



Analyse et recherche d'oeuvres d'art 2D selon le contenu pictural

Thomas Hurtut

► To cite this version:

Thomas Hurtut. Analyse et recherche d'oeuvres d'art 2D selon le contenu pictural. Interface homme-machine [cs.HC]. Télécom ParisTech, 2008. Français. NNT : . tel-00426645

HAL Id: tel-00426645

<https://pastel.hal.science/tel-00426645>

Submitted on 27 Oct 2009

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

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



Thèse

présentée pour obtenir le grade de docteur
de l'Ecole nationale supérieure des télécommunications
Spécialité : Traitement de signal et images

Thomas Hurtut

Analyse et recherche d'oeuvres d'art 2D selon
le contenu pictural

Soutenu le 4 Mars 2008 devant le jury composé de :

Alberto Del Bimbo
Jean-Marc Chassery
Jean-Michel Morel
Gilles Pesant
Farida Cheriet
Yann Gousseau
Francis Schmitt

Rapporteur
Examineur
Rapporteur
Examineur
Co-directrice de thèse (Canada)
Co-directeur de thèse (France)
Co-directeur de thèse (France)

*“Mon dessin au trait est la traduction
directe et la plus pure de mon émotion.”*

*“Je sens par la couleur, c’est donc par elle
que ma toile sera toujours organisée.”*

Henri Matisse

Remerciements

Je n'aurais jamais pu mené à bien cette thèse sans l'encadrement, les discussions et les encouragements dont j'ai bénéficiés. C'est avec un sentiment de profonde gratitude que je souhaite remercier toutes les personnes qui ont participé à ce projet. Si je remonte la chronologie des événements qui ont permis d'arriver jusqu'ici, un des premiers est sans nul doute ce jour de septembre 2001 où j'ai franchi le seuil du bureau de Farida Cheriet. Farida m'a inlassablement soutenu entre autres dans ce projet de thèse dont je lui avais déjà esquissé les contours à notre première rencontre. Je la remercie pour son courage, ses nombreux conseils et ses précieux encouragements. Plus tard, c'est la porte du bureau de Francis Schmitt que j'ai également eu la chance de franchir¹, lui-même m'ouvrant ensuite celle de Yann Gousseau. Je les remercie de m'avoir donné cette chance et d'avoir également accepté de m'encadrer sur ce sujet relativement marginal. Leur patience, leur investissement et leur foi inébranlable en leurs motivations me resteront gravés. Christian Lahanier du C2RMF a également participé à la genèse de ce projet. Je le remercie de m'avoir accueilli dans son bureau pour discuter et réfléchir ensemble. Je remercie vivement Alberto del Bimbo, Jean-Marc Chassery, Jean-Michel Morel et Gilles Pesant de s'être intéressé à nos travaux, et d'avoir accepté de participer au jury de cette thèse.

Cette thèse a été soutenue financièrement par le Fond Québécois de la Recherche sur la Nature et les Technologies (FQRNT), l'Institut Télécom (ex-GET), le laboratoire LIV4D de l'Ecole Polytechnique de Montréal, et le projet Européen Muscle.

Ma gratitude va également à Maria Chronopoulos et Jacynth Tétrault de l'Atelier Circulaire à Montréal qui ont toujours su écouté mes idées parfois assez saugrenues avec beaucoup d'attention. Merci à tous les artistes de l'atelier de m'avoir accueilli chez eux et confié leurs oeuvres pour qu'elles soient numérisées ; Carlos Calado à Montréal ainsi qu'Anne-Marie Prache à Metz pour m'avoir transmis un peu de leur passion avec un génie pédagogique rare.

Je mesure pleinement aussi tout l'importance de l'aide technique patiemment réalisée par Sophie-Charlotte Barrière à Paris et Luc Lalonde à Montréal. Combien de fois m'avez-vous sauvé après avoir débarqué dans votre bureau ! Je tiens à remercier l'équipe administrative de choc constituée par Patricia Friedrich à Paris et Jeanne Daunais à Montréal. Combien d'heures sauvées grâce à vous, et quel plaisir à vous cotoyer quotidiennement. Je pense aussi à Madeleine Guillemet dont j'ai monopolisé le scanner durant

¹ Avec Christian Lahanier et Henri Maître, j'espère qu'ils n'en veulent plus à ce "Canadien obstiné" de les avoir inondé de courriels et d'appels téléphoniques.

plusieurs jours pour numériser une des bases de données utilisée dans cette thèse. Philippe Débanné qui a constitué une aide multidisciplinaire en de nombreuses circonstances, a toujours été d'une écoute amicale attentive.

Ma reconnaissance va également à mes collègues et amis du laboratoire LIV4D et du laboratoire TII : Jérémie Anquez (vive la F.A. team), Rafik Bourezak (frerot!), Mihai Costache (alors combien?), Julien Delliére (bientôt la quille), Loïc Denis (une noisette?), Pascal Fallavollita (*pretty good!*), Rola Harmouche (*bashte*), Ivan Kyrgyzov (bons TDs!), Gabrielle Lehureau (merci encore pour le bananier-sitting), Marie Lienou (l'éléphant sera sur mon prochain bureau), Hervé Lombaert (pardon pour la sculpture de cannettes sur ton bureau), Pénélope Lopez Quiroz (merci pour les conseils sur le Chiapas), Bin Luo (bonne chance à Grenoble!), Alexandre Mallet (t'achètes la rue de la paix?), Antonio Moreno (un grand amateur de noisettes), Thibault Napoleon (tu nous prêtes tes clés?), Olivier Nempont (la bodegita del dude), Julien Rabin $(-)\rightarrow<(*v^o)>!$, Tony Tung (tu dois encore m'expliquer tes courbes), Gui-Song Xia (*hui jiàn*).

Je tiens également à remercier mes proches de leur présence et pour m'avoir encouragé toutes ces longues années d'études secondaires. Mes soeurs et mes parents, ainsi que Didier pour m'avoir tant soutenu. Je ne peux oublier Mathieu, incroyable complice. Enfin je tiens à remercier celle qui m'accompagne. Cléo a été capable de me soutenir et de me critiquer en même temps, de me conseiller et de me guider sur cette thèse comme sur beaucoup d'autres choses.

Résumé

Les institutions culturelles mènent depuis une vingtaine d'années une politique de sauvegarde numérique exhaustive de leurs collections conduisant à la création de bases d'images de plus en plus grandes. Les oeuvres d'art se différencient des images naturelles car elles sont souvent stylisées. Cette caractéristique influence notre interprétation de l'image et l'impression visuelle qui nous est transmise. Nous proposons dans cette thèse des descripteurs et mesures de similarité spécifiques au contenu pictural et les testons dans le cadre de la recherche par le contenu d'images similaires. Le premier aspect du contenu pictural auquel nous nous sommes intéressés concerne l'organisation spatiale globale des couleurs. Nous modélisons le problème de la comparaison entre deux organisations spatiales des couleurs par un problème de transport optimal appliqué à des imagerie. Dans le cadre de la recherche d'images dans une base, nous proposons également un seuil adaptatif sur cette distance de transport fondé sur une approche *a contrario*. Le deuxième aspect du contenu pictural que nous étudions concerne les caractéristiques du contenu géométrique lié aux lignes dans les dessins au trait. Pour cette approche nous développons un détecteur de contours de trait sans paramètre reposant sur un filtrage topologique de l'arbre des lignes de niveau significatives de Desolneux et al., ainsi qu'un ensemble de méthodes d'extraction de caractéristiques visuelles incluant les extrémités de trait, les jonctions et les coins. Nous évaluons ces deux approches en utilisant sept bases différentes totalisant environ 65000 images.

Abstract

Cultural institutions around the world have, over the past two decades, implemented a policy of digitally safeguarding their artwork collections, building extremely large image databases. Works of visual art are quite distinct from natural images in that they are often stylized. This property introduces a subjectivity that influences one's understanding of the scene as well as the impression conveyed to the viewer when he/she looks at the image. Such database can benefit from specific content-based features. The first aspect of the pictorial content that we studied concerns the spatial organization of colours within a work of visual art. We proposed a methodology to model the difference between the spatial arrangements of two images as an optimal transportation problem. We also proposed an adaptive matching criterion based on a statistical approach, itself founded on an *a contrario* approach. The second part of my doctoral research focused on characterising the geometrical information and pictorial impact conveyed by linear strokes. The methodology we developed here is based firstly on a novel parameter-less method for the extraction of stroke boundaries in line artworks based on a selection of the level-lines of the topographic map. We also developed a set of methods that analyzes these boundaries to extract several intuitive geometrical cues. We evaluated the performance of these methods on seven different databases comprising a total of about 65000 images.

Table des matières

Remerciements	vi
Résumé	vii
Abstract	ix
1 Introduction	1
2 État de l’art des méthodes manuelles et automatiques d’analyse des œuvres d’art 2D	5
2.1 Description et classification du contenu pictural en histoire de l’art	6
2.1.1 Modes de description du contenu artistique	6
2.1.2 Le style d’un artiste et ses influences	7
2.1.3 Processus cognitif de l’artiste	9
2.1.4 Vers un modèle de représentation de l’art	10
2.1.5 Style et sujet	12
2.1.6 Structure géométrique des peintures	13
2.1.7 Impact pictural	14
2.1.8 Conclusion	15
2.2 Analyse automatique du contenu artistique	16
2.2.1 Taxonomie du contenu image et des besoins d’utilisateurs d’images	17
2.2.2 Authentification d’une œuvre ou identification d’un artiste	17
2.2.3 Primitives bas-niveaux	20
2.2.4 Primitives géométriques	21
2.2.5 Relations visuelles 2D globales	22
2.2.6 Relations visuelles 3D	24
2.2.7 Unités sémantiques	25
2.2.8 Abstraction culturelle, style	25
2.2.9 Abstraction émotionnelle	27
2.2.10 Abstraction technique	27
2.2.11 Métadonnées	27
2.2.12 Conclusion	28
2.3 Rendu non-photoréaliste	29
2.4 Conclusion générale	31

I	Organisation spatiale des couleurs	33
3	Recherche d'images selon l'organisation spatiale des couleurs	35
3.1	Travaux connexes à la recherche d'images selon l'organisation spatiale des couleurs	36
3.1.1	Espaces couleur	36
3.1.2	De la couleur à l'information couleur-espace	37
3.1.3	Mesures de similarité	39
3.1.4	Utilisation d'histogrammes couleurs	41
3.1.5	Utilisation de méthodes de segmentation d'images	43
3.2	Comparaison de l'OSC sous la forme d'un problème d'optimisation global	45
3.2.1	Appliquer l'EMD aux pixels	46
3.2.2	Utilisation d'imagettes	47
3.2.3	Distance d_e inter-pixels exponentielle	48
3.2.4	Ajout d'informations supplémentaires	48
3.3	Expériences	53
3.3.1	Comportement de l'EMD et effet <i>taquin</i>	53
3.3.2	Facteur de pondération α	56
3.3.3	Variabilité des formats d'images	57
3.3.4	Taille des imagettes	60
3.3.5	Comparaison avec l'utilisation d'une méthode de segmentation	61
3.3.6	Apport des imagettes bicolores	65
3.3.7	Base d'œuvres en noir et blanc	68
3.3.8	Plusieurs bases d'essais	68
3.4	Conclusion	69
4	Seuil automatique pour la recherche d'images selon l'OSC	81
4.1	Méthodes a contrario pour la recherche dans des bases de données	82
4.1.1	Méthodes a contrario	82
4.1.2	Application à la recherche dans des bases de données	82
4.2	Critère de décision non supervisé	83
4.2.1	Une décision sous forme de test statistique	83
4.2.2	Mise en correspondance ε -significative	84
4.2.3	Modèle de fond	85
4.3	Expériences	87
4.3.1	Estimation numérique de δ_Q	87
4.3.2	Influence de la taille de la base d'imagettes bruitées	87
4.3.3	Variabilité du seuil	88
4.3.4	Optimisation de l'algorithme de calcul de l'EMD	91
4.3.5	Exemples de requêtes par l'exemple	92
4.4	Conclusion	94

II	Geométrie des lignes	107
5	Extraction des contours des traits	109
5.1	Travaux connexes sur les méthodes d'extraction des contours	110
5.1.1	Méthodes locales	110
5.1.2	Contours actifs	111
5.1.3	Utilisation de la carte topographique et des lignes de niveaux	111
5.1.4	Reconnaissance automatique d'écriture cursive	113
5.1.5	Reconstruction et analyse de dessins d'ingénierie	115
5.2	Détection a contrario des contours dans une image	115
5.2.1	Lignes de niveaux significatives	116
5.2.2	Lignes redondantes et principe de maximalité	118
5.3	Extraction des contours de traits dans les dessins	119
5.3.1	Principe de maximalité non supervisé appliqué aux arbres mono- tones maximaux	121
5.3.2	Ensemble d'adhérence	123
5.4	Expériences	124
5.4.1	Comparaison avec une méthode de squelettisation	126
5.4.2	Comparaison avec le principe de maximalité de Desolneux et al. . .	126
5.4.3	Variations d'intensités le long des traits	129
5.4.4	Images réelles	131
5.5	Conclusion	132
6	Analyse de l'impact pictural des œuvres au trait	137
6.1	Travaux connexes	138
6.1.1	Le choix de la courbure	138
6.1.2	Calcul de la courbure locale	139
6.1.3	Modes d'utilisation de la courbure	139
6.1.4	Points d'intérêts, coins et jonctions	140
6.1.5	Caractérisation du style d'écriture	141
6.2	Analyse du contenu géométrique	143
6.2.1	Lissage des contours	143
6.2.2	Estimation du rayon de l'outil	144
6.2.3	Calcul de la courbure	144
6.2.4	Extraction des points d'inflexions	147
6.2.5	Extraction des extrema de courbure	147
6.2.6	Croisements de traits	148
6.2.7	Coin	149
6.2.8	Extrémités de traits de dessins	152
6.3	Descripteurs et mesure de similarité	153
6.3.1	Descripteurs	154
6.3.2	Mesure de similarité	155
6.4	Expériences	155
6.4.1	Estimation de la courbure	156
6.4.2	Estimation du rayon de l'outil	156

TABLE DES MATIÈRES

6.4.3	Paramètre k_c	158
6.4.4	Paramètres k_e et k_j	159
6.4.5	Base de donnée classée selon l'impact pictural	159
6.4.6	Détection des extrémités, croisements et coins	162
6.4.7	Résultats d'indexation comparés	165
6.4.8	Exemples de requête sur la base classée	171
6.4.9	Exemples de requête sur une base non classée	171
6.4.10	Discrimination de textons	171
6.5	Conclusion	177
7	Conclusion et perspectives	181
A	Bases d'images utilisées dans les expériences	185
A.1	Bases d'images utilisées pour l'étude de l'organisation spatiale des couleurs	185
A.2	Bases d'images utilisées pour l'étude du contenu géométrique	187
A.3	Remerciements	188
B	Problème de transport optimal à deux classes de poids total égaux	193
B.1	Notations et définition du problème	193
B.2	Chemins de départs	194
B.3	Condition d'optimalité	195
	Bibliographie	215
	Publications	218

Chapitre 1

Introduction

Contexte

Les institutions culturelles du monde entier ont démarré depuis une vingtaine d'années une politique de sauvegarde exhaustive de leurs collections. Leurs objectifs vont de la préservation numérique des œuvres et du suivi de leur état, à l'amélioration de l'accessibilité des œuvres au public et aux professionnels pour toutes sortes d'applications. Ce dernier aspect est perçu comme très important par les musées car il fait partie intégrante de leur mission. Cette politique de numérisation a conduit les musées à constituer des bases de données d'œuvres extrêmement riches et souvent disponibles sur Internet avec de plus ou moins bonnes résolutions. Par exemple,

- le Louvre mets en ligne sur son site les 30000 œuvres exposées dans ses salles¹ et les 140000 œuvres de son département d'arts graphiques²,
- le musée de l'Ermitage à Saint-Petersbourg a mis en ligne environ 10000 œuvres 2D (peintures, dessins, estampes)³,
- le Rijksmuseum à Amsterdam a mis en ligne la totalité de sa collection d'œuvres 2D, soit environ 3000 œuvres⁴.

D'autres institutions rassemblent et publient électroniquement de très grandes quantités d'œuvres d'origines diverses. Par exemple,

- la base Joconde regroupe 300000 notices des musées de France souvent accompagnées

¹cartelfr.louvre.fr

²arts-graphiques.louvre.fr

³www.hermitagemuseum.org/

⁴www.rijksmuseum.nl

-
- d'une image de l'œuvre⁵
- le Musée Virtuel du Canada regroupe environ 420000 œuvres de 1200 institutions culturelles Canadiennes⁶.
 - ARTstor est une association à but non lucratif regroupant les collections de plusieurs musées américains et disponibles à des fins pédagogiques pour l'enseignement supérieure et les musées du monde entier. Le site de l'association donne accès à environ 550000 images d'œuvres⁷.
 - le site Artprice, leader mondial pour l'archivage et l'analyse du marché de l'art vend un accès à 290000 catalogues regroupant les œuvres de 400000 artistes différents⁸
 - la banque d'images privée Corbis regroupe environ 100000 œuvres d'arts muséales dont l'agence a acquis les droits⁹
 - l'agence Getty Images regroupe environ 300000 images artistiques¹⁰

Problématique

Pouvoir consulter et parcourir des bases d'images aussi volumineuses est une tâche complexe [Tsai, 2007, Stanchev et al., 2006]. Du point de vue de l'utilisateur, plusieurs stratégies de recherche peuvent être envisagées selon qu'il ait une idée précise d'une image qu'il souhaite retrouver, ou au contraire qu'une catégorie d'œuvres correspondent à son critère de recherche (toutes les œuvres d'un artiste par exemple). Les images d'œuvres d'art ont ceci de particulier par rapport aux images naturelles qu'elles sont très souvent stylisées. Cette propriété introduit une subjectivité qui peut se traduire non seulement au niveau de la compréhension du contenu de la scène (par le biais de multiples ambiguïtés), mais aussi dans l'impression ressentie lorsqu'on regarde l'image. Ce dernier aspect est d'ailleurs également spécifique aux œuvres d'arts. Nous ne regardons pas un tableau avec la même attention qu'une image naturelle. Le visiteur d'un musée peut attendre une certaine expérience esthétique très abstraite et éphémère. Plus concrètement, le visiteur d'une galerie peut de même chercher à trouver l'œuvre qui ornera au mieux une pièce de sa maison selon cette expérience. Ce contenu et cette utilisation particulière des œuvres appellent des critères de recherche spécifiques complémentaires aux critères classiques que nous pouvons rencontrer en recherche d'image.

A l'heure actuelle, les bases de données muséales utilisent le plus souvent une indexation par métadonnées, c'est-à-dire par des champs textuels saisis pour chaque œuvre par un expert. Cette démarche coûteuse est valide et utile pour des données factuelles telles que le nom de l'artiste ou la date de création. En revanche, pour tout ce qui est lié au contenu de l'œuvre (palette de couleurs, style, sujets représentés), cette stratégie est trop

⁵www.culture.gouv.fr/documentation/joconde

⁶www.museevirtuel.ca

⁷www.artstor.org

⁸web.artprice.com/catalogue

⁹www.corbis.com

¹⁰www.gettyimages.com

fortement soumise à la subjectivité et la non reproductibilité d'une telle tâche. Ceci a été mis notamment en évidence par [David, 2004].

L'indexation automatique d'une base d'images selon le contenu consiste à utiliser une méthode permettant de calculer pour chaque image un ensemble d'attributs descriptifs. La définition d'une mesure de similarité utilisant ces descripteurs permet de comparer deux images entre elles ou de soumettre une requête sur l'ensemble de la base pour effectuer une recherche. Cette requête peut avoir été exprimée sous la forme d'un exemple choisi à l'intérieur ou à l'extérieur de la base, ou esquissé manuellement par l'utilisateur. Habituellement, les images les plus proches sont ensuite retournées à l'utilisateur.

