterns used to describe the images. Thus, images are described with most common bags of patterns, and bags of stroke. For these two layers of information, we have also extracted graphs of regions that allow extracting a more structural knowledge of the images. A more complex description is then inserted into the knowledge base in order to allow complex queries. The purpose of this database is to offer the possiblity to make either query by example, or query by specific features of the images, to user.

Keywords : Old documents, Complex image analysis, Indexing, Ontologies