Objectifs et contributions

L'objectif principal de cette thèse a été de proposer de nouveaux descripteurs adaptés au contenu spécifique des œuvres d'arts. La majorité des approches proposées dans la littérature scientifique repose sur des descripteurs statistiques caractérisant certains aspects locaux du contenu pictural comme l'étude de la couleur des pixels dans l'image, de leur organisation spatiale locale en textures à l'aide d'histogrammes, ou d'outils de décomposition ou de segmentation. Cela a suscité une littérature relativement riche depuis une dizaine d'années. Cependant les approches s'efforçant de dépasser le cadre des méthodes directement dérivées de celles appliquées à l'analyse des images naturelles sont assez rares.

Afin de mieux comprendre les caractéristiques spécifiques du contenu pictural, nous nous sommes dans un premier temps inspirés d'une littérature portant sur l'étude et l'analyse du contenu artistique d'une œuvre en histoire de l'art. Plusieurs historiens, psychologues et spécialistes de la cognition se sont en effet penchés sur ce medium pour comprendre les mécanismes de l'œil de l'artiste et du regard du spectateur. Une œuvre d'art 2D a en effet la particularité d'avoir été visuellement produite par le biais d'un système visuel humain, celui de l'artiste.

Cette étude nous a conduit vers deux approches différentes que nous exposons dans les deux parties de cette thèse. Ces approches proposent d'étudier deux aspects complémentaires du contenu pictural qui peuvent répondre à des besoins ou critères de recherche différents. La première partie porte sur l'étude de l'organisation spatiale globale des couleurs dans une œuvre. Cette caractéristique décrit l'agencement spatiale des régions homogènes par rapport au support utilisé. Cette caractéristique est très proche de ce que nous appelons souvent la composition picturale d'une œuvre. La manière dont s'organisent les régions colorées dans une œuvre est un descripteur artistique qui s'applique à tout type d'œuvres (peintures, photos, dessins etc.) et qui est liée à la notion d'équilibre. La composition d'un tableau peut influencer l'impression ressentie indépendamment du contenu sémantique de l'image. Cette organisation est souvent au cœur du processus créatif lors de la construction de l'œuvre.

La deuxième partie porte sur l'étude du contenu géométrique des traits visibles dans les dessins. Particulièrement adaptée aux œuvres d'art graphique, la méthodologie proposée introduit une description possible du contenu pictural lié aux lignes tracées par l'artiste et de l'impact visuel transmis par leur biais. Ce type de contenu pictural est proche du style de l'artiste mais il est également indissociable du sujet représenté.

Les méthodes proposées dans ces deux parties s'intéressent et se limitent à deux aspects bien particulier du contenu pictural. Les résultats expérimentaux ont été réalisées sur un grand nombre de bases de tests et d'images : 5 bases différentes regroupant au total environ 60000 images pour la première partie, et 2 bases d'environ 200 dessins pour la deuxième partie (Annexe A). Ce nombre élevé de résultats est d'autant plus appréciable qu'il est difficile de constituer ou obtenir des bases d'images dans ce domaine où les droits d'auteurs imposent souvent une réticence forte de la part des institutions. Nous remercions donc vivement les institutions qui ont collaboré avec nous dans ce projet.

Organisation du document

Nous présentons tout d'abord dans le **Chapitre 2** les principales approches utilisées pour décrire le contenu pictural des œuvres d'art en histoire de l'art ainsi que celles étudiant les mécanismes de construction utilisés par l'artiste pour organiser le contenu pictural d'une œuvre. Puis nous présentons les méthodes automatiques issues de l'analyse d'image qui ont été proposées pour étudier le contenu pictural.

La Partie I débute avec le **Chapitre 3** où nous proposons d'étudier le problème de la comparaison de l'organisation spatiale des couleurs entre images sous la forme d'un problème d'optimisation global de transport. Nous complétons cette méthode en introduisant un seuil automatique de similarité dans le **Chapitre 4**. Ce seuil permet de fixer pour chaque requête et chaque base la limite entre les images similaires à une requête et les images différentes.

Nous débutons la Partie II avec le **Chapitre 5** où nous proposons une méthode non supervisée d'extraction des contours des traits dans les œuvres au trait. Puis nous introduisons au **Chapitre 6** un ensemble de méthodes analysant la géométrie des contours extraits et permettant le calcul d'un ensemble de 11 attributs.

En annexes de ces travaux sont présentés en détail les bases qui ont été constituées pour tester l'ensemble des méthodes proposées dans cette thèse, et les comparer d'autres. Les différentes contributions apportées dans cette thèse ont fait l'objet de plusieurs publications dont la liste est visible à la page 217.

Chapitre 2

État de l’art des méthodes manuelles et automatiques d’analyse des œuvres d’art 2D

Dans cette thèse nous nous intéressons à l’analyse automatique du contenu pictural dans des œuvres 2D. Par contenu pictural nous entendons ce qui a rapport à la peinture, les techniques utilisées, la touche de l’artiste et ses effets, et d’une manière générale toutes les caractéristiques visibles liées au processus technique de création artistique. Le contenu pictural se rencontre dans tous les arts dit visuels. Historiquement, ce contenu a été étudié selon diverses approches. L’Histoire de l’art est notamment l’histoire de l’évolution des caractéristiques du contenu pictural à travers les époques. En traitement et analyse automatique d’images, de nombreux travaux ont aussi cherché à étudier les différentes modalités du contenu pictural.

Dans ce chapitre, nous commençons par décrire en Section 2.1 plusieurs angles d’approche explorés par les théoriciens en arts pour analyser une œuvre, ou par les artistes pour produire une œuvre. Cette riche et longue expérience nous permet à nous chercheurs en traitement de l’image de mieux comprendre les cheminements par lesquels sont passés les artistes dans leur apprentissage. Elle nous permet aussi de contextualiser les méthodes proposées à l’aide d’une terminologie familière aux professionnels des arts, qu’ils soient historiens, artistes ou conservateurs. Dans la Section 2.2, nous ferons un état de l’art que nous avons voulu aussi exhaustif que possible des travaux qui ont cherché à analyser automatiquement le contenu pictural d’œuvres 2D. Nous terminerons à la Section 2.3 par bref aperçu du domaine de la synthèse d’images non photo-réalistes. Ce domaine de recherche s’intéresse entre autres au processus créatif des artistes pour produire des images artistiques de synthèse qui soient plus *réalistes*.

2.1 Description et classification du contenu pictural en histoire de l'art

Dans cette section nous allons présenter quelques approches qui permettent non seulement de décrire une œuvre et son contenu pictural (Section 2.1.1 à 2.1.2) mais aussi de comprendre le processus créatif de l'artiste et les caractéristiques visibles qui en résultent (Section 2.1.3 à 2.1.7). Cette deuxième préoccupation est doublement justifiée. Les méthodes de vision qui devront être mises en œuvre se substituent premièrement au regard du spectateur quand il regarde une certaine œuvre. Mais dans le contexte des arts visuels, et à la différence de l'imagerie médicale par exemple, l'image elle-même que nous étudions est issue d'un processus purement cognitif et sensoriel.

2.1.1 Modes de description du contenu artistique

En histoire de l'art, une œuvre peut être tout d'abord analysée d'une manière iconographique. L'approche iconographique consiste à étudier le contenu sémantique d'une œuvre et à y reconnaître les stéréotypes visuels qui permettent d'identifier le sujet représenté au delà de la dimension esthétique [Gombrich, 2000]. Les résultats actuels de la recherche en analyse d'image par ordinateur permettent difficilement d'envisager des méthodes automatiques d'analyse iconographique.

Les approches non iconographiques décrivent les œuvres selon différents niveaux. Par exemple en analysant les caractéristiques contextuelles [Kirsh and Levenson, 2000] de l'œuvre : nom de l'artiste, médium utilisé, date de création, etc. Nous trouvons aussi des descriptions selon des niveaux plus subjectifs tels que le titre, le type de sujet représenté (portrait, paysage, ...), le courant artistique auquel appartient l'œuvre. Enfin nous trouvons aussi des descriptions selon des caractéristiques plus formelles telles que la couleur, la composition, la lumière, les contrastes, les formes, les lignes [Barnet, 1985, Sayre, 1989, Carr and Leonard, 1992]. Ces dernières approches utilisent un vocabulaire très familier à la communauté du traitement d'images.

Ces différentes catégories ne sont ni clairement identifiées, ni indépendantes. Un courant artistique peut influencer la palette de couleur. Le clair-obscur se définit par exemple en terme de caractéristiques formelles. Le médium utilisé peut aussi influencer à lui seul des caractéristiques tels que la ligne ou la texture. Malgré tout, nous voyons donc que l'histoire de l'art s'accommode d'un niveau de description formel [Wolfflin, 1950, Wollheim, 1993, Shapiro, 1994] et aussi depuis longtemps utilise un niveau de description fondé sur des caractéristiques susceptibles de nous offrir des pistes d'analyse automatique.

2.1.2 Le style d'un artiste et ses influences

Le contenu pictural est intimement lié à la notion de style comme en témoignent les nombreuses définitions du style que nous pouvons trouver dans la littérature (voir Tableau 2.1). La notion de style recouvre des types d'influences qui ont évolué dans l'histoire de l'art [Gombrich, 2000]. Jusqu'au XVIII^e siècle, le style était collectif dans le sens où les artistes s'efforçaient de suivre des règles iconographiques communes de représentation de la réalité, dictées par leur capacité à la reproduire fidèlement, mais aussi par les traditions de leur époque. Par exemple les égyptiens avaient pour habitude de représenter certaines parties des personnages de face et d'autres de profil (Figure 2.1). Nous savons aujourd'hui que ceci n'était pas le reflet d'une incapacité à savoir utiliser certaines méthodes de représentation comme les raccourcis¹, qui aurait donné un effet plus réaliste, mais bien souvent un choix délibéré d'un mode de représentation. L'art n'a pas toujours poursuivi une quête de réalisme. Les égyptiens peignaient ce qu'ils savaient, contrairement par exemple aux impressionnistes qui cherchaient à ne peindre que ce que leurs yeux voyaient. Malgré cela, comme l'analyse Gombrich, le style domine et commande, même quand l'artiste veut imiter fidèlement la nature [Gombrich, 2002]. L'énigme du style serait en partie dans les limitations imposées par : la technique (les outils utilisés), son apprentissage (initial et continu), et l'objectif visé par la représentation. A partir du XVIII^e siècle, l'histoire de l'art selon Gombrich est l'histoire d'une guerre contre le schéma, la mort de l'art narratif et la naissance de l'art contemplatif qui offre un regard neuf sur la nature. Par la suite le style est devenu de plus en plus personnel à chaque artiste, même au sein d'un courant artistique, avec un contenu descriptif de moins en moins présent (jusqu'à l'art dit abstrait). L'artiste s'est progressivement libéré des schémas de la tradition (Figure 2.1). Depuis la fin de l'art abstrait, nous assistons inversement à une perte du contenu stylistique vers une œuvre purement liée à son contenu ou son message [Lucie-Smith, 1999].

Les artistes subissent pleinement l'influence de leur système visuel. Ils l'ont souvent beaucoup étudié eux-même, et ils ont aussi parfois voulu s'en affranchir². Ils ont souvent cherché à jouer avec les principes perceptuels et compris l'œil du peintre³, comme celui du spectateur. Notre regard puise dans l'imagination, et l'artiste a appris à laisser à celle-ci une part de liberté dans ses tableaux, ce qui s'est traduit par de nombreuses recherches en particulier après l'apparition de la photographie [Gombrich and Eribon, 1998].

¹Terme désignant un effet visuel qui tend à exagérer la perspective par une réduction de celle-ci [Bersani, 1968].

²En particulier les impressionnistes. Cézanne : *"Monet n'est qu'un œil, mais quel œil !"* [Gombrich, 2002]

³*"Il [l'artiste] s'agit d'un homme qui a su apprendre à regarder les choses d'un œil critique, à sonder les divers éléments de ses perceptions en procédant à plusieurs interprétations, soit par jeu, soit d'une façon sérieuse"* [Gombrich, 2002]

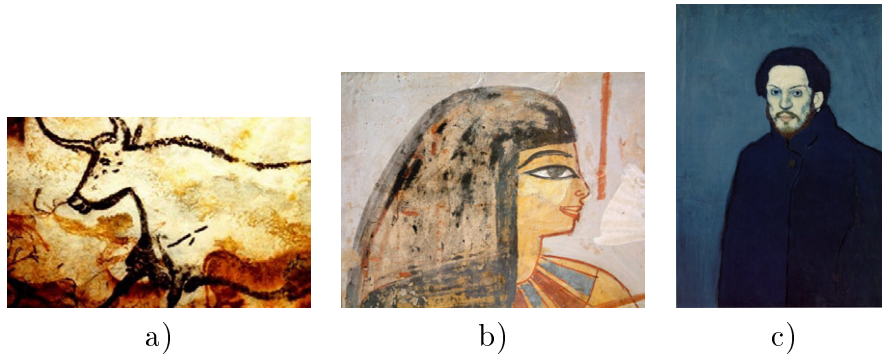


FIG. 2.1: Histoire de l'art. a) l'histoire a commencé dans les grottes préhistoriques où déjà, les *artistes* faisaient preuve d'une grande aptitude à représenter des éléments de la réalité (25000 ans avant Jesus-Christ). b) l'art égyptien (environ 3000 ans avant Jesus-Christ) comme l'art religieux du moyen âge est un bon exemple d'art impersonnel, régi par des règles picturales au service d'une autorité (divine, ou royale par exemple). c) Un portrait de la période bleue de Picasso (1903). Depuis environ 200 ans, l'art n'est plus narratif, il est contemplatif. Il est le lieu d'une guerre contre le schéma selon [Gombrich, 2002].

Selon ...	... le style, c'est :
Lucie-Smith [Lucie-Smith, 1999]	[...] un langage visuel spécifique se distinguant des autres formes de langage visuel
Lucie-Smith [Lucie-Smith, 1999]	[...] les déformations qui surviennent lorsque le subjectif empiète sur l'objectif
American Dictionary	The combination of distinctive features of artistic expression, execution or performance characterizing a particular person, group, school or era
De Kooning	Le style n'est qu'une supercherie.

TAB. 2.1: Différentes définitions du style.

2.1.3 Processus cognitif de l'artiste

Lorsqu'un artiste analyse son œuvre en cours de création, ou lorsque nous regardons nous même une œuvre nous reconstruisons visuellement ce que nous y voyons. Nous détectons en effet les objets qui y sont représentés, qu'ils soient figuratifs (un fruit) ou abstrait (des lignes, formes 2D, etc.). Ce processus de reconstruction visuel a été étudié par l'école de la Gestalt à partir de 1923, durant une cinquantaine d'années [Köhler, 1967, Ales et al., 1991, Kanizsa, 1996]. Les lois qui ont été proposées par cette école tentent de modéliser le fonctionnement de notre système visuel lorsque par exemple, nous voyons une droite là où notre rétine n'a physiquement qu'un ensemble de cellules alignés stimulés par les photons perçus. De même, une image numérique n'est qu'une quantité d'information représentée par un ensemble de pixels discrets. Pourtant notre œil est capable, avec une grande facilité, de reconstruire les éléments géométriques visuellement représentés par cette image.

A l'époque de leur publication, les lois de la Gestalt n'ont bénéficié d'aucune traduction mathématique. Et par conséquent elles ont été rarement étudiées et formalisées au début de la recherche en vision par ordinateur. [Marr, 1982] par exemple se fonde sur une approche plus neurophysiologique que les travaux de la Gestalt. Probablement comme le pense [Ales et al., 1991] parce que les lois Gestaltistes de regroupement se basent sur une détection préliminaire et implicite des primitives géométriques élémentaires tels que les points, droites et courbes. Or il s'est avéré que l'extraction de ces éléments étaient une tâche extrêmement difficile à réaliser de manière automatique. Aujourd'hui, les lois de la Gestalt bénéficient d'un regain d'intérêt [Desolneux et al., 2008]. Nous utiliserons à plusieurs reprises dans cette thèse des résultats de ces travaux récents.

[Arnheim, 1954] a également cherché à comprendre la perception visuelle et en particulier celle de l'artiste et du spectateur face à une œuvre. Il formule plusieurs concepts haut niveaux, comme l'équilibre, la forme, l'espace, la lumière, la couleur et le mouvement, que l'artiste utilise. Pour chacun d'eux, il recherche les outils à sa disposition [Arnheim, 1969]. Il appuie notamment ses hypothèses par une analyse des dessins d'enfants et leurs évolution au cours de la croissance. Ces recherches témoignent d'un effort de compréhension et de formalisation des modes de représentation en art visuel, quel qu'il soit.

Plus récemment, dans le domaine des neurosciences et de la cognition, Cavanagh a observé que les artistes utilisent un nombre limité de lois physiques par rapport à celles connues, sans que l'œil humain s'en aperçoive [Cavanagh, 1999, Cavanagh, 2005]. Ces lois peuvent même être différentes comme dans le cas des dessins au trait. Nous sommes capables de reconnaître et de reconstruire un objet dessiné au trait bien que les objets réels décrit puissent bien souvent n'en contenir aucun comme l'observe aussi [Willats, 2006]. Il a été montré également que cette propriété du système visuel humain est indépendante de l'origine de l'utilisateur (occidentale, tribus primitives, et même chez certains animaux). De même, les artistes utilisent des lois physique plus simples concernant l'optique, les projections 3D, et la gestion des ombres et sources lumineuses [Cavanagh, 2003]. Ces lois amènent à des *incohérences* visuelles qui sont souvent invisibles au premier regard, bien

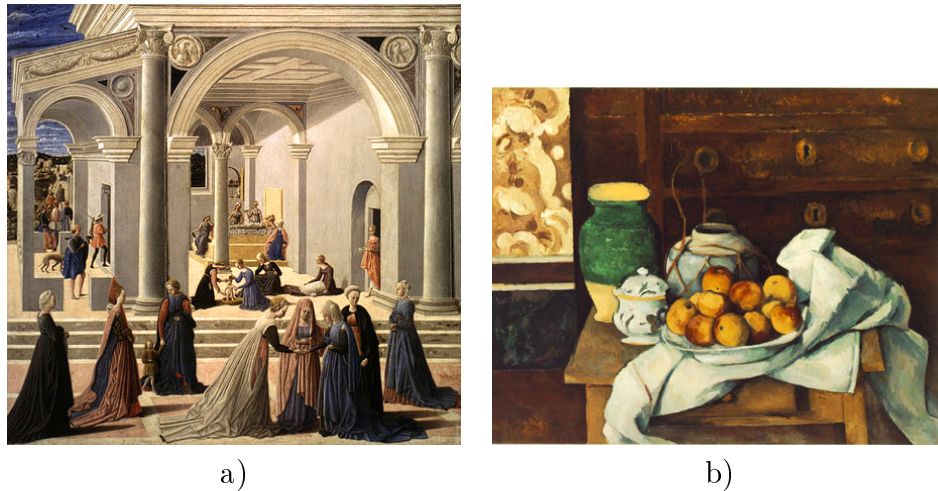


FIG. 2.2: Les artistes adaptent les lois de l'optique. a) sur ce détail de la naissance de la Vierge de Fra Angelico (1467), les ombres qui commencent à faire leurs apparitions dans les tableaux de la pré-renaissance sont non uniformément représentées. Grandes et prononcées au premier plan, elles sont parfois inexistantes au second plan et dans la pièce où il règne une grande luminosité malgré une faible ouverture. b) Cézanne en 1883 et bien avant la période du cubisme déstructure volontairement la perspective frontale de la commode dans cette nature morte.

qu'elles soient en contradiction avec les lois de Physique (Figure 2.2a)). En quelque sorte l'artiste utilise un registre perceptif simplifié, mais suffisant pour obtenir une impression visuelle non choquante pour le spectateur. Enfin, l'artiste peut aussi consciemment déformer la réalité pour produire un effet visuel ou organiser sa toile comme Cézanne sur la Figure 2.2b).

2.1.4 Vers un modèle de représentation de l'art

Parallèlement à Arhneim, des artistes en arts visuels ont cherché quelque fois à expliquer et comprendre leur propre manière de travailler. Ceci a notamment été fait dans un but pédagogique à l'école du Bauhaus avec par exemple les cours de Klee [Klee, 1964, Klee, 2004], Moholy-Nagy [Moholy-Nagy, 1955] ou Kandinsky [Kandinsky et al., 1991, Kandinsky, 1975]. Leurs écrits souvent très riches souffrent d'un manque de synthèse permettant une utilisation formelle pour les problématiques de vision par ordinateur qui nous intéressent. Il est malgré tout intéressant qu'une grande partie de ces études se concentrent sur la ligne [Klee, 2004] et le point en particulier dans le cadre du dessin [Kandinsky et al., 1991].

[Willats, 1997] dans sa monographie a prolongé et formalisé les travaux d'Arhneim. Il s'agit probablement d'un des modèles les plus complets de représentation de l'art visuel.

Les premiers travaux de Willats remontent aux années 1970 avec Dubery [Dubery and Willats, 1972]. De même que la linguistique et la rhétorique ont leurs modèles pour décrire un texte et son style, Willats propose un modèle pour la représentation picturale. Celui-ci est inspiré et validé par de multiples observations de dessins et peintures d'artistes, mais aussi d'enfants [Willats, 1977]. L'évolution des dessins d'enfants entre 0 et 15 ans environ, témoigne de la progression de ses mécanismes de représentation pour décrire une scène 3D sur un support 2D. Les travaux de Willats ont été complétés ensuite par Durand [Willats and Durand, 2005]. Le modèle a été récemment transféré et appliqué en informatique dans un moteur procédural de synthèse semi-automatique de dessins [Grabli et al., 2004]. Willats et Durand proposent quatre niveaux d'analyse. Le premier est le système de projection utilisé par l'artiste, puis le système de dénotation, le système d'attributs, et le système de marques. Ces choix faits par l'artiste sont généralement identifiables par l'œil humain et peuvent donc faire l'objet d'une analyse par l'œil du spectateur ou de l'historien d'art.

1. **Système de projection** : La première étape dans la représentation d'une scène réelle est selon Willats et Dubery le choix d'un système de projection pour représenter en 2D une scène 3D. Le plus connu est la perspective linéaire, mais il en existe d'autres : projection orthogonale, verticale, horizontale, oblique, naïve, inversée, et topologique. Les deux derniers ne sont pas des systèmes de projection au sens géométrique primaire du terme (explicable et visible selon certaines règles sur les rayons lumineux de la scène 3D), mais font partie des modes de représentation possible. Toute représentation peut être décrite selon un de ces modes de projection (l'art abstrait sera souvent topologique par exemple, tout comme les dessins de très jeunes enfants). Willats a observé une évolution du choix de système de projection entre 5 et 13 ans environ, qui va du système topologique à la perspective (linéaire ou naïve). Dans l'histoire de l'art, tous ces systèmes sont présents comme l'illustre également [Panofsky, 1976].
2. **Système de dénotation** : Le système de dénotation est le système de primitives utilisées sur l'image pour représenter les primitives de la scène 3D. C'est à dire le fait que l'artiste (ou un appareil électronique) utilise des primitives sans dimensions (des points, par exemple : la télévision, les photographies, la période pointilliste en peinture), des primitives 1D (lignes, par exemple les dessins au trait) ou 2D (régions en aplats), pour représenter des primitives de la scène 3D de dimension 0, 1, 2 ou 3. Chacune de ces primitives de la scène 3D ou 2D a des propriétés d'étalement selon chaque dimension. Savoir quel système de dénotation a utilisé l'auteur n'est pas toujours évident, et plusieurs interprétations peuvent être parfois possibles (par exemple, un contour au trait 1D dans l'image peut délimiter un volume ou une région plane de la scène).
3. **Système d'attributs** : Le système d'attributs correspond aux attributs de chaque primitive de l'image : couleur, intensité, géométrie du squelette pour les traits, épaisseur, etc.
4. **Système de marques** : Le système de marques, proposé par [Willats and Durand, 2005], est lié au choix de l'association physique outil/support utilisée et la texture

qui en résulte (fusain, crayon, peinture, feutre, papier photo couleur, etc.).

2.1.5 Style et sujet

Willats a proposé un ensemble de règles et de choix précis offerts à l'artiste lorsque celui-ci cherche à représenter une scène 3D sur l'espace 2D de son support. Ces choix sont généralement reconnaissables a posteriori par un œil humain aguerri même si il peut subsister quelques ambiguïtés. Ces choix peuvent-ils être détectés analysés de manière automatique sur des œuvres ?

Cela paraît difficile car cela nécessiterait d'avoir en notre possession la scène 3D à l'origine de l'œuvre (lorsqu'elle existe). Même si cela était le cas, nous avons vu dans les sections précédentes qu'un artiste déforme parfois la réalité de ce qu'il voit. De plus, les méthodes de vision par ordinateur sont encore aujourd'hui loin de pouvoir faire de la reconnaissance de générique contenu dans une scène photographiée. Dans le cas des œuvres d'art, la scène est bien souvent stylisée. Cette déformation de la réalité est souvent propre à l'artiste.

Les systèmes de projection et dénotation de Willats sont donc indétectables de manière automatique dans la grande majorité des cas. Les choix liés au système d'attributs peuvent être parfois étudiées en détectant et analysant les caractéristiques des primitives (traits, régions, points) dans un tableau, mais ces caractéristiques ne sont pas toujours dissociables des propriétés liées au sujet. Le système de marques peut être analysé de manière automatique dans une œuvre réelle puisque sa traduction picturale est essentiellement texturale.

Si nous définissons le style comme étant ce qui distingue le contenu subjectif du contenu objectif, la reconnaissance de style est une tâche extrêmement difficile car ces deux contenus se traduisent souvent par les mêmes effets picturaux. Cette tâche demande donc des connaissances avancées en histoire de l'art en plus d'une capacité à comprendre et interpréter une image. Si nous prenons l'exemple d'un tableau de Picasso de sa période dite *bleue* (Figure 2.1c)), le style de cette œuvre est liée à la combinaison de sa couleur, de son auteur, mais aussi des éléments expressifs des personnages, d'éléments iconographiques et chronologiques.

La même remarque peut être appliquée aux dessins au trait. Si nous imaginons le dessin d'une branche de sapin dessinée de manière réaliste, son contour chaotique peut être très proche d'un autre dessin où l'artiste aurait stylisé une feuille de chêne mais d'un trait tout aussi énergique et chaotique. Isoler le style du sujet, ou le sujet du style paraît être un objectif extrêmement difficile à atteindre [Robinson, 1981] bien qu'il ait été poursuivi par un grand nombre de publications en analyse automatique d'image comme nous allons le voir à la Section 2.2. C'est pourquoi, nous proposons dans cette thèse de nous placer essentiellement du point de vue du spectateur, et de chercher à comprendre

et analyser l'impact visuel d'une œuvre du à la combinaison du style de l'artiste et des caractéristiques visuelles liées à la scène représentée.

2.1.6 Structure géométrique des peintures

Leyton a proposé en 2006 des règles d'analyse purement géométriques de la structure des œuvres d'art pour décrire l'impact visuel d'une œuvre sur le spectateur [Leyton, 2006]. Plus précisément il s'intéresse aux contours des formes 2D et primitives 1D. L'idée de Leyton est de voir une œuvre comme un espace mémoire ayant emmagasiné un historique de déformations élastiques. Il énonce une équivalence entre ces transformations élastiques, la courbure le long des formes géométriques résultantes et la *tension émotionnelle* provoquée dans le regard du spectateur.

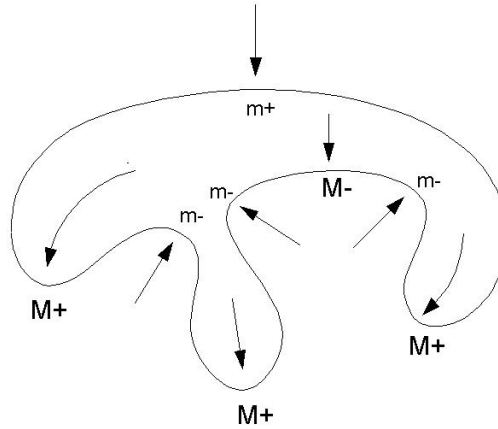


FIG. 2.3: Analyse des extremums par [Leyton, 2006]. Leyton distingue quatre types d'extremums selon qu'ils correspondent à une force venant de l'intérieur (M) ou de l'extérieur (m) d'une forme. Le signe + correspond à une protubérance, alors que le signe – correspond à un creusement.

Plus précisément, il distingue quatre types d'extremums de courbure selon qu'il s'agit de minimum ou maximum locaux, et à l'intérieur ou l'extérieur d'une forme au sens topologique (Figure 2.3). Il énonce aussi des lois d'évolutions de ces extremums le long de l'axe de symétrie local. Un maximum local de courbure crée une tension dite de *pénétration*, alors qu'un minimum local est le lieu d'une *compression* en terme de déformation.

L'attention portée à ces extremums et en particulier aux points de fortes courbures et points de rebroussements rejoint les nombreuses études empiriques et théoriques sur la courbure le long d'un contour que nous verrons à la section 6.1.1. A la différence de ces études, Leyton applique et observe ces propriétés sur la courbure à l'intérieur de dessins ou peintures complexes, et non pas sur un seul contour d'un seul objet.

Leyton propose une analyse des éléments géométriques dans l'œuvre (ceux liés aux

extremums de courbures). La présence à des échelles éventuellement différentes d'extremums, ainsi que leur structure (angle, force, nombre d'inflexions) lui sert de support pour décrire le niveau de tension des différentes régions de l'œuvre du point de vue du spectateur.

2.1.7 Impact pictural

Dans cette section nous énonçons quelques lignes directrices sur l'approche que nous allons adopter dans le reste de cette thèse compte tenu des sections précédentes. Nous entendons par *impact pictural* une partie de l'impact psychologique délivré chez le spectateur lorsqu'il regarde une œuvre. L'impact pictural est du à la combinaison du style pictural et des propriétés visuelles des sujets représentés. Ces deux types de contenu sont traduits dans une œuvre par l'artiste à l'aide d'éléments picturaux tels que des lignes, formes, couleurs et textures.

Nous pouvons *définir* l'impact pictural par opposition à l'impact psychologique dû au contenu descriptif, c'est à dire à ce qui est objectivement compris et interprété par le spectateur. Cette deuxième partie de l'impact psychologique sera appelée *impact descriptif*. Elle est souvent reliée à la mémoire du spectateur et ce qu'il reconnaît dans une œuvre (une pomme, un enfant qui pleure, l'animal favori du spectateur...).

L'art abstrait provoque un impact psychologique qui est purement pictural. “*In an abstract painting, ideas, emotions, and visual sensations are communicated solely through lines, shapes, colours, and textures that have no representational significance*” [Owen, 2007]. Robinson propose une expérience imaginative où un extra-terrestre nous amènerait une de ces œuvres [Robinson, 1981]. Nous la regarderions probablement comme une œuvre abstraite si cette image ne nous évoque rien de reconnaissable, alors que pour l'extra-terrestre, cette œuvre aurait sûrement aussi un impact descriptif.

Inversement, il est presque impossible d'imaginer une œuvre qui provoquerait un impact psychologique qui soit uniquement descriptif. Si cela existait, ce devrait être une œuvre qui transmette uniquement l'idée d'un objet ou d'une scène en laissant au spectateur l'entière tâche d'en imaginer la représentation visuelle. Un tableau au fond homogène avec pour seul motif la phrase *un enfant pleure* s'approcherait de cet objectif. Néanmoins, le format, la manière d'écrire cette phrase, la couleur du fond et celle de l'écriture auraient des influences sur l'effet produit (voir Figure 2.4). Si nous essayons de normaliser ces paramètres pour être le plus neutre possible, nous en venons à finalement à considérer l'équivalent d'un texte imprimé.



FIG. 2.4: Vers un impact purement descriptif?. De gauche à droite, et de haut en bas, un tableau de l'artiste Ben, Magritte et deux images synthétiques.

2.1.8 Conclusion

L'impact psychologique total produit par une œuvre est souvent appelé *impact esthétique* [Silverman, 1984, Eakins et al., 2004]. Dans cette thèse nous faisons l'hypothèse que cet impact est dû à la combinaison d'un impact pictural (que nous pouvons isoler dans une œuvre abstraite par exemple) et d'un impact descriptif. Cet impact est idiosyncratique, c'est à dire qu'il varie d'un spectateur à un autre.

Nous pensons que l'impact descriptif peut être très difficilement analysé ou extrait automatiquement dans une œuvre car cela supposerait que nous soyons capable d'analyser sémantiquement la scène décrite dans l'œuvre. Cette scène est très souvent stylisé à travers le regard et la volonté de l'artiste. Comme les éléments stylistiques et ceux liés à aux objets décrits se traduisent par des primitives communes, il est particulièrement difficile de les dissocier.

L'écart sémantique (*semantic gap* en anglais) [Lew et al., 2006] en traitement d'images naturelles désigne l'écart entre les primitives bas niveaux que les méthodes sont capables d'analyser (par exemple une image présentant un ovale rose avec quelques discontinuités sombres...), et les concepts hauts niveaux représentés dans l'image (cette image est un portrait d'une personne connue). Cet écart sémantique occupe très activement la communauté de recherche en traitement d'images depuis environ 10 ans, et les solutions sont encore très loin d'être satisfaisantes. Dans le cas des œuvres d'art, combler l'écart sémantique est beaucoup plus difficile dès que les œuvres sont stylisées, (Figure 2.5).

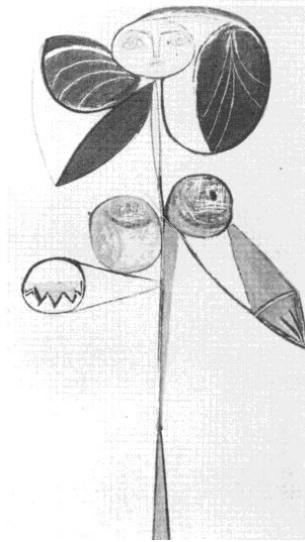


FIG. 2.5: Ecart sémantique. Dessin de Picasso dont le titre est *La femme-fleur*, qui est aussi un portrait de Françoise Girod. Ce titre ambigu illustre la difficulté à extraire le contenu descriptif d’une œuvre stylisée. Les saisons d’Arcimboldo sont aussi un exemple connu d’ambiguïté picturale.

2.2 Analyse automatique du contenu artistique

Dans cette section, nous présentons un état de l’art des méthodes automatiques proposées pour analyser le contenu artistique global d’une œuvre. Même si notre objectif est de se limiter au contenu pictural (voir Section 2.1.8), cette étude permet un tour d’horizon des méthodes et angles d’approches déjà réalisés. Pour nous permettre d’organiser cet état de l’art, nous utiliserons une taxonomie du contenu d’une image (Section 2.2.1). Nous isolerons tout d’abord les travaux qui ont proposés des méthodes d’authentification d’œuvres (Section 2.2.2) car ce besoin est particulier et assez éloigné du notre. Ensuite dans les Sections 2.2.3 à 2.2.11, nous explorerons les différentes catégories de la taxonomie du contenu image.

Remarque : En 2002 et 2005, le groupe de recherche DELOS/NSF publia deux *manifestes* visant à souligner les multiples intérêts et impacts de la recherche en vision dans le domaine du patrimoine culturel [Chen et al., 2002, Chen et al., 2005] et en particulier des méthodes de description et recherche automatique dans une base de données d’images.

2.2.1 Taxonomie du contenu image et des besoins d'utilisateurs d'images

La taxonomie du contenu d'une image de [Burford et al., 2003] définit 10 catégories qui sont présentées dans le tableau 2.2. Une autre organisation possible est d'utiliser une taxonomie des besoins utilisateurs, organisées selon le type d'utilisations. La taxonomie VISOR utilise ce point de vue et définit 7 catégories [Conniss et al., 2000], présentées dans le tableau 2.3.

Catégorie	Notes
Primitives bas niveaux	contours locaux, texture, couleur, netteté
Primitives géométriques	lignes et contours d'objets fermés
Relations visuelles 2D	l'organisation spatiale des objets dans une scène 2D
Relations visuelles 3D	l'organisation spatiale dans la 3ème dimension (profondeur)
Unités sémantiques	noms d'objets de classes, spécifique (ex. <i>Mona Lisa</i>) ou généraux (ex. <i>cheval</i>)
Abstraction culturelle	information dérivée d'une connaissance culturelle (ex. contenu iconographique)
Abstraction émotionnelle	impact émotionnel dérivé (ex. froid, chaud, tension ou calme)
Abstraction technique	information dérivée d'une expertise technique (ex. tumeur cancéreuse sur une IRM)
Abstraction contextuelle	information dérivée d'une certaine connaissance de l'environnement (ex. reconnaître une scène de jour ou de nuit)
Métadonnées	information externe à l'image (ex. taille, date de création, médium, auteur)

TAB. 2.2: Taxonomie du contenu d'une image de [Burford et al., 2003]

Dans l'état de l'art de l'étude automatique du contenu artistique, nous pouvons distinguer trois grandes applications : l'authentification (ou identification de l'auteur), la recherche d'images, et la classification de bases de données. Ces trois applications mettent souvent en œuvre des techniques similaires d'analyse du contenu. Nous voyons que la taxonomie VISOR nécessiterait de segmenter la recherche d'images en sous-catégories qui sont rarement explicités dans les références du domaine.

L'authentification d'œuvres est une application qui demande une réponse à un problème bien posé : cette œuvre est-elle oui ou non de tel artiste. C'est pourquoi nous traiterons cette problématique dans une section isolée. Les méthodes proposées s'appliquent souvent à un seul artiste (ex. Pollock, Van Gogh) et utilisent des descripteurs spécifiques au travail de chacun d'eux. Ces descripteurs sont néanmoins intéressants d'un point de vue général car il reflète une capacité à cibler un aspect très personnel chez un artiste (comme en authentification calligraphique). Le reste de l'état de l'art sera articulé selon la taxonomie de Burford. Ceci nous permettra d'organiser les références selon des thèmes assez familiers pour la communauté de l'analyse automatique d'images.

2.2.2 Authentification d'une œuvre ou identification d'un artiste

L'authentification d'une œuvre peut être vue comme une classification binaire selon l'auteur attribué : *vrai* ou *faux*. Cette approche s'applique au problème de la contrefaçon,

2.2. Analyse automatique du contenu artistique

Catégorie	Notes
Illustration	les images sont utilisées à des fins d'illustration en conjonction avec d'autres formes de média (ex. journaux)
Information Processing	les informations contenues dans l'image sont destinées à être analysées et interprétées (ex. IRM médicales)
Information dissemination	les images sont destinées à être transmises à d'autres personnes
Learning	les images permettent un apprentissage académique (ex. œuvres d'art en histoire de l'art)
Generation of ideas	les images sont utilisées pour susciter une certaine inspiration ou création
Aesthetic value	les images sont utilisées à des fins décoratives
Emotive	les images sont utilisées pour leurs forces visuelles (ex. publicité)

TAB. 2.3: Taxonomie VISOR des utilisations des images de [Conniss et al., 2000]

mais aussi plus simplement celui de l'attribution d'une œuvre. Cette seconde problématique apparaît souvent pour les peintres qui dirigeaient un atelier où de nombreux élèves peignaient selon le style du maître une partie ou la totalité d'une œuvre. Ces problématiques sont d'une grande importance sur le marché de l'art car leurs enjeux fixent les valeurs monétaires d'une œuvre avec parfois des variations extrêmes. Depuis une quinzaine d'années, la recherche en analyse de peinture par ordinateur s'est intéressée à ces problématiques non sans provoquer quelques réticences de la part des experts.

Depuis 1999, Taylor et al. appliquent une analyse fractale du contenu géométrique de peintures *dripping* de Jackson Pollock [Taylor et al., 1999, Taylor et al., 2007]. Ces œuvres réalisées par écoulements de peintures à l'aide d'un pinceau ou d'un bâton à la verticale d'une toile posée à terre représente la période caractéristique du style de Pollock. Une approche fractale analyse les répétitions multi-échelles et récursives d'un motif géométrique [Mandelbrot, 1982]. L'approche fractale de Taylor et al. est appliquée aux images après une binarisation. Cette méthode a suscité plusieurs controverses quant à sa validité jusqu'à aujourd'hui [Jones-Smith and Mathur, 2006, Taylor et al., 2006]. Elle étudie seulement deux dimensions fractales sur chaque image, l'ensemble de tests est réduit à seulement 14 peintures, et elle n'a pas été appliquée à d'autres peintres. Depuis, ce manque a été partiellement comblé par [Mureika et al., 2005], et un nombre d'œuvres potentielles de Pollock plus élevé [Taylor et al., 2007]. Dans cette dernière référence 14 œuvres récemment apparues sur le marché ont été analysées. Ces œuvres sont encore en attente d'authentification par des experts. La méthode de Taylor les a classées comme étant fausses.

[Mureika et al., 2005] ont montré que la méthode de Taylor n'est pas efficace pour distinguer des peintures de Pollock avec d'autres peintures de la même période (expressionnisme) tels que celles du Canadien Jean Paul Riopelle. Par contre, lorsque la même méthode est appliquée non plus aux images binaires, mais à la carte des contours (obtenue avec un filtre de Sobel), les différentes classes sont bien séparées. Les auteurs y voient une confirmation du fait que les contours ont une grande importance en perception comme le soulignèrent [Hubel and Wiesel, 1962].

Une méthode d'authentification de dessins de Delacroix à base de calculs de ratios sur

des sous parties de l'image binaire de tailles 20×20 a été proposée par [Kroner and Lattner, 1998]. Les ratios calculent le pourcentage de pixels noirs par rapport au nombre de pixels blancs, normalisé par le ratio sur l'image complète. Les distributions des directions des traits dans le dessin sont aussi calculées. Une combinaison des histogrammes des ratios ainsi que des directions décrit les dessins. Sur une base de 41 dessins dont 19 de Delacroix et 22 d'autres artistes, la méthode permet d'authentifier les dessins de Delacroix avec un taux moyen de 87% par validation croisée de 25 dessins parmi les 41.

Un projet de recherche débuté en 2000 à Amsterdam appelé Authentic a pour objectif l'authentification automatique d'œuvres de Van Gogh [van den Herik and Postma, 2004]. Dans le cadre de ce projet, plusieurs travaux ont été publiés, s'intéressant à la caractérisation du style personnel de Van Gogh. Une première étude utilise et compare différents descripteurs : des histogrammes de couleurs, le spectre de Fourier, l'orientation spatiale obtenue après filtrage Gaussien, des statistiques sur les niveaux de gris, un indice fractal, les composantes indépendantes, et enfin un réseau de neurones comme classifieur [van den Herik and Postma, 2000]. L'objectif de cet article est de trouver les meilleurs descripteurs permettant de dresser la *signature picturale* d'un artiste, par analogie à la graphologie. Leur conclusion est que les descripteurs les plus discriminatoires sont l'indice fractal, le spectre de Fourier, la moyenne des niveaux de gris et l'histogramme de couleur global calculé sur l'image. La classification dans le cadre de cette étude est faite par artiste. Il semble surprenant dans ce cadre que la moyenne des niveaux de gris soit un descripteur véritablement robuste pour reconnaître un artiste par exemple. Par ailleurs la base de test pour cette étude est constituée de 6 classes de 10 œuvres de 6 artistes, ce qui est assez faible pour pouvoir conclure, dans la mesure aussi où la majorité des descripteurs sont des indices statistiques.

Dans une autre publication du même projet, [Berezhnoy et al., 2005] présentent une étude statistique portant sur la forme des coups de pinceaux dans l'ensemble du corpus de Van Gogh. Van Gogh offre un support idéal pour cette étude car ses coups de pinceaux sont en général très apparents sur ses œuvres. Chaque trait de pinceau est extrait par un filtre circulaire, puis modélisé par un polynôme du troisième degré. Les auteurs visualisent ensuite la répartition des coefficients d'ordres 2 et 3 dans un espace à deux dimensions. Ils observent que les quelques 65000 coups de pinceaux extraits sur 169 tableaux se regroupent sous la forme d'un nuage de type gaussienne à deux dimensions. Ceci reflète selon les auteurs une certaine homogénéité dans la forme des coups de pinceaux de Van Gogh. Ils concluent que cela témoigne de la signature picturale de Van Gogh, utilisable pour l'authentification d'œuvres, ce qui est discutable puisqu'aucune comparaison n'est faite avec un autre artiste, ni même avec des faux tableaux de Van Gogh.

Toujours dans le projet Authentic, Berezhnoy et al. font une étude exhaustive des distributions des couleurs complémentaires dans un corpus presque complet de 145 tableaux de Van Gogh [Berezhnoy et al., 2004, Berezhnoy et al., 2007].

Plus récemment, [Lyu et al., 2004] proposent une approche texturale utilisant une décomposition en ondelettes. Leur méthode après avoir été entraînée sur un corpus suffisant,

2.2. Analyse automatique du contenu artistique

permettrait non seulement d'authentifier mais aussi de décomposer les parties de l'œuvre qui aurait été réalisées par un apprenti et non par le maître lui même.

Auteur	Artiste étudié	Techniques utilisées
[Taylor et al., 2007]	Peintures de Pollock	Dimensions fractales sur images binaires
[Mureika et al., 2005]	Peintures de Pollock	Dimensions multi-fractales sur cartes de contours
[Kroner and Lattner, 1998]	Dessins de Delacroix	Histogrammes augmentées sur image partitionnée
[van den Herik and Postma, 2000]	Peintures de Van Gogh	Ensemble de descripteurs locaux (Fourier, histogrammes couleurs, indice fractale, ...)
[Berezhnoy et al., 2005]	Peintures de Van Gogh	Statistiques sur la forme des coups de pinceaux, modélisation polynomiale
[Berezhnoy et al., 2007]	Peintures de Van Gogh	Distributions de couleurs complémentaires
[Lyu et al., 2004]	Bruegel le vieux	Analyse multi-échelles en ondelettes

TAB. 2.4: Récapitulatif des méthodes d'authentications.

2.2.3 Primitives bas-niveaux

Contenu colorimétrique

La couleur est un élément essentiel du contenu pictural d'une œuvre et c'est pourquoi de nombreuses publications ont cherché à l'étudier. [Corridoni et al., 1998] ont proposé l'utilisation des catégories colorimétriques de Itten pour enrichir sémantiquement le contenu analysée. Ils utilisent une segmentation K-moyennes (appelées aussi nuées dynamiques). Stanchev et al. propose sensiblement la même méthode en 2003 [Stanchev et al., 2003]. Lay et Guan propose une grammaire colorimétrique très similaire aussi, inspiré de la théorie des couleurs opposées de Hering, des sept contrastes de couleurs de Itten, et des règles harmoniques de Birren [Lay and Guan, 2004]. La catégorisation des couleurs dans l'image se fait en utilisant là aussi un algorithme K-moyennes. Yelizaveta et al. propose la même grammaire mais en utilisant une segmentation Blobworld [Carson et al., 2002] pour obtenir une meilleure précision [Yelizaveta et al., 2004, Yelizaveta et al., 2005].

Texture et organisation spatiale locale

[Keren, 2002, Keren, 2003] a proposé une méthode de classification utilisant une description par transformée en cosinus discrète (DCT) et un classificateur naïf de Bayes. La méthode atteint un taux de succès de 86% sur une base 50 peintures de 5 artistes différents.

[Lewis et al., 2004] utilisent aussi des ondelettes et proposent une plateforme exploitant conjointement les méta données et le contenu des images. [Kobayasi and Muroya, 2003] utilisent des ondelettes de Haar pour mesurer les variations spatiales colorimétriques. [Kushki et al., 2004] s'intéressent plus particulièrement à un cadre de décision fusionnant les descripteurs proposés par le standard MPEG-7.

Plusieurs approches ont cherché à étudier l'organisation spatiale locale des couleurs en utilisant des descripteurs à base d'histogrammes. [Pass et al., 1996] ont proposé des vecteurs de couleurs cohérentes. Ces histogrammes couleurs intègrent pour chaque couleur (sur une palette quantifiée) la proportion de celles appartenant à une classe dite cohérente, c'est à dire appartenant à un chemin connexe d'une taille supérieur à un seuil. En 1999, Pass et Zabih utilise des histogrammes appelés histogrammes joints, qui consistent à compléter chaque couleur avec la valeur du module du gradient et un descripteur de texture [Pass and Zabih, 1999]. Huang et al. proposent d'utiliser les matrices de cooccurrence d'Haralick qu'ils appellent corrélogrammes [Huang et al., 1997, Huang et al., 1999]. Il s'agit d'un histogramme où pour chaque couleur c_i , nous calculons la probabilité d'avoir la présence d'un pixel de la même couleur (cas de l'autocorrélogramme) ou d'une couleur différente c_j (cas général) à une distance pixelique k . En pratique la distance k ne dépasse pas 7 pixels sur des images 200×200 . [Williams and Yoon, 2007] réutilise les autocorrélogrammes en y incluant des informations locales de gradient.

Auteur	Techniques utilisées
[Corridoni et al., 1998]	Segmentation K-moyennes de l'espace colorimétrique
[Stanchev et al., 2003]	Segmentation K-moyennes de l'espace colorimétrique
[Lay and Guan, 2004]	Segmentation K-moyennes de l'espace colorimétrique
[Yelizaveta et al., 2005, Yelizaveta et al., 2004]	Segmentation BlobWorld de l'espace colorimétrique
[Keren, 2002, Keren, 2003]	Analyse transformée DCT
[Lewis et al., 2004]	Analyse en ondelettes
[Kobayasi and Muroya, 2003]	Analyse en ondelettes
[Kushki et al., 2004]	Descripteurs MPEG-7
[Pass et al., 1996]	Vecteurs de couleurs cohérentes
[Pass and Zabih, 1999]	Histogrammes joints
[Huang et al., 1997, Huang et al., 1999]	Corrélogrammes de couleurs
[Williams and Yoon, 2007]	Autocorrélogrammes + gradients

TAB. 2.5: Récapitulatif des méthodes reposant sur des primitives bas-niveaux.

2.2.4 Primitives géométriques

La méthode de Berezhnoy et al. déjà mentionnée pour l'authentification de peintures de Van Gogh fait partie également de la catégorie des méthodes utilisant les primitives géométriques [Berezhnoy et al., 2005]. Elle utilise une modélisation polynomiale des traits de pinceaux, puis des statistiques sur les coefficients de ces polynômes.

Rémi et al. ont proposé un outil d'analyse automatique de la structure et du tracé de dessins d'enfants en ligne [Remi et al., 2002]. Le but d'un tel outil est de réussir à détecter les enfants sujets à des difficultés motrices neurobiomécaniques, pour ensuite les prendre en charge suffisamment tôt au moment de l'apprentissage de l'écriture. Les dessins sont des copies de deux modèles simples composés de traits rectilignes et d'un cercle. La détection se fait par transformée de Hough, et un critère de linéarité. La structure de ces éléments ajoutés à plusieurs descripteurs bas niveaux (hauteur, largeur, surface, moyenne et écart type de la vitesse du trait de stylet, nombre et durées des pauses) permettent une classification selon un certain niveau d'aisance à l'écriture de l'élève.

2.2. Analyse automatique du contenu artistique

[Kampel and Sablatnig, 2007] ont proposé une méthode de classification d’images de poteries selon une étude géométrique de leur contour. Leur description intègre la détection et la caractérisation des points d’intérêts le long du contour tels que points d’inflexions, extremums par rapport à l’axe principal des poteries, points de fortes courbures etc.

Widjaja a proposé en 2003 une étude sur les dégradés transversaux de la peau le long des visages ou des mains dans les portraits [Widjaja et al., 2003]. Cette étude est de par sa définition extrêmement limitative. Mais elle renseigne sur certains aspects parfois inattendus, comme les dégradés, qui peuvent témoigner du style de l’artiste.

Auteur	Techniques utilisées
[Berezhnoy et al., 2005]	Modélisation polynomiale des traits de pinceaux
[Remi et al., 2002]	Détection des traits par transformée de Hough
[Kampel and Sablatnig, 2007]	Analyse géométrique des contours de poteries
[Widjaja et al., 2003]	Analyse des profils transversaux de la peau dans des portraits

TAB. 2.6: Récapitulatif des méthodes reposant sur les primitives géométriques.

Nous proposerons dans la seconde partie de cette thèse une méthode d’analyse reposant également sur les primitives géométriques. Adaptée au dessin, cette méthode détecte les traits curvilignes tracés par l’artiste. L’analyse des points d’intérêts géométriques extraits à partir de ces traits nous permet de proposer une description du contenu pictural.

2.2.5 Relations visuelles 2D globales

Dans cette section, les relations visuelles 2D globales correspondent à l’organisation spatiale des régions dans une image. Dans le cadre des arts visuels 2D, cette organisation est très proche du concept de composition artistique picturale. Cette composition est le témoin de l’agencement des régions homogènes réalisé durant l’élaboration de l’œuvre. Cet aspect est également lié aux notions d’équilibre visuel étudié par [Arnheim, 1954].

Nous proposerons dans les deux chapitres suivants une méthode étudiant les relations visuelles 2D globales. C’est pourquoi nous détaillerons un peu plus certaines des méthodes présentées dans cette section.

Tanaka et al. étudient en 1999 la composition géométrique des tableaux [Tanaka et al., 1999, Tanaka et al., 2000]. Leur outil extrait la proportion et la localisation des différentes formes. Ces formes sont tout d’abord issues d’une segmentation par flux de contours [Ma and Manjunath, 1997]. La segmentation est suivie d’une étape de sélection automatique des régions en utilisant des critères de contraste, texture et couleur. L’organisation spatiale des figures ainsi retenues est ensuite analysée selon des critères d’équilibre et de positionnement. Cet outil est utilisé comme outil d’aide à la création de support multimédia attrayants.

Corridoni et al. ont étudié en 1996 et 1997 les relations visuelles liées aux couleurs [Corridoni et al., 1996, Corridoni et al., 1997]. Ils proposent une grammaire faisant le lien entre l'ensemble chromatique bas niveau et quelques concepts hauts niveaux comme la température, l'accordance et l'harmonie, fondée sur la sémantique développée par Itten en 1961 [Itten, 1961]. Ils étudient aussi l'organisation spatiale des régions colorées après une segmentation K-moyennes dans l'espace CIE-Luv des couleurs. Le plus gros inconvénient de ce type d'approche est qu'en créant un pont entre les attributs bas niveaux et quelques concepts hauts niveaux comme les couleurs chaudes et froides ou le clair-obscur, les requêtes ont un champ d'application limité à la grammaire développée.

IBM a conçu au début des années 1990 un système appelé QBIC⁴ utilisant la distribution des couleurs indépendamment de leur position spatiale et quelques informations sur la texture de l'image [Flickner et al., 1995]. Il fut amélioré en 1997 en intégrant la position des pixels sur un partitionnement de l'image en 108 régions [Holt et al., 1997, Bird et al., 1999]. Il est utilisé encore aujourd'hui au musée de l'Hermitage à Saint-Petersbourg.

[Del Bimbo et al., 1998] ont proposé le système PICASSO. Ce système intègre une description spatiale globale des couleurs utilisant une segmentation multi-niveaux de l'image en régions colorées. Cette représentation multi-niveaux permet de proposer ensuite à l'utilisateur plusieurs échelles de descriptions de l'image selon son intérêt de recherche. Les régions segmentées à chaque niveau sont regroupées dans un graphe $\mathcal{G}$, où chaque région R_i est décrite par sa surface relative, sa couleur moyenne, la position globale de son centroïde, les 13 moments centraux, et un vecteur $[c_1^{R_i}, \dots, c_{128}^{R_i}]$ où $c_j^{R_i}$ vaut 1 si la j^{eme} couleur est présente dans la région ou 0 sinon (les couleurs sont exprimées dans l'espace CIE-Luv quantifié sur 128 couleurs). La distance utilisée pour comparer deux images se fonde sur le coût minimal entre les deux graphes pour comparer chaque région requête à l'ensemble des régions candidates, à l'aide d'une distance pondérée entre les différents descripteurs de chaque région.

Dans d'autres travaux Del Bimbo et Vicario ont également proposé une mesure de similarité de l'agencement spatiale relatif de deux ensembles de pixels non nécessairement connexes [Del Bimbo and Vicario, 1998, Berretti et al., 1999, Berretti et al., 2003]. Pour deux points a et b il y a neuf directions possibles (appelées *walkthroughs*) représentées par les couples $\langle i, j \rangle$ avec $i, j \in \{-1, 0, 1\}$ pour aller du point a au point b . La distance entre deux ensembles de points A et B proposée par Del Bimbo et Vicario utilise tout d'abord le calcul des poids $w_{i,j}(A, B)$ avec $i, j \in \{-1, 0, 1\}$. Ces poids sont les intégrales sur l'ensemble des points de A et B :

$$w_{i,j}(A, B) = K_{ij}(A, B) \int_A \int_B C_i(x_b - x_a) C_j(y_b - y_a) dx_b dy_b dx_a dy_a \quad (2.1)$$

où $C_i, i \in \{-1, 0, 1\}$ est la fonction caractéristique de l'ensemble $\mathbb{R}^{-*}, \{0\}, \mathbb{R}^{+*}$ respectivement. Les facteurs $K_{ij}(A, B)$ sont des facteurs de normalisation. Les auteurs construisent ensuite une distance métrique $\mathcal{D}_S(w, w')$ permettant de comparer l'organisation spatiale

⁴Query By Image Content

2.2. Analyse automatique du contenu artistique

de deux paires d'ensembles de points. Ils montrent que cette distance est plus représentative vis à vis de la perception de l'organisation spatiale entre objets que des distances utilisant les centroïdes des régions par exemple. Cette distance a été ensuite utilisée au sein d'un système générique de recherche d'images [Berretti et al., 2002, Berretti et al., 2003]. Chaque image dans ce système est modélisée par un graphe relationnel où chaque noeud représente une région de l'image non nécessairement connexe. Chaque noeud q_i peut être associé à plusieurs attributs comme la couleur moyenne ou des indices de textures. Ce système utilise une distance globale entre deux images Q et D représentées par leurs graphes relationnels [Berretti et al., 2000]. Les auteurs illustrent les performances de $\mu^{\Gamma^*}(Q, D)$ par deux applications. La première application utilise une segmentation manuelle de l'image, et la deuxième utilise la méthode de segmentation proposée dans le système PICASSO présenté dans le paragraphe précédent. Cette méthodologie permet principalement de comparer l'agencement **relatif** des régions entre elles, ce qui est différent d'une étude de l'organisation spatiale globale des régions.

De même, [Onkarappa and Guru, 2007] ont étudié les relations visuelles relatives des primitives géométriques dans les dessins et schéma à base de traits. L'objectif visé est de permettre la recherche automatique de dessins similaires selon ce critère. Ils proposent une modification des matrices 9DLT [Chang, 1991]. Ces matrices traduisent de manière symbolique l'agencement spatial relatif d'éléments présents dans une image.

Auteur	Techniques utilisées
[Tanaka et al., 1999, Tanaka et al., 2000] [Holt et al., 1997, Bird et al., 1999] [Del Bimbo et al., 1998] [Berretti et al., 2003]	Segmentation de l'image par flux de contours QBIC : distance pixel à pixel d'une image sous-échantillonnée Segmentation pyramidale par K-moyennes Distance <i>walkthrough</i> sur une segmentation manuelle ou automatique de l'image
[Onkarappa and Guru, 2007]	Matrice 9DLT modifiée pour l'étude de l'agencement relatif de primitives 1D dans un dessin

TAB. 2.7: Récapitulatif des méthodes ciblant les relations visuelles 2D globales.

2.2.6 Relations visuelles 3D

Tonder proposa en 2007 une méthode de description automatique du niveau de perspective et de la composition spatiale dans les tableaux de jardins japonais [van Tonder, 2007]. Il montre que ces aspects descriptifs permettent de distinguer différents paysagistes japonais classiques.

Criminisi et Stork ont exploré la perspective dans plusieurs tableaux, en particulier de certains maîtres de la Renaissance. Ceci les a conduit à revendiquer l'idée que ces maîtres auraient utilisé l'aide d'un système optique à base de lentilles grossissantes pour réaliser leurs peintures [Criminisi and Stork, 2004, Stork, 2006].

2.2.7 Unités sémantiques

Les unités sémantiques selon la taxonomie de Burford (Tableau 2.2) correspondent au contenu descriptif d'une image et donc souffrent de l'écart sémantique entre descripteurs bas niveaux et concepts descriptifs haut niveaux. Une des rares publications appliquées au domaine des arts visuels que nous avons trouvé et celle de [Valle et al., 2006]. Ceux-ci proposent une méthode de recherche d'images culturelles utilisant la détection de points d'intérêt SIFT. Leur méthode est appliquée aux images en niveaux de gris et est particulièrement adaptée à la recherche de doublons d'images ayant subies des transformations (déformation, bruit, changement de la netteté, compression etc).

2.2.8 Abstraction culturelle, style

La catégorie de l'abstraction culturelle selon la taxonomie de Burford (Tableau 2.2) correspond aux informations dérivées d'une connaissance culturelle présente dans l'œuvre. C'est pourquoi nous plaçons dans cette section tous les travaux ayant pour but d'analyser et reconnaître le style d'une œuvre, d'un artiste ou d'un courant artistique.

[Pareti and Vincent, 2005, Ogier, 2006] ont proposé dans le cadre du projet Madonne une méthode de reconnaissance de style de lettres graphique. Elle utilise les contrastes locaux et une loi en puissance de type Zipf [Zipf, 1949].

[Nack et al., 2002] proposent une grammaire faisant le lien entre quelques attributs bas niveaux simples comme le contraste, les couleurs, et l'histogramme d'intensité et des courants artistiques connus. Ils définissent trois styles (le clair obscur, le cubisme, et l'impressionnisme) à partir des attributs. Les attributs bas niveaux utilisés permettent des résultats satisfaisants uniquement avec le clair-obscur comme le concluent les auteurs. Le même type d'approche et de limitations se retrouvent dans [Yelizaveta et al., 2004].

Seldin et al. ont proposé en 2003 une méthode de classification non supervisée par segmentation conjointe, qu'ils appliquent à la classification de quelques styles picturaux [Seldin et al., 2003]. Il s'agit d'une approche texturale, par ondelettes. La segmentation est par ailleurs menée conjointement sur l'ensemble de la base pour établir un dictionnaire de texture commun à toutes les images. Cela est un inconvénient, car cette segmentation est coûteuse. De plus chaque ajout d'image dans la base nécessite que l'ensemble des segmentations soit recalculé. Les résultats permettent de bien classer une petite base de 35 peintures réalisées par 5 peintres aux styles très différents (Picasso période cubiste, Van Gogh, Rembrandt, Shishkin et Aivazovsky). De part sa complexité cette méthode est inapplicable à de grandes bases de données. Malgré tout, ce type d'approche, à base d'outils de décomposition multi échelles comme des ondelettes ou des filtres de Gabor est efficace pour caractériser la texture de la touche picturale.

La classification de dessins à l'encre chinoise a été abordée par plusieurs chercheurs.

2.2. Analyse automatique du contenu artistique

[Zhang et al., 2004] ont étudié quelques techniques classiques pour modéliser des dessins. [Li and Wang, 2004] proposent une méthode d'analyse et d'indexation utilisant une décomposition multi-échelles par ondelettes et des mélanges de modèles de Markov cachés pour étudier des dessins à l'encre noire. [Jiang et al., 2006] utilisent des vecteurs de cohérence couleurs, un histogramme sur la taille des contours (extrait avec un détecteur de Sobel) et un classifieur SVM. [Pham, 2005] a proposé une méthode mixte utilisant des classes iconographiques prédéfinies, et une description automatique du contenu couleur et des formes par modèles déformables.

Lombardi a étudié en 2005 dans sa thèse un système d'indexation selon le style utilisant un mélange de plusieurs descripteurs bas-niveaux classiques (histogrammes couleurs et luminance, gradients, indices de texture) [Lombardi, 2005, Lombardi et al., 2004, Lombardi et al., 2005].

[Icoglu et al., 2004, Gunsell et al., 2005] ont proposé une méthode d'indexation selon le style à l'aide d'un descripteur à six dimensions. Celui ci contient des informations sur le ratio de pixels noirs sur l'image binarisée, la moyenne du gradient, le nombre d'extremums sur l'histogramme de luminance, la largeur de l'histogramme des couleurs, et quelques mesures statistiques sur l'image en niveau de gris.

[Yan and Jin, 2006] proposent une méthode basée uniquement sur la couleur, mais avec pour objectif d'atteindre spécifiquement le contenu stylistique. Ils définissent six styles différents, chacun étant décrit par un vecteur à sept dimensions dont les éléments mesurent les proportions de noir, blanc, gris, couleur, teinte, ombre, et tonalité.

Auteur	Techniques utilisées
[Pareti and Vincent, 2005, Ogier, 2006] [Nack et al., 2002]	Contrastes locaux et lois en puissance de Zipf Contrastes, couleurs et histogrammes + catégorisation en courants artistiques
[Yelizaveta et al., 2004]	Contrastes, couleurs et histogrammes + catégorisation en courants artistiques
[Seldin et al., 2003]	Méthode de segmentation + catégorisation en courants artistiques
[Zhang et al., 2004] [Li and Wang, 2004] [Jiang et al., 2006]	Histogrammes appliqués à l'étude des dessins chinois Ondelettes appliqués à l'étude des dessins chinois Vecteurs de cohérences couleurs + SVM appliqués à l'étude des dessins chinois
[Pham, 2005]	Segmentation par modèles déformables appliqués à l'étude des dessins chinois
[Lombardi et al., 2004]	Histogrammes couleurs, contrastes, textures + classificateur K-moyennes
[Icoglu et al., 2004, Gunsell et al., 2005] [Yan and Jin, 2006]	Ratios de pixels noirs, gradient, histogramme d'intensités Vecteurs de mesures colorimétriques à six dimensions

TAB. 2.8: Récapitulatif des méthodes étudiant l'abstraction culturelle

2.2.9 Abstraction émotionnelle

A notre connaissance, un des rares travaux s'étant intéressé à l'impact émotionnel dérivé d'une œuvre et celui de [Wei-ning et al., 2004]. Ces auteurs étudient la *dynamique* des dessins en proposant deux catégories *statique* ou *dynamique*. Leur approche assez succincte utilise la forme de l'histogramme des orientations des traits dans le dessin. Si l'histogramme reflète la présence de nombreux contours horizontaux, l'œuvre est classée comme statique. Au contraire si l'histogramme contient une majorité de contours verticaux, l'œuvre est classée comme dynamique.

Nous pouvons aussi citer dans cette section les travaux déjà évoqués sur la couleurs utilisant un espace colorimétrique catégorisé avec des classes de couleurs chaudes et froides [Lay and Guan, 2004, Yelizaveta et al., 2005, Yelizaveta et al., 2004] vu à la Section 2.2.3.

Les travaux présentés dans la seconde partie de cette thèse pourraient également être placés dans cette catégorie puisque nous cherchons à décrire l'impact pictural délivré chez le spectateur.

2.2.10 Abstraction technique

Le domaine de l'analyse et de la reconnaissance automatique des types d'outils (crayon, fusain etc) a été abordé par Kammerer et al. sur des images en lumière visible ou infrarouge [Kammerer et al., 2003a, Kammerer et al., 2003b, Lettner et al., 2004, Lettner and Sablatnig, 2005, Kammerer et al., 2007]. L'outil développé vise à aider le conservateur en arts dans l'étude des outils utilisés sur différents type de supports et médium. Pour extraire les traits de dessins, une combinaison de filtre de Canny et de contours actifs est utilisée. Soulignons par ailleurs qu'il s'agit d'un outil semi-automatique, et que les traits extraits sont manuellement choisis isolés et rectilignes. L'analyse d'image subséquente met l'accent sur la texture des outils caractérisée par les variations des contours actifs extraits et l'utilisation conjointe de lumière visible et infrarouge.

2.2.11 Métadonnées

Les métadonnées sont généralement utilisées pour indexer et chercher des images dans des bases de données. Pour des articles de synthèse résumant l'état de l'art concernant les problématiques actuelles, enjeux principaux, défis et perspectives dans le milieu de l'indexation d'images en général, nous pouvons nous référer aux synthèses récentes de [Lew et al., 2006] et [Liu et al., 2007] ou encore à la monographie de [Del Bimbo, 1999].

Dans le domaine culturel en ligne en particulier, deux études de synthèse ont été publiées en 2004 et 2007 sur les méthodes de recherche d'images par métadonnées et par

le contenu [Mattison, 2004, Tsai, 2007]. Comme les approches par métadonnées sont très éloignées des méthodes d'analyse du contenu nous ne détaillerons pas plus ces approches. La principale contrainte réside dans le fait qu'il faut enrichir manuellement les métadonnées pour chaque œuvre. La seconde limite est que ces méthodes s'appliquent mal aux descriptions non factuelles, ou subjectives (comme l'abstraction culturelle, émotionnelle, les relations visuelles ...). Ces catégories présentent une cohérence inter-indexeurs (personnes chargés d'indexer manuellement à l'aide de mots clés des images) très faible dans le cadre d'une approche cognitive d'indexation de base de données [David, 2004].

Quelques approches plus abouties réalisent une combinaison de méthodes par le contenu, et d'utilisation de métadonnées (ex. [Lewis et al., 2004]).

2.2.12 Conclusion

De nombreuses études ont été proposées pour prendre en compte plusieurs des aspects du contenu artistique. Plusieurs travaux cherchent à identifier l'artiste à l'origine de l'œuvre à des fins d'authentification, ou de classification de bases de données par artiste. Ces méthodes utilisent parfois le terme de *style* pour désigner le contenu décrit. Dans ce contexte, le style est vu comme une empreinte personnelle, parfois inconsciente, laissée par l'artiste et que nous pourrions appeler *signature picturale* par analogie avec la graphologie.

Nous voyons dans ces références qu'identifier un artiste pour classer une base hétérogène est différent d'une classification binaire, et extrêmement difficile dès que le nombre d'artistes est important ou que ceux-ci ont un style proche. Par ailleurs, le nom de l'artiste est souvent une méta-donnée qui accompagne l'œuvre.

Un autre groupe de travaux s'intéresse à certains aspects du contenu pictural comme les relations visuelles, le *style* de l'œuvre (ou du courant artistique auquel elle a été associé), l'impact émotionnel. Malheureusement ces approches se limitent souvent à quelques styles réducteurs (impressionnisme, clair-obscur etc.). Elles correspondent souvent à des méthodes de description bas-niveau des éléments caractérisant l'écriture picturale d'un artiste comme dans le cadre du projet Authentic. L'ensemble de ces travaux limitent leur étude à des bases réduites composées par exemple de 5 courants artistiques et 10 œuvres par courants. Nous remarquons aussi qu'un grand nombre de ces méthodes utilisent des descripteurs bas-niveau. Ce sont souvent des combinaisons de mesures locales sur les distributions des couleurs, gradients, et textures. Un petit groupe de méthodes s'intéresse aux relations visuelles 2D (c'est à dire la composition artistique de l'œuvre) ce qui représente un niveau conceptuel intermédiaire entre des mesures bas niveau et des concepts haut niveaux.

2.3 Rendu non-photoréaliste

Un domaine de recherche assez important dans la communauté de rendu graphique est celui du rendu non-photoréaliste (NPR). Les chercheurs de ce domaine développent des techniques de synthèse de dessins ou peintures, à partir de photographies ou de modèles 3D. Pour une introduction à ce milieu, voir la monographie de [Gooch and Gooch, 2001]. Les motivations de ce domaine de recherche sont multiples. Le dessin est un outil pédagogique. L'abstraction permet de communiquer efficacement. Les dessins restent intelligibles sur des petites surfaces d'affichage. Un contenu simple ou incomplet peut être suggéré. Et enfin, l'abstraction peut être plus esthétique que le réalisme.

Rapidement, dans un souci de *non-réalisme*, cette communauté a cherché à comprendre les différents aspects du processus artistique pour mieux le reproduire. C'est pourquoi il est intéressant d'étudier la littérature de ce domaine. L'objectif ultime de ces applications est en effet de pouvoir capturer directement le style d'une œuvre, pour le transférer ensuite à des photographies ou des objets. L'attente vis à vis de la communauté de la vision, de méthodes d'analyse automatique est par conséquent très forte. Cette attente est explicite dans de nombreuses références [Jodoin et al., 2002, P. Barla, 2005, Grabli et al., 2004] et des programmes de recherche comme le projet ARTIS de l'IMAG⁵.

Les contributions du NPR peuvent être classées en deux groupes. Un ensemble de travaux s'intéressent uniquement à reproduire le plus fidèlement possible les détails visibles et locaux d'un rendu artistique. Les résultats les plus efficaces sont souvent issus d'un modèle physique du comportement de la matière sur le support. Pour un bon aperçu de l'ensemble de ces techniques voir [Ostromoukhov, 2002]. Ces approches directes commencent à atteindre leurs limites aujourd'hui, car elles ne prennent pas en compte le processus créatif sous-jacent de l'artiste. C'est pourquoi depuis quelques années un certain nombre de travaux de recherche initiés par cette communauté ont été effectués pour comprendre comment voit, fonctionne et crée un artiste pour représenter une scène sur un support 2D [Willats and Durand, 2005]. Une session de cours Siggraph en 2002 a été dédiée à l'ouverture de plusieurs réflexions concernant ces aspects [Durand et al., 2002].

Parallèlement à ces travaux, plusieurs approches se sont efforcées d'intégrer certains des aspects du processus créatif dans les méthodologies développées pour le rendu NPR. Précisons toutefois que cette communauté utilise rarement directement des œuvres, et n'utilisent pas des outils de vision ou d'analyse d'images. Les paramètres stylistiques choisis sont soit extraits de l'artiste informaticien qui utilise un logiciel de rendu informatique [Hamel and Strothotte, 1999, Hertzmann et al., 2002], soit souvent en partie inférés par un modèle 3D de l'objet que nous souhaitons représenter [Hamel and Strothotte, 1999, Grabli et al., 2004] (voir Figure 2.6), soit en attente d'un système d'analyse automatique pour les extraire à partir d'œuvres réelles telles que des dessins ou peintures [Jodoin et al., 2002, Freeman et al., 2003, P. Barla, 2005]. Le moteur de rendu développé dans

⁵Institut d'informatique et mathématiques appliquées de Grenoble, France, <http://artis.imag.fr/Research/NPR/>



FIG. 2.6: Génération automatique de dessins stylisés [Grabli et al., 2004]

Ces dessins produits à partir d'un modèle 3D. Dans cette méthode, Grabli et al. ont développé un système procédural et multi-couche de génération fondé sur les systèmes de [Willats, 1997] présentés à la Section 2.1.4.

ces publications devient plus structuré et il tend à modéliser de plus en plus le processus créatif réel. Une des rares publications proposant une méthode d'extraction du style d'une œuvre réelle dans le but de le transférer à un autre est celle de Bae et Durand [Bae and Durand, 2004]. Cette méthode consiste à utiliser un outil de décomposition multi-échelles à base de filtre de Gabor [Freeman and Adelson, 1991]. Les descripteurs sont les moyennes des coefficients de décomposition à petite et moyenne échelle de ces filtres. Puis les coefficients de l'image sur laquelle nous souhaitons transférer le style sont ajustés pour obtenir les mêmes statistiques. Cette approche est très similaire à celles adoptées pour l'authentification d'œuvres vues en Section 2.2.2.

La principale problématique du NPR est le rendu artistique par image de synthèse à partir de photographies ou de modèles 3D. Étant situés du côté créatif de l'œuvre les acteurs de cette communauté en sont rapidement venus à l'étude du processus créatif du peintre ou dessinateur pour le comprendre et le reproduire. Ainsi par leurs études bibliographiques, nous sommes sur la piste d'un modèle créatif qui nous permet de mieux cerner la structure des images que nous devons, de notre côté, analyser automatiquement. A plus long terme, ces deux domaines peuvent se rejoindre car le NPR souhaite aussi pouvoir analyser et extraire une modèle de contenu artistique, style ou impact visuel pour pouvoir l'insérer dans leur processus de synthèse automatique.

2.4 Conclusion générale

Dans cette thèse, nous proposons d'étudier l'impact pictural produit par une œuvre évoqué à la Section [2.1.7](#). Nous proposerons dans les deux parties de la thèse deux approches ciblant des types différents de contenu.

Dans la **Partie I** du manuscrit, nous essayons d'améliorer les approches existantes qui étudient les relations visuelles 2D. La composition artistique correspond à l'organisation spatiale globale des couleurs dans l'œuvre. Cet aspect du contenu pictural influence l'impact pictural des peintures et a l'avantage de pouvoir se traduire en termes assez bas niveaux.

Dans la **Partie II** du manuscrit, nous proposons une méthodologie pour extraire et analyser les primitives monodimensionnelles dans une œuvre. Ce type de primitives correspond aux traits tracés par l'artiste. L'étude des primitives 1D indiquée par Willats, a été rarement réalisée en vision par ordinateur (voir Section [2.2.4](#)). En revanche, elle a déjà suscité indirectement l'intérêt de la communauté de recherche en synthèse d'images comme nous l'avons vu à la Section [2.3](#).

Première partie

Organisation spatiale des couleurs

Chapitre 3

Recherche d'images selon l'organisation spatiale des couleurs



FIG. 3.1: Trois exemples de regroupements d'images selon leur organisation spatiale des couleurs.

Nous définissons dans ce chapitre l'organisation spatiale des couleurs (ci-après notée **OSC**) comme étant l'agencement spatial global des couleurs dans l'image. Trois exemples de regroupement d'images selon l'organisation spatiale des couleurs sont présentés sur la Figure 3.1. Une description intuitive et manuelle selon l'OSC de la première image en haut à gauche sur cette figure serait par exemple de dire que cette œuvre est composée d'une région rose au centre de taille moyenne par rapport à l'image, au dessus d'une petite région blanche, et le tout sur un fond brun sombre. Dans le cadre des arts visuels 2D, cette organisation est très proche du concept de composition artistique picturale. La composition d'un tableau est en effet liée à l'agencement des régions colorées tout au long du processus créatif de l'artiste. L'esthétique d'un tableau se construit notamment sur l'agencement spatial des formes et certaines de leurs caractéristiques (couleurs, texture,

contours), parfois plus que par ce qu'elles représentent sémantiquement. La majorité des artistes, comme Cézanne, n'hésitent pas à déformer volontairement la réalité pour mieux équilibrer la composition d'un tableau (voir Figure 2.2b)). La recherche d'image par un tel critère favorise aussi la sérendipité. C'est à dire la découverte fortuite d'une association d'images intéressantes dont le contenu sémantique est différent mais dont l'organisation spatiale des couleurs est très proche. Ces associations délivrent un impact pictural similaire qui précède notre compréhension objective de l'œuvre. Elles mettent également en relief des similarité de construction de l'espace picturale chez des artistes ou des époques qui peuvent être éloignés, traitant de sujets qui peuvent être différents.

Nous verrons à la Section 3.1 quelques travaux connexes à cette problématique. Pour comparer l'organisation spatiale des couleurs de deux images, nous proposons à la Section 3.2 d'utiliser une mesure de similarité reposant sur l'optimisation d'un problème de transport entre une imagerie requête et une imagerie cible.

3.1 Travaux connexes à la recherche d'images selon l'organisation spatiale des couleurs

Cette section présente des travaux connexes à l'étude de l'OSC, **non spécifiques aux arts visuels**. Les méthodes spécifiques aux œuvres d'arts ont déjà été présentées à la Section 2.2.5. L'histoire de l'art a aussi abordé cet aspect d'une œuvre à plusieurs reprises (Section 2.1). Aborder la recherche d'images par la couleur seule a été le point de départ dans les années 1990 d'une littérature extrêmement prolifique qui a progressivement intégré des informations spatiales (Section 3.1.2). La grande majorité des méthodes proposées se placent selon le schéma d'un système de recherche d'image par le contenu (Figure 3.2). Ces approches doivent tout d'abord choisir un espace colorimétrique adapté (Section 3.1.1). Différentes stratégies ont été proposées à base d'histogrammes (Section 3.1.4 et 3.1.4). Plus tard des méthodes de segmentation d'images ont été utilisées pour capturer l'organisation spatiale d'une image et indexer une base de donnée (Section 3.1.5). Modéliser l'organisation spatiale d'une image est une étape importante mais il est essentiel de pouvoir mesurer correctement la similarité entre deux modélisations. Nous verrons ainsi en détail à la Section 3.1.3 une mesure de similarité s'inspirant d'un problème de transport appelée *Earth Mover Distance* (EMD).

3.1.1 Espaces couleur

Toutes les couleurs du spectre visible peuvent être représentées par la donnée de trois valeurs, appelées valeurs tristimulus. Les expériences psychophysiques réalisées dans les années 1920 ont montré que n'importe quelle couleur vue par un observateur pouvait en effet être exprimée par la combinaison linéaire de trois couleurs primaires [Wyszecki

and Stiles, 2000]. Les valeurs tristimulus représentent l'énergie relative associée à chaque primaire par rapport à l'énergie de ces primaires lorsqu'elles sont réglées pour le blanc de référence choisi. Ce constat est lié à la trivariance visuelle qui résulte de la présence de trois types différents de cônes sur la rétine de l'œil.

Les couleurs primaires peuvent être choisies de manière arbitraire tant que ces trois couleurs sont linéairement indépendantes. Pour ne pas privilégier un système colorimétrique particulier physiquement réalisable par rapport à un autre, la Commission Internationale de l'Eclairage (CIE) a choisi trois couleurs primaires standards virtuelles et définit ainsi l'espace colorimétrique CIE 1931 XYZ. Cet espace standard est indépendant de toutes contraintes d'application matérielle. Il peut être converti en d'autres espaces colorimétriques tels que l'espace sRGB (adaptés aux écrans d'affichages), ou l'espace CMY (adapté aux imprimantes). La CIE a également proposé l'espace psychométrique CIE 1976 $L^*a^*b^*$ (plus communément appelé CIE Lab). Dans cet espace, L^* , a^* et b^* représentent la luminance, la proportion de rouge/vert et la proportion de jaune/bleu respectivement.

L'espace CIE Lab est obtenu par transformations non linéaires des primaires X, Y, Z. Il a été conçu avec l'objectif de linéariser l'espace des couleurs par rapport aux petites différences entre couleurs que nous percevons sous une illumination donnée. Le blanc de référence D65 est généralement choisi. Une métrique telle que la distance Euclidienne peut être ainsi utilisée sur cet espace tridimensionnel dans le but de mesurer perceptuellement les différences entre couleurs. C'est pour cette raison que l'espace CIE Lab est souvent utilisé en analyse d'images. Nous l'utiliserons également dans nos expériences.

3.1.2 De la couleur à l'information couleur-espace

L'indexation d'images selon la couleur seule est une approche étudiée et utilisée depuis relativement longtemps [Swain and Ballard, 1991]. Celle-ci est fondée sur le fait que nous pouvons souvent représenter les objets contenus dans une image par leur couleur prédominante (par exemple le ciel est bleu, les forêts sont vertes, le soleil est jaune ...). Inversement, les couleurs seules ne permettent pas d'inférer le contenu sémantique. Par conséquent, cette approche peut créer de nombreux faux positifs si l'objectif est celui longtemps poursuivi de l'indexation du contenu sémantique.

Il est donc historiquement apparu nécessaire d'introduire d'autres caractéristiques de l'image comme l'organisation spatiale locale des couleurs, des indices de formes, textures etc. La littérature dans ce domaine est extrêmement riche et continue de se développer. Pour un état de l'art récent, nous pouvons nous référer à celui de [Schettini et al., 2001]. Citons quelques systèmes commerciaux intégrant des informations spatiales locales : QBIC (IBM) [Flickner et al., 1995], Watson [Smith and Li, 1999] (IBM), VIRAGE [Bach et al., 1996], AMORE (NEC) [Mukherjea et al., 1999], WALRUS (Bell) [Natsev et al., 2004], IKONA (INRIA) [Boujemaa et al., 2001]. Dans le domaine universitaire citons les systèmes classiques Photobook (MIT) [Pentland et al., 2003], BlobWorld (Berkeley) [Carson et al.,

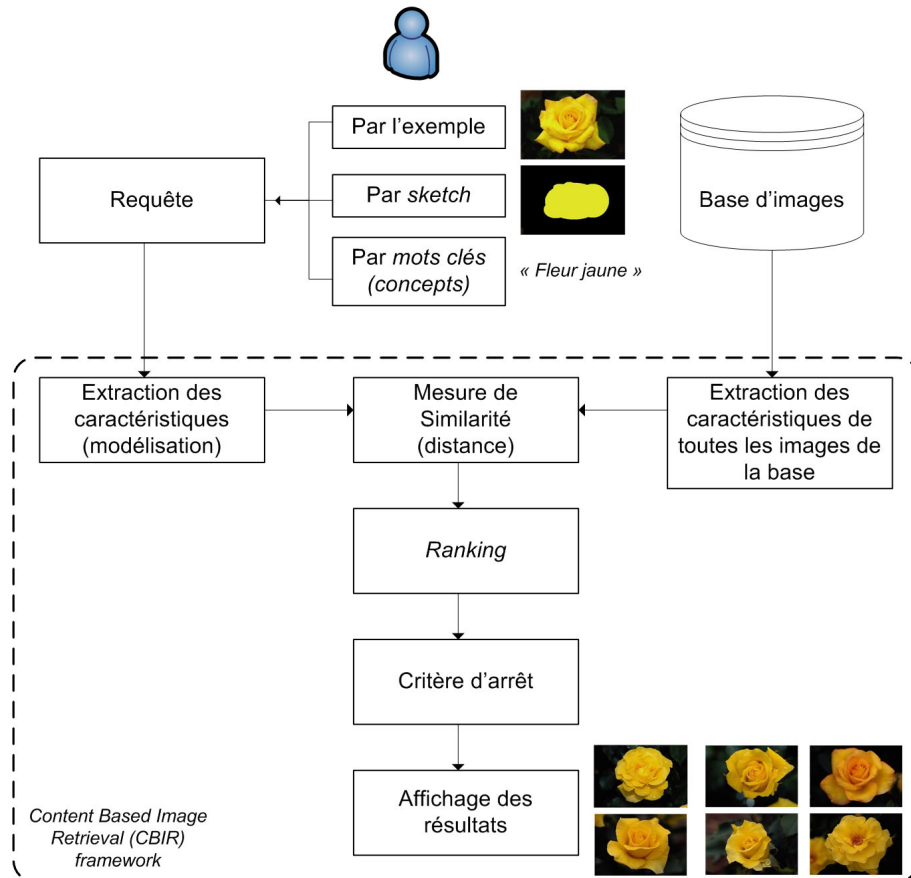


FIG. 3.2: Schéma classique d'un système CBIR. L'utilisateur peut effectuer une requête en fournissant une image exemple (*Query by Example*), en dessinant grossièrement une image (*Query by Sketch*), ou bien encore à partir de concepts prédéfinis (les couleurs, et certains objets sémantiques comme fleurs, bateau, visages ciel ...) [Lew, 2000]. Les descripteurs de cette requête sont extraits et comparés à ceux de l'ensemble de la base à l'aide d'une mesure de similarité. Cette mesure permet d'ordonner la base par distance croissante à partir de la requête (*Ranking*). Un critère d'arrêt en général supervisé renvoie un nombre limité d'images (par exemple les 10 plus proches images). Nous proposerons au Chapitre 4 un critère non supervisé.

2002], VisualSEEK et WebSEEK (Columbia) [Smith and Chang, 1997], Netra (UCSB) [Ma and Manjunath, 1999], WBIIS (Stanford) [Wang et al., 1998] et SIMPLicity (PSU) [Wang et al., 2001] et plus récemment [Chen et al., 2007].

Dans ce chapitre nous nous intéressons uniquement aux méthodes prenant en compte l'organisation spatiale **globale** des couleurs. Pour caractériser la composition d'une œuvre il est effectivement primordial de tenir compte de cette organisation (Figure 3.3). Les méthodes proposées dans le cadre de cette problématique peuvent être organisées en deux groupes. Les méthodes utilisant une quantification uniforme de l'espace colorimétrique et de l'espace de l'image peuvent être placées dans la catégorie des approches par histogrammes couleur-espace (Section 3.1.4). La deuxième catégorie de méthodes repose sur l'utilisation de méthodes de segmentation d'images (Section 3.1.5).



FIG. 3.3: Distribution spatiale de la couleur [Schettini et al., 2001] Les distributions couleurs de ces images sont identiques. Mais les organisations spatiales sont très différentes.

3.1.3 Mesures de similarité

Les méthodes que nous présenterons aux Sections 3.1.4 et 3.1.5 modélisent souvent une image sous la forme d'une distribution pondérée d'éléments (histogramme, vecteur, ensemble de régions, etc.). Pour comparer ces modèles, une mesure de similarité entre deux modèles doit être choisie. Nous présentons dans cette section différentes distances souvent présentes dans la littérature dans des approches basées sur des histogrammes ou des méthodes de segmentation d'images.

Méthodes intra-bins et inter-bins classiques

Pour comparer deux distributions, le choix de la distance est déterminant. Nous pouvons grouper les distances entre distributions en deux catégories. Les distances dites *intra-bins* comparent uniquement et directement entre eux les éléments de même position dans la distribution. Dans le cas d'un histogramme à une dimension par exemple, la distance L^2 intra-bins entre deux histogrammes est la distance quadratique cumulée entre chaque case de l'histogramme. Quelle que soit la modélisation choisie, ce type de distance est généralement très peu robuste aux changements de contraste, au pas de quantification de la distribution et au bruit dans l'image.

3.1. Travaux connexes à la recherche d'images selon l'organisation spatiale des couleurs

L'autre famille de distances est formée par les distances dites *inter-bins*. Une des premières a été proposée par [Stricker and Orengo, 1995]. Elle consiste à utiliser une distance L^p directement sur une distribution cumulative. Ce type de distance adaptée aux histogrammes 1D est plus robuste aux changements de luminosité. Cependant il produit en général des faux positifs car les différences entre deux histogrammes peuvent s'estomper au fur et à mesure de l'accumulation.

Une autre distance inter-bins et la distance quadratique proposée par [Hafner et al., 1995]. Cette méthode utilise une matrice de similarité inter-bins A pour mesurer la distance entre deux histogrammes h_i et h_j : $d(h_i, h_j) = \sqrt{h_i - h_j A(h_i - h_j)}$. Cette distance est aussi équivalente à la distance de Mahalanobis. Le choix de la matrice A est déterminant. La similarité mutuelle entre les distributions de couleurs peut être surestimée, et mener à de nombreux faux positifs.

Pour plus de détails sur ces distances classiques et d'autres non exposées ici, le lecteur peut par exemple consulter la monographie de [Del Bimbo, 1999]. Notons qu'un travail préliminaire à cette partie de la thèse a été effectué par un stagiaire en 2004 nommé Haroldo Dalazoana. Ce travail consistait à étudier et comparer un grand nombre de ces distances, combinées également à l'utilisation de différents espaces couleurs. Cette étude nous a permis de privilégier l'espace CIE-Lab, ainsi que la distance EMD que nous présentons dans la section suivante.

Distance EMD

Enfin, une distance inter-bins s'inspire d'un problème connu en optimisation sous le nom de *problème de transport*. Une manière d'illustrer le problème de transport est de considérer le problème qui consiste à déplacer un ensemble de tas de terre vers un ensemble de trous répartis dans l'espace en réalisant un minimum de travail. Cette image est proche de celle utilisée par Monge au XVIII^e siècle qui fut le premier à formaliser et étudier les problèmes de transport. Le travail réalisé correspond à la somme des masses de terre déplacées, multipliées par leurs distances de déplacement. Chaque tas de terre peut être éventuellement divisé et envoyé vers des trous différents si cela permet de minimiser le travail effectué. Ce travail optimal est une distance métrique lorsque la masse totale de la terre et la capacité totale des trous à accueillir cette masse sont égales.

[Rubner et al., 2000] ont baptisé cette distance *Earth Mover's Distance* (EMD). Il a été montré que cette distance est équivalente à la distance de Mallows entre deux distributions probabilistes [Levina and Bickel, 2001]. Sous sa forme générale, la distance EMD est vue comme le travail minimal nécessaire pour associer un ensemble de descripteurs pondérés $\mathcal{F} = \{(f_i, w_i)_{i=1, \dots, n}\}$ à un autre ensemble du même type. Le travail se calcule sur la base d'une distance au sol d_e . Les descripteurs f_i pondérés par des poids $w_i \geq 0$ appartiennent à un espace E sur lequel la métrique d_e est définie pour comparer deux descripteurs.

Mathématiquement, l'EMD entre deux ensembles $\mathcal{F}^1 = \{(f_i^1, w_i^1)\}_{i=1, \dots, n_1}$ et $\mathcal{F}^2 = \{(f_i^2, w_i^2)\}_{i=1, \dots, n_2}$ ayant le même poids total¹ $\sum_i w_i^1 = \sum_j w_j^2$ est définie par :

$$d(\mathcal{F}^1, \mathcal{F}^2) = \min_{x_{ij}} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} d_e(f_i^1, f_j^2) x_{i,j}, \quad (3.1)$$

où les *flux* $x_{i,j}$ suivent les contraintes suivantes :

$$\forall i = 1, \dots, n_1, \forall j = 1, \dots, n_2, x_{i,j} \geq 0$$

,

$$\begin{aligned} \forall i = 1, \dots, n_1, \quad & \sum_{j=1}^{n_2} x_{ij} = w_i^1, \\ \forall j = 1, \dots, n_2, \quad & \sum_{i=1}^{n_1} x_{ij} = w_j^2. \end{aligned}$$

En suivant l'illustration des déplacements de terre, la deuxième et la troisième contrainte impose que toute la terre soit déplacée et que tous les trous soit remplis. Dans le cadre de l'organisation spatiale des couleurs, les descripteurs f_i correspondent à des vecteurs incluant la couleur et des informations spatiales.

3.1.4 Utilisation d'histogrammes couleurs

Cette section présente tout d'abord l'approche générale utilisée dans la littérature pour utiliser des histogrammes dans la recherche d'images selon l'OSC. Puis nous verrons un exemple récent de ce type de méthode.

Intégration des informations spatiales

Pour modéliser le contenu des intensité d'une image en niveaux de gris, nous pouvons utiliser un histogramme à une dimension. Les valeurs de l'histogramme représentent la probabilité d'apparition d'un niveau de gris dans l'image. Dans le cas des images couleurs, nous pouvons soit séparer les trois canaux d'un espace ABC quantifié et construire trois histogrammes, soit mélanger les canaux sur un seul vecteur en les plaçant les uns à la suite des autres, soit construire directement un histogramme en 3 dimensions sous la forme d'un cube discrétisé où la probabilité d'une couleur $(a, b, c) \in ABC$ dans l'image s'écrit directement $h_{ABC}(a, b, c) = Pr\{A = a, B = b, C = c\}$.

¹Il est possible de généraliser le problème au cas où les poids totaux sont différents, voir [Rubner et al., 2000]. Nous ne rencontrerons jamais cette situation dans cette thèse.

Pour intégrer les informations spatiales, une approche directe consiste à construire un histogramme 5D où l'espace de représentation 5D est lui même quantifié. Alors, la probabilité s'écrit $h_{ABCDXY}(a, b, c, x, y) = Pr\{A = a, B = b, C = c, X = x, Y = y\}$. Dans ce cas, cette distribution représente la probabilité d'une couleur $(a, b, c) \in ABC$ d'apparaître à la position (x, y) dans l'image. Cela conduirait à des histogrammes extrêmement grands pour pouvoir représenter correctement l'ensemble des couleurs et positions possibles.

Une manière plus économique de procéder est de quantifier adaptativement chaque image pour ne considérer qu'un nombre faible de couleurs différentes, puis de conserver pour chaque couleur trouvée une information spatiale telle que le centroïde des pixels associés à cette couleur.

Toutes les distances présentées à la section précédente peuvent être utilisées pour comparer deux histogrammes. [Rubner et al., 2000] montrent que l'EMD est la plus performante pour comparer deux histogrammes ou distributions pondérées d'ensembles de pixels de même couleurs. Dans leurs travaux, ils utilisent une quantification de l'espace couleur par arbre k-dimensionnel [Bentley, 1975]. Les autres distances sont généralement peu robustes à la quantification des couleurs et au changement de contraste.

Histogrammes spatiaux chromatiques dynamiques

Nous présentons en détails dans cette section un exemple récent de méthode modélisant le contenu couleur-espace en utilisant un histogramme et une distance inter-bins. [Ciocca et al., 2002] proposent d'utiliser des histogrammes sur un nombre quantifié faible de couleurs. La quantification des couleurs se fait dans sRGB, de manière statique sur 11 couleurs prédéfinies $\{\mathbf{c}_i\}_{(i=1,\dots,11)}$ ou dynamique en utilisant un classifieur K-moyennes. Pour chaque ensemble de pixels $\mathcal{G}_i$ de la même couleur $\mathbf{c}_i$, un descripteur à 4 dimensions $S_i = \{h_i, \mathbf{b}_i, \sigma_i, \mathbf{c}_i\}$ est constitué où h_i est le ratios de la taille de $\mathcal{G}_i$ par rapport à l'ensemble des pixels, $\mathbf{b}_i$ est le centroïde de $\mathcal{G}_i$, et σ_i est la dispersion chromatique au sein de $\mathcal{G}_i$ avant quantification.

Les auteurs proposent une mesure de similarité entre descripteurs d_S qui pondère les contributions spatiales et couleurs de la manière suivante :

$$d_S(S_i, S_j) = \frac{1}{3} \left[\frac{\min(h_i, h_j)}{\max(h_i, h_j)} + \frac{\sqrt{2} - d_{euc}(\mathbf{b}_i, \mathbf{b}_j)}{\sqrt{2}} + \frac{\min(\sigma_i, \sigma_j)}{\max(\sigma_i, \sigma_j)} \right] \times \left[1 - \frac{1}{2} d_{euc}(\mathbf{c}_i, \mathbf{c}_j)^2 \right],$$

où d_{euc} est la distance euclidienne et $\mathbf{c}_i$ est exprimé dans l'espace HSV. La fonction d^* proposée pour mesurer la similarité entre deux images I et J est :

$$d^*(I, J) = \sum_{S_i} \max\{d_S(S_i, S_j) | S_j \in J\}.$$

Comme cette mesure n'est pas symétrique, la distance utilisée est

$$d(I, J) = \max\{d^*(I, J), d^*(J, I)\}$$

. Remarquons qu'il s'agit d'une distance inter-bins où le maximum de la distance d_S entre deux bins est retenues pour le calcul d^* . Ces histogrammes sont une amélioration des histogrammes spatiaux chromatiques déjà proposés par [Cinque et al., 2001], qui n'intégrait pas la couleur moyenne dans les descripteurs et qui utilisait une distance intra-bins.

3.1.5 Utilisation de méthodes de segmentation d'images

Rappelons qu'un certain nombre de méthodes spécifiquement appliquées aux arts visuels ont été présentées à la Section 2.2.5. La plupart de ces méthodes sont des méthodes utilisant une segmentation ou un partitionnement de l'image.

Un certain nombre de travaux s'intéressent à l'agencement relatif de groupes de pixels dans une image, obtenus souvent par des méthodes de segmentation ou manuellement en laissant l'utilisateur indiquer les régions d'intérêt. L'agencement relatif étudié peut se traduire tout d'abord par des relations topologiques analysées à partir des intersections des espaces délimités par les formes [Egenhofer and Franzosa, 1991]. Des relations symboliques de type *au dessus de* ou *au Nord-Ouest de*, sont aussi souvent exploitées, par exemple par projection des régions sur les axes de l'espace dans la méthode *2D-strings* [Chang et al., 1987]. Nous avons présenté une de ces méthodes à la Section 2.2.5 qui a été appliquée aux œuvres d'art [Berretti et al., 2003]. Ces méthodes s'intéressent à l'agencement spatial relatif des régions entre elles, indépendamment de leur position globale dans l'image et souvent indépendamment de leur surface, par le biais de termes de normalisation. Il s'agit donc d'une problématique différente de celle l'étude de l'OSC.

Plusieurs méthodes de partitionnement d'images en un nombre restreint de régions rectangulaires ont été aussi proposées à partir de 1995 pour essayer de saisir la composition spatiale globale de l'image. Elles peuvent être statiques [Flickner et al., 1995, Gong et al., 1996, Stricker and Dimai, 1997, Lipson et al., 1997] ou dynamique [Yamamoto et al., 1999] (voir Figure 3.4). Chaque région peut être représentée par sa couleur moyenne ou son histogramme couleur. Ces méthodes de découpage de l'image utilisent une distance *région à région* de même position ou de même niveau, et intègrent donc aucune information de dépendance mutuelle. De même elles sont peu robustes à la stratégie de partitionnement ou au pas de quantification.

Plus récemment, les approches de découpage en régions homogènes ou RBIR² utilisent des méthodes de segmentation automatique d'images. Ces méthodes effectuent d'abord une segmentation de l'image en régions qui sont ensuite utilisées pour comparer deux

²Region Based Image Retrieval

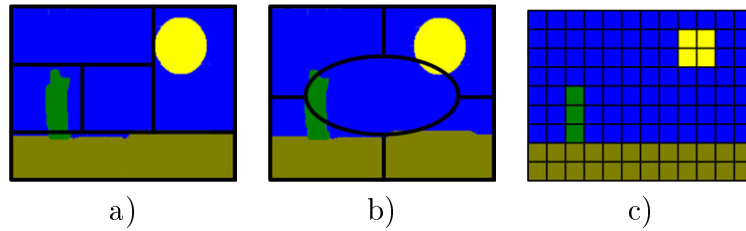


FIG. 3.4: Différentes méthodes de partitionnement d'images a) Partitionnement dynamique et construction d'un arbre binaire de régions incluses les unes dans les autres [Yamamoto et al., 1999]. b) Partitionnement flou en 5 régions dont une région centrale [Stricker and Dimai, 1997]. c) Partitionnement statique de type imagerie Qbic 9×12 pixels [Holt et al., 1997].

images. De nombreuses méthodes de segmentation ont été utilisées, par exemple [Rugna and Konik, 2002, Prasad et al., 2004, Aggarwal et al., 2002]. Ces méthodes souffrent souvent d'un grand nombre de paramètres. Par ailleurs les erreurs de segmentation sont très fréquentes et ont souvent une incidence sur l'indexation. Pour résoudre cet inconvénient, deux niveaux de segmentation ou description sont parfois utilisés [Dai and Zhang, 2005].

Le modèle *Blobworld* segmente l'image sous la forme de mélanges de Gaussiennes multidimensionnelles [Carson et al., 2002]. Le nombre de Gaussiennes (appelées *blobs* dans ce contexte) est un paramètre important et fixé par une méthode MDL (*Minimum Description Length*). La technique MDL est issue de la théorie de l'information [Rissanen, 1978]. Elle consiste à chercher le meilleur compromis entre un minimum de paramètres utilisés (le nombre de blobs) pour représenter un ensemble de données (l'image) et un maximum de fidélité aux données. La mesure de similarité pour comparer deux distributions de blobs est une distance de Mahalanobis.

[Wang et al., 2001] ont proposé le système appelé SIMPLicity. Ce système utilise un algorithme de segmentation k -moyennes appliqué aux bloc 4×4 pixels de l'images et où k est le nombre de régions recherchées. Pour fixer k , l'algorithme utilise des valeurs croissantes de k jusqu'à ce que la distance quadratique entre les blocs et les centres des clusters trouvées soit inférieure à un seuil fixe. Chaque bloc de 16 pixels est caractérisé par 6 descripteurs correspondant à la couleur moyenne du bloc dans l'espace LUV, et 3 descripteurs issus d'une analyse en ondelettes. La mesure de similarité entre deux groupes de régions est une distance de Mahalanobis.

Plusieurs approches utilisent une méthode de segmentation conjointement avec la distance EMD vue à la Section 3.1.3. [Liu et al., 2005] utilise la segmentation JSEG [Deng and Manjunath, 2001] avec un étiquetage sémantique des couleurs de chaque région trouvée. Ils utilisent la région EMD pour comparer deux ensembles de régions où chacune est pondérée par la surface relative de la région par rapport à l'image. [Jing et al., 2004] utilisent également la méthode de segmentation JSEG avec l'EMD. Ils ajoutent de plus un schéma de contrôle de pertinence (*Relevance Feedback*).

[Greenspan et al., 2000, Dvir et al., 2002] utilisent la segmentation BlobWorld décrite ci-dessus et l'EMD pour comparer deux distributions de blobs. La distance au sol utilisée pour comparer deux blobs entre eux est une distance de Fréchet [Fréchet, 1906]. Les auteurs illustrent cette méthode pour deux applications. La première est une problématique de recherche d'images. Nous utiliserons cette méthode pour la comparer à notre méthodologie. La deuxième application est celle de l'adaptation du modèle d'une image requête en fonction d'une cible. Pour adapter une possible sur-segmentation ou sous-segmentation introduite par la méthode BlobWorld dans la requête, les auteurs suggèrent une stratégie itérative de fusion ou scission de blobs lorsque l'EMD indique que des blobs ont été effectivement fusionnés ou scindés lors du transport optimal vers une image cible.

[Gousseau, 2003] a proposé une comparaison de l'organisation spatiale de deux images en noir et blanc en utilisant leurs ensembles de niveaux. Soit une image u , les ensembles de niveaux dit inférieurs sont les ensembles $\chi_\lambda(u) = \{x \in \mathbb{R}^2, u(x) \leq \lambda\}$ où λ parcourt le spectre des niveaux de gris. En considérant ces ensembles de niveaux comme des réalisations d'ensembles fermés aléatoires, et en utilisant un test topologique d'indépendance entre ces réalisations pour deux images, il introduit une mesure de l'indépendance topologique et spatiale entre deux images. Plus deux images ont leurs ensembles de niveaux dépendants, plus elle ont une organisation spatiale similaire. Cette méthode peut difficilement se généraliser aux images couleurs, car les relations d'inclusions sont alors perdues.

3.2 Comparaison de l'OSC sous la forme d'un problème d'optimisation global

Nous proposons de comparer directement l'organisation spatiale de deux images différentes en considérant cette comparaison comme étant l'optimisation globale du transport des pixels d'une image requête vers les pixels d'une image cible. Ce problème de transport se place dans un espace à cinq dimensions où les trois premières dimensions sont les canaux de l'espace colorimétrique CIE-Lab, et les deux dernières dimensions sont les dimensions spatiales (X, Y) de l'image. Nous verrons à la Section 3.2.1 que ce problème se simplifiera en un *problème d'affectation* de pixels sans possibilité de scission ou fusion de pixels. Idéalement, nous souhaiterions que ce problème soit appliqué à l'ensemble des pixels des images originales, mais les coûts calculatoires d'une telle approche sont prohibitifs. C'est pourquoi nous choisissons de calculer le transport optimal entre deux images fortement sous échantillonnées. Nous exposerons les caractéristiques de ce sous échantillonnage à la Section 3.2.2. L'EMD requiert le choix d'une distance au sol entre deux pixels. Nous avons testé deux distances au sol différentes, présentées dans les Section 3.2.3 et 3.2.4.

3.2.1 Appliquer l'EMD aux pixels

En imaginant que nous puissions déplacer les pixels d'une image et changer leurs couleurs, quel est le travail minimal à fournir pour envoyer une image vers une autre ? La traduction mathématique d'un tel problème est celle d'un problème de transport tel qu'introduit à la Section 3.1.3. Ce travail minimal à fournir pour transporter une image vers une autre peut être vu comme une mesure de similarité. Si le travail est faible, alors nous dirons que les deux images sont assez similaires selon leur OSC.

Plutôt que d'utiliser une segmentation préliminaire de l'image en régions, nous proposons d'appliquer l'EMD directement aux pixels. Pour comparer deux images nous appliquons directement l'EMD aux descripteurs $\mathcal{F} = \{(f_i, 1/n)_{i=1,\dots,n}\}$ où n est le nombre de pixels dans l'image et chaque f_i est un vecteur à 5 dimensions contenant la couleur et la position du i^{me} pixel. Un poids unitaire $1/n$ est attribué à chaque descripteur car chaque pixel couvre la même surface dans l'image. Dans ce cas particulier où tous les descripteurs ont le même poids, le problème de transport est simplifié en un problème dit *d'affectation* [Ford and Fulkerson, 1956] où les descripteurs ne peuvent être divisés lors de leur transport. La distance entre deux ensembles de descripteurs est alors définie par :

Si $\mathcal{F}^1 = \{(f_i^1, 1/n)\}_{i=1,\dots,n}$ et $\mathcal{F}^2 = \{(f_i^2, 1/n)\}_{i=1,\dots,n}$ sont deux ensembles de descripteurs pondérés dans un espace métrique (E, d_e) , alors

$$d(\mathcal{F}^1, \mathcal{F}^2) = \min_{\phi} \sum_{i=1}^n d_e(f_{\phi(i)}^1, f_i^2), \quad (3.2)$$

où ϕ est une permutation de l'ensemble des $\{1, \dots, n\}$.

La solution optimale est une solution où les descripteurs n'ont pas été divisés ou fusionnés lors du déplacement dans E . Dans notre cas, chaque pixel d'une image requête est donc entièrement assigné à la position d'un pixel cible. Chaque flux x_{ij} dans l'Equation (3.1) vaut soit 0 soit 1. Dans le cadre d'un tel problème d'affectation, le calcul de l'EMD est grandement simplifié [Burkard and Çela, 1998, Burkard et al., 2008].

La matrice des coûts d'un tel problème correspond à la matrice des valeurs $d_e(f_i^1, f_j^2)$ pour $(i, j) \in \{1, \dots, n\}^2$. Cette matrice n'a aucune structure particulière, et en particulier elle n'est pas creuse. Par conséquent l'algorithme le plus efficace pour résoudre un tel problème d'affectation est l'algorithme de Kuhn-Munkres³ [Kuhn, 1955] en $O(n^3)$, appelé aussi méthode Hongroise. Une manière d'accélérer ce calcul pourrait être de rendre creuse cette matrice en imposant une limite supérieure aux déplacements possibles des pixels dans E [Burkard and Çela, 1998, Burkard et al., 2008].

³Code disponible à l'adresse : <http://www.informatik.uni-freiburg.de/~stachnis/resources.html>

3.2.2 Utilisation d'imagettes

Contraintes calculatoires

Bien que la réduction à un problème d'affectation et l'utilisation de l'algorithme de Kuhn-Munkres permettent de réduire fortement les temps de calculs sans approximations, ceux ci restent encore prohibitifs lorsque la méthode est appliquée à une image de résolution normale au sein d'une base de données de taille classique. C'est pourquoi nous appliquons cette approche à des images ayant une résolution beaucoup plus faible que les images originales. Ces images seront dans toute la suite appelées imagettes, voir la Figure 3.5.

Ceci n'est pas une limitation trop forte de la méthode compte tenu que nous nous intéressons à l'organisation spatiale globale et à grande échelle des couleurs dans l'image. Notre objectif est de saisir la composition générale des régions colorées et non d'étudier les textures par exemple. Nous verrons par ailleurs dans les résultats expérimentaux qu'accroître la taille de ces imagettes a un impact faible sur l'indexation.

En gardant les mêmes notations qu'à la section précédente, le nombre total de pixels de ces imagettes est fixé à n . Ce paramètre peut être fixé en fonction des contraintes calculatoires et de la taille de la base de données utilisée. L'influence de n sur les résultats sera étudié à la Section 3.3.4. Un ordre de grandeur pour n est quelques centaines, et dans la majorité de nos expériences nous utiliserons $n = 150$.



FIG. 3.5: A gauche une image originale (3200×2200 pixels). A droite une imagette de cette image (10×15 pixels).

Calcul des imagettes

Le calcul des imagettes est réalisé de la manière suivante. Lorsqu'il est possible de le faire, nous corrigeons tout d'abord les canaux RGB du facteur Gamma pour obtenir les valeurs tristimulus. Selon les lois additives de Grassman [Wyszecki and Stiles, 2000], les valeurs des tristimulus peuvent alors être moyennées pour obtenir les imagettes. Ayant ensuite pour objectif de comparer ces imagettes entre elles via l'EMD, et la distance d_e entre pixels, nous convertissons ces imagettes vers l'espace psychométrique CIE-Lab. La

description globale d'une image est donc composée des n pixels $f_i = \{L_i, a_i, b_i, X_i, Y_i\}$ de son imagerie où L, a, b sont les coordonnées dans CIE-Lab, et X, Y la position du pixel dans l'imagerie. Chacun de ces pixels est pondéré par un poids $1/n$.

Remarque : l'utilisation d'images rapproche la méthode proposée des méthodes de partitionnement présentées à la Section 3.1.5 comme celle de Qbic [Flickner et al., 1995]. La différence majeure entre notre méthode et celles ci réside dans la mesure de similarité. Les méthodes de partitionnement classiques utilisent une distance entre régions de même position.

3.2.3 Distance d_e inter-pixels exponentielle

Nous avons testé deux distances au sol d_e différentes dans nos expériences pour la résolution de l'Equation 3.2. La distance d_e permet de comparer deux descripteurs pixeliques f_i et f_j . Nous présentons dans cette section une distance reposant sur les métriques Euclidiennes dans les espaces couleurs et spatial :

$$d_e(f_i, f_j) = \alpha * (1 - \exp \{-\Delta_c/\sigma_c\}) + (1 - \alpha) * (1 - \exp \{-\Delta_{xy}/\sigma_s\}), \quad (3.3)$$

où α est un paramètre permettant de pondérer les contributions couleur et spatiales, et où $\Delta_c = \sqrt{(L_i - L_j)^2 + (a_i - a_j)^2 + (b_i - b_j)^2}$ et $\Delta_{xy} = \sqrt{(X_i - X_j)^2 + (Y_i - Y_j)^2}$. Les paramètres σ_c et σ_s sont choisis en fonction des dynamiques que nous souhaitons imposer aux déplacements couleurs et spatiaux. Des valeurs typiques pour ces paramètres sont $\sigma_c = 15$ et σ_s égal à un quart de la diagonale de l'imagerie. L'utilisation des exponentielles a été choisie car elles permettent de fixer un majorant à l'EMD. Ceci repose sur l'idée qu'au delà d'une certaine distance, deux images sont simplement considérées comme différentes. Cette distance non linéaire est adaptée à la recherche automatique d'images où une image cible est soit proche d'une requête (selon une certaine valeur de similarité) soit simplement différente.

3.2.4 Ajout d'informations supplémentaires

Sachant que le coût calculatoire de l'EMD dans le cadre de notre problématique varie en $O(n^3)$, nous présentons dans cette section une approche permettant de conserver cette complexité tout intégrant plus d'information en chaque pixel que la couleur moyenne seule.

Imagettes bicolores

Dans les sections précédentes nous avons présenté l'utilisation de l'EMD en l'appliquant directement au transport des pixels d'une imagerie. Chaque pixel est représenté par sa position et la couleur moyenne de la région qu'il représente dans l'image originale. Quelles solutions pouvons-nous envisager si nous souhaitons conserver plus d'informations que seule cette couleur moyenne? Nous pouvons envisager des informations caractérisant la texture de la région sous-jacente (coefficients de décomposition en ondelettes par exemple). Pour plus de simplicité, nous nous intéressons dans cette section à la conservation et l'utilisation non plus d'une, mais de deux couleurs principales de la région sous-jacente. Nous construisons alors ce que nous appellerons des imagettes *bicolores*, ayant en chaque position (X, Y) deux couleurs, ainsi qu'un poids associé à chacune.

A l'aide d'un classifieur K -moyennes à deux classes ($K = 2$), pour chaque élément i de l'imagette bicolore, nous calculons les deux couleurs moyennes principales des pixels associés dans l'image originale. Le classifieur associe à chacune de ces deux couleurs un poids correspondant au nombre de points de la classe trouvée. Une de ces deux classes a nécessairement un poids supérieur ou égal à celui de l'autre classe. Nous noterons $\mathcal{C}_M$ la couleur associée à cette classe (couleur majoritaire, choisie arbitrairement lorsqu'il y a égalité des poids) et $\mathcal{C}_m$ l'autre couleur (couleur minoritaire). Nous normalisons aussi ces poids (notés p_M et p_m) par le nombre total de pixels associés dans l'image originale de telle sorte que $p_M + p_m = 1$.

Pour appliquer l'EMD entre deux imagettes bicolores, nous notons toujours f_i le descripteur pixelique composé désormais de $f_i = \{\mathbf{C}_i, X_i, Y_i\}$, où $\mathbf{C}_i$ est la bicouleur :

$$\mathbf{C}_i = \{(\mathcal{C}_M^i, p_M^i), (\mathcal{C}_m^i, 1 - p_M^i)\}.$$

Sur la Figure 3.6 sont présentés trois exemples d'imagettes bicolores, avec de gauche à droite, l'image originale, l'imagette à une seule couleur (calculée selon la méthode présentée à la Section 3.2.2), l'imagette de couleur majoritaire, l'imagette de couleur minoritaire, et l'imagette en niveaux de gris des poids p_M^i normalisés du noir au blanc sur 8bits. Remarquons que le ratio des imagettes est fixe, et qu'il peut être différent de celui des images originales. Nous discuterons de ce choix en détail à la Section 3.3.3. Nous voyons sur ces trois exemples que l'utilisation des bicouleurs permet logiquement de garder des couleurs plus saturées, moins dégradées par moyennage. Lorsque la région sous-jacente est homogène, le poids associé à cette région p_M^i est généralement élevé (pixels blancs).

Remarque : les deux couleurs d'un pixel bicolore ont la même position X_i, Y_i dans ce modèle. Nous appellerons ce couple de couleurs une bicouleur. Nous pourrions recalculer deux positions sub-pixeliques à l'aide des positions moyennes des deux classes issues du classifieur. Nous discuterons de cette approche dans la discussion.

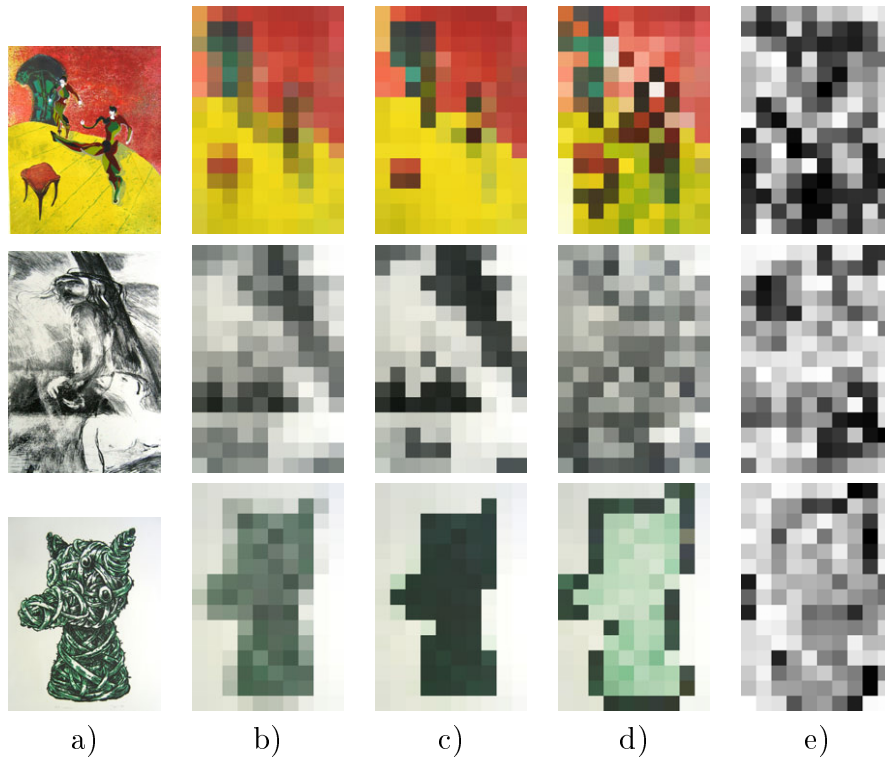


FIG. 3.6: Imagettes bicolores. a) image originale. b) imagette classique dont chaque pixel a la couleur moyenne calculée sur la région sous-jacente dans l'image originale. c) imagette des couleurs majoritaires obtenue avec un classifieur 2-moyennes. d) imagette des couleurs minoritaires. d) imagette de la pondération p_M entre les deux classes issues du classifieur (normalisée sur $[0, 255]$).

Reformulation du problème d'affectation entre deux imagerie bicolores

La formulation du problème d'affectation pour déplacer une imagerie bicolore à une autre ne change pas :

Si $\mathcal{F}^1 = \{(f_i^1, 1/n)_{i=1, \dots, n}\}$ et $\mathcal{F}^2 = \{(f_i^2, 1/n)_{i=1, \dots, n}\}$ sont deux imagerie bicolores répartis dans un espace métrique (E, d_e) , alors

$$d(\mathcal{F}^1, \mathcal{F}^2) = \min_{\phi} \sum_{i=1}^n d_e(f_{\phi(i)}^1, f_i^2), \quad (3.4)$$

où ϕ est une permutation de l'ensemble des $\{1, \dots, n\}$.

Pour résoudre ce problème d'affectation nous avons néanmoins besoin de redéfinir la distance d_e entre deux pixels bicolores. Comme les deux couleurs d'un pixel bicolore ont la même position, nous pouvons toujours calculer le déplacement spatial par une distance euclidienne Δ_{xy} comme dans l'Équation 3.3. En revanche, pour le déplacement couleur d'une bicouleur à une autre, nous proposons d'utiliser à nouveau une distance de transport d' imbriquée dans d_e . Nous définissons la distance d_e entre deux pixels bicolores par :

$$d_e(f_i, f_j) = \alpha * (1 - \exp\{-d'(\mathbf{C}_i, \mathbf{C}_j)/\sigma_c\}) + (1 - \alpha) * (1 - \exp\{-\Delta_{xy}/\sigma_s\}), \quad (3.5)$$

où $d'(\mathbf{C}_i, \mathbf{C}_j)$ est la distance de transport entre les bicouleurs $\mathbf{C}_i = \{(\mathcal{C}_M^i, p_M^i), (\mathcal{C}_m^i, 1 - p_M^i)\}$ et $\mathbf{C}_j = \{(\mathcal{C}_M^j, p_M^j), (\mathcal{C}_m^j, 1 - p_M^j)\}$. Il s'agit ici d'un problème de transport général. En effet il faudrait avoir $p_M = p_m$ pour retrouver un problème d'affectation. Néanmoins dans le cas d'un problème à deux classes de poids total constant ($p_M + p_m = 1$), celui ci se simplifie grandement.

EMD entre deux bicouleurs

Rappelons tout d'abord la notation de l'EMD (voir Equation (3.1)), dans le cadre du problème 2×2 correspondant au transport d'une bicouleur $\mathbf{C}_i$ vers une bicouleur $\mathbf{C}_j$. Mathématiquement, l'EMD entre deux bicouleurs $\mathbf{C}_i = \{(\mathcal{C}_M^i, p_M^i), (\mathcal{C}_m^i, 1 - p_M^i)\}$ et $\mathbf{C}_j = \{(\mathcal{C}_M^j, p_M^j), (\mathcal{C}_m^j, 1 - p_M^j)\}$ ayant le même poids total $p_M^i + p_m^i = p_M^j + p_m^j = 1$ est défini par :

$$d'(\mathbf{C}_i, \mathbf{C}_j) = \min_{x_{kp}} \sum_{k \in \{M, m\}} \sum_{p \in \{M, m\}} d'_e(\mathcal{C}_k^i, \mathcal{C}_p^j) x_{kp}, \quad (3.6)$$

où les flux x_{kp} suivent les contraintes suivantes :

$$\forall k \in \{M, m\}, \forall p \in \{M, m\}, x_{kp} \geq 0$$

3.2. Comparaison de l'OSC sous la forme d'un problème d'optimisation global

,

$$\forall k \in \{M, m\}, \quad \sum_{p \in \{M, m\}} x_{kp} = p_k^i,$$

$$\forall p \in \{M, m\}, \quad \sum_{k \in \{M, m\}} x_{kp} = p_p^j,$$

où la distance $d'_e(\cdot, \cdot)$ est la distance au sol entre deux couleurs, associée à la distance EMD $d'(\cdot, \cdot)$ ci dessus.

Alors, le calcul de $d'(\mathbf{C}_i, \mathbf{C}_j)$ est réalisé par la vérification de une à deux inéquations résumées dans l'Algorithme 1. Pour alléger cette section, la preuve de ce résultat est fournie en Annexe B.

Algorithme 1 Calcul de l'EMD entre deux bicouleurs

En notant $c_{kp} = d'_e(\mathcal{C}_k^i, \mathcal{C}_p^j)$ pour $k, p \in \{M, m\}$ le coût associé au déplacement d'une couleur $\mathcal{C}_k^i$ vers une couleur $\mathcal{C}_p^j$ dans CIE-Lab, nous avons :

si $D = c_{MM} + c_{mm} - c_{mM} - c_{Mm} > 0$ **alors**

$$d'(\mathbf{C}_i, \mathbf{C}_j) = p_m^i c_{mM} + (p_M^j - p_m^i) c_{MM} + p_m^j c_{Mm}$$

sinon

si $p_M^i < p_M^j$ **alors**

$$d'(\mathbf{C}_i, \mathbf{C}_j) = p_M^i c_{MM} + p_M^j c_{mm} + (p_M^j - p_M^i) c_{mM}$$

sinon

$$d'(\mathbf{C}_i, \mathbf{C}_j) = p_M^j c_{MM} + p_M^i c_{mm} + (p_M^j - p_M^i) c_{mM}$$

fin si

fin si

Distance au sol entre deux couleurs

Dans un souci d'homogénéité avec la distance utilisée dans le cadre des imageries à une seule couleur, la distance d'_e qui est utilisée notamment pour calculer la matrice des coûts est la distance suivante $d'_e(\mathcal{C}_k, \mathcal{C}_p) = \sqrt{(L_k - L_p)^2 + (a_k - a_p)^2 + (b_k - b_p)^2}$.

Coût calculatoire

Cette approche est sensiblement équivalente à augmenter artificiellement n d'un facteur 2. Le coût calculatoire est par contre bien moindre puisque la complexité de problème de transport global entre deux imageries reste la même. Seul la distance au sol $d_e(\cdot, \cdot)$ est un peu plus coûteuse par rapport à la distance au sol utilisée dans le cadre des imageries à une seule couleur par pixel. Cette augmentation de n est artificielle car elle contraint néanmoins les deux couleurs de chaque pixel bicouleurs à rester ensemble lors du transport global.

3.3 Expériences

Nous présentons dans cette section l'ensemble des expériences et résultats liés à l'étude de l'OSC. Une série d'expériences pour mieux comprendre le fonctionnement de l'EMD dans le cadre du problème d'affectation et des imagerie est présentée à la Section 3.3.1. L'effet du paramètre de pondération α entre contribution spatiale et couleur sera illustré à la Section 3.3.2. La gestion des différents ratios hauteur/largeur d'images sera abordée à la Section 3.3.3. Nous étudierons ensuite à la Section 3.3.4 l'influence de la taille des imagerie sur les résultats. Nous comparerons ensuite dans la Section 3.3.5 notre méthode à une méthode utilisant une segmentation classique appelée BlobWorld. Nous comparerons l'utilisation d'images avec une seule couleur par pixel ou avec une bicouleur à la Section 3.3.6. Puis nous illustrerons la particularité du cas des bases uniquement constituées d'images en noir et blanc à la Section 3.3.7. Enfin, de nombreux exemples de requêtes sur différentes bases d'images seront présentés et discutés à la Section 3.3.8.

Les expériences sur images réelles utilisent cinq bases d'images différentes. L'ensemble de ces bases ainsi que leurs contenus sont présentés dans l'Annexe A. La première base d'images que nous avons utilisé est celle qui nous a été fournie par l'Institut de Recherche et d'Histoire des Textes (IRHT)⁴. Cette base contient 1500 enluminures numérisées avec une grande résolution et dont les couleurs ont été calibrées. Cette base est utilisée pour illustrer la plupart des expériences précédant celles de recherche d'images. Hormis la Section 3.3.4 où l'influence du paramètre n est discuté, toutes les expériences utilisent des imagerie de $n = 150$ pixels (10×15 ou 15×10). Les paramètres de la distance d_e sont fixés à $\sigma_c = 15$ et σ_s égal au quart de la diagonale. (voir Equation (3.3)). Le choix du paramètre α sera discuté à la Section 3.3.2. En dehors de cette section il sera toujours fixé à $\alpha = 0.55$. Enfin toutes les expériences utilisent des imagerie à une seule couleur, en dehors de la Section 3.3.6 où celles-ci seront comparées aux imagerie bicolores.

3.3.1 Comportement de l'EMD et effet *taquin*

Pour mieux comprendre le comportement de l'EMD, nous proposons de représenter l'affectation optimal trouvé par cette distance sous la forme de déplacements de pixels sur une imagerie de leurs positions de départ (celles de l'imagerie requête) à leurs positions d'arrivée (celles qui leur a été assignées dans l'imagerie cible). Sur chaque ligne de la Figure 3.7, deux images sont visibles en colonne a) et b). Nous calculons l'affectation optimal de l'image a) vers l'image b) avec l'EMD. Sur la colonne c), nous illustrons cette affectation par un ensemble de déplacements symbolisés par des segments en pointillés où la position finale est représentée par un point. Nous voyons que sur des formes simples les déplacements se font par commutation de deux pixels.

⁴IRHT, CNRS, 40 avenue d'Iena, 75116 Paris. URL : <http://www.irht.cnrs.fr/>

3.3. Expériences

Sur la colonne d), la distance d_e de l'Equation 3.3 a été modifiée, en prenant le carré des distances Euclidiennes : $\Delta_c = (L_i - L_j)^2 + (a_i - a_j)^2 + (b_i - b_j)^2$ et $\Delta_{xy} = (X_i - X_j)^2 + (Y_i - Y_j)^2$. Cette distance d_e avait été étudiée dans nos tests préliminaires, puis abandonnée car elle induisait un effet caractéristique visible sur la colonne d) de la Figure 3.7. Cet effet s'apparente à celui du jeu classique appelé *taquin* (permutation de pixel voisin en pixel voisin). Il est instable car il n'apparaît que dans certains cas particuliers où les images présentent un *chemin* entre deux pixels où le déplacement par taquin est moins coûteux que la permutation directe des deux pixels. Pour qu'un effet taquin puisse apparaître entre deux pixels, la distance minimale entre ces deux pixels doit être égale à environ $0.8\sigma_s$ (Figure 3.8).

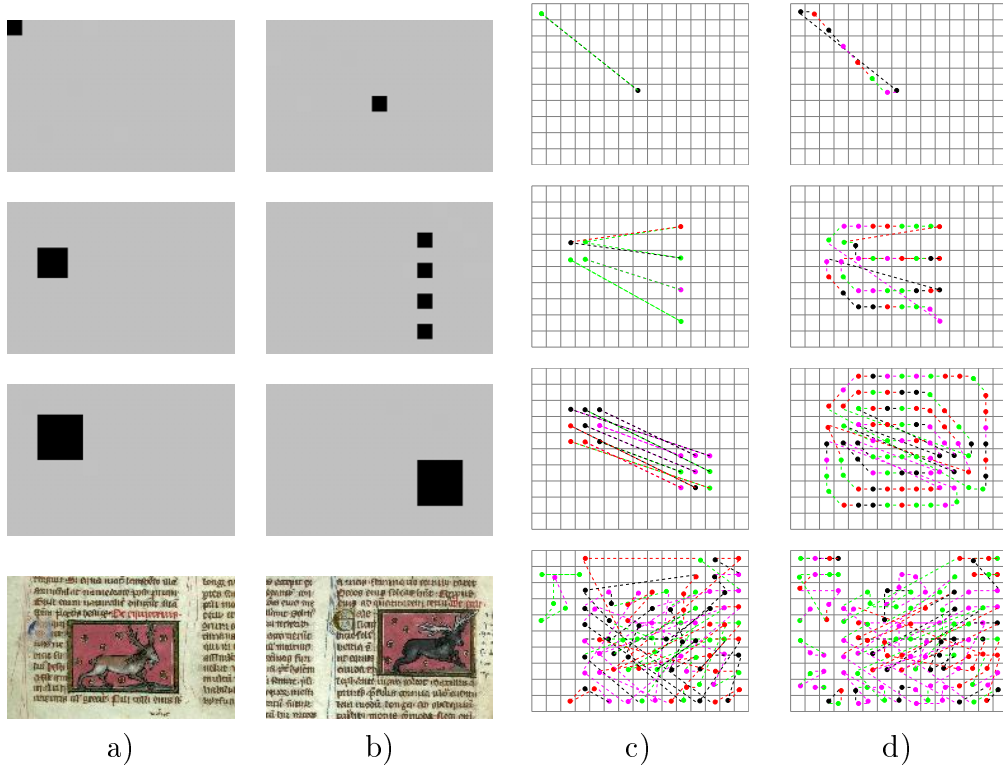


FIG. 3.7: Affection des pixels par l'EMD et effet taquin. Pour chaque ligne, l'affectation de l'image en colonne a) à l'image visible en colonne b) est représenté sur la colonne c). Chaque déplacement est représenté par un segment en pointillé où un point symbolise le pixel d'arrivée. Nous voyons sur la colonne c) que les affectations se font par permutation de deux pixels entre eux. Sur la colonne d), le même test est effectué avec une distance d_e utilisant le carré des distances Euclidiennes. Un effet taquin est visible sur chaque exemple de cette colonne. Sur un exemple réel (dernière ligne), les affectations ne sont pas facilement lisibles. Cette même transformation est représentée d'une manière différente sur la Figure 3.9.

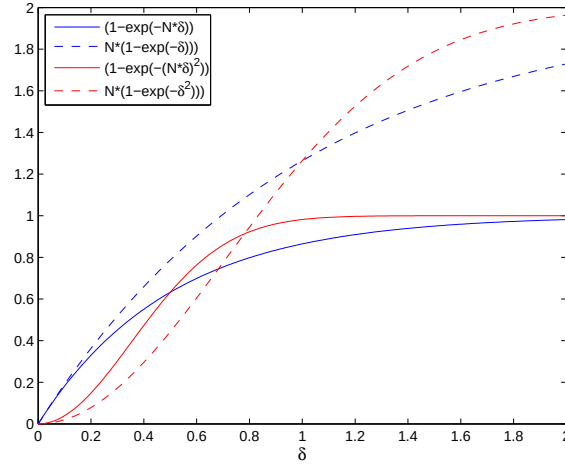


FIG. 3.8: Effet taquin. Lorsque le carré d'une distance Euclidienne est utilisé dans l'Equation 3.3 (courbes rouges), il est moins coûteux d'effectuer un déplacement $N\delta$ en N étapes (courbe rouge en pointillée) que directement entre deux pixels d'imagettes (courbe rouge continue), pour $\delta < 0.8$. Sur la figure est utilisée la valeur $N = 2$. Dans le cadre des imagettes 10×15 et σ_s égal au quart de la diagonale, ce déplacement correspond à 3.6 pixels. C'est pourquoi lorsqu'un taquin de moindre coût existe potentiellement entre deux pixels à déplacer, l'EMD peut choisir de l'utiliser pour assigner deux pixels (Figure 3.7d)). Dans le cas d'une distance Euclidienne normalisée (courbes bleues), cette situation n'arrive jamais car $N(1 - \exp(-\delta)) > (1 - \exp(-N\delta))$ pour tout $\delta > 0$.



FIG. 3.9: Affectation des pixels par l'EMD entre deux images réelles. L'imagette en bas à droite représente l'affectation des pixels de l'imagette en haut au centre aux positions optimales minimisant l'EMD avec l'imagette en bas au centre. Les pixels de la région rouge requête par exemple ont été assignés au niveau de la région rouge cible. Les marges blanches verticales et horizontales de l'imagette requête ont été assignées au niveau des deux marges verticales de l'imagette cible (voir Figure 3.7c) en bas).

3.3.2 Facteur de pondération α

Influence de α

Nous avons présenté à la Section 3.1.3 l'EMD que nous utilisons pour mesurer la similarité entre deux images. La distance d_e présentée en Section 3.2.3 et en Section 3.2.4 utilise un paramètre de pondération $\alpha \in [0, 1]$ entre la contribution des couleurs et celle du déplacement spatial (voir l'Equation (3.3)). Si l'utilisateur souhaite privilégier l'organisation spatiale, α doit être fixé à une valeur proche de 0. Alors, les mouvements de pixels seront pénalisés par rapport à un *déplacement* dans l'espace colorimétrique. Inversement, plus α s'approche de 1, plus les pixels peuvent se déplacer *librement* dans l'espace 2D de l'image vers des pixels de couleurs plus fidèles. Le cas extrême $\alpha = 1$ correspond à une requête selon le seul contenu couleur, et est équivalent à une méthode par histogramme des couleurs de l'image utilisant l'EMD entre histogrammes. Deux résultats sont visibles sur la Figure 3.10 illustrant l'influence de α .



FIG. 3.10: Pondération spatiale couleur avec le paramètre α . Deux requêtes (encadrées en rouge) sont suivies de leurs 3 plus proches voisins. a) $\alpha = 0.5$ est utilisé. Nous voyons que l'organisation spatiale est respectée. b) $\alpha = 1$, la méthode devient alors équivalente à une recherche par histogramme couleur utilisant l'EMD. L'organisation spatiale y est effectivement perdue.

Discussion sur le choix général de α

Il n'est pas évident de choisir une valeur de α répondant aux besoins d'une comparaison intuitive de l'OSC. Le mélange entre espace et couleur n'est pas naturel, et le choix $\alpha = 0.5$ n'a pas une grande signification selon l'OSC. Soulignons aussi que la valeur de α à choisir dépend aussi indirectement des paramètres σ_c et σ_s .

Dans cette section, nous avons constitué 10 classes de 5 images sélectionnées dans la base IRHT dans des séries d'images très similaires selon l'OSC. Ces classes sont visibles sur la Figure 3.11. Par ailleurs, nous avons construit 10 autres classes de 5 images constituées des mêmes images tournées de 180° . Ces classes ont exactement le même contenu couleur mais elles ont été choisies de manière à ce qu'une fois tournées de 180° , elles n'aient plus du tout la même OSC (les OSC choisies ne sont pas invariantes par rotation). Nous imposons donc que le choix de α soit suffisamment élevé pour retrouver les classes (en étant tolérant avec les déplacements de pixels), mais aussi suffisamment bas pour ne pas trop s'approcher d'une recherche par le seul contenu couleur.

Pour évaluer la qualité de la classification en fonction de α nous utilisons l'indice appelé *décompte de gain cumulé* (ou *discounted cumulative gain*, DCG). Cette mesure cumule les contributions g_r associés à chaque image de rang r dans la liste des résultats ordonnée selon l'EMD. La contribution g_r de la r^{eme} image vaut $1/\log_2(r)$ si elle appartient à la même classe que la requête et 0 sinon. Ce cumul de gain est normalisé par le cumul maximal possible (si toutes les images de la classe arrivent dans les premières). Pour une requête appartenant à une classe de taille C dans une base de taille N , le DCG est alors donné par :

$$DCG = \frac{1 + \sum_{r=2}^N g_r}{1 + \sum_{j=2}^C 1/\log_2(j)}.$$

La courbe DCG en fonction de α est visible sur la Figure 3.12. Cette courbe suggère que le choix de $\alpha = 0.55$ correspondant au maximum du DCG est à privilégier. Nous notons une nette rupture de pente du DCG pour $\alpha > 0.6$.

3.3.3 Variabilité des formats d'images

Le ratio entre hauteur et largeur d'une image peut varier beaucoup d'une image à une autre. Les œuvres d'art n'échappent pas à cette grande variabilité comme en témoigne la Figure 3.13. Lorsque nous souhaitons chercher des œuvres similaires selon l'OSC, comment devons-nous gérer cette variabilité ?

Le choix d'un ratio est bien souvent fait par l'artiste avant même de commencer

3.3. Expériences



FIG. 3.11: Les 10 classes de 5 images utilisées pour choisir la valeur de α .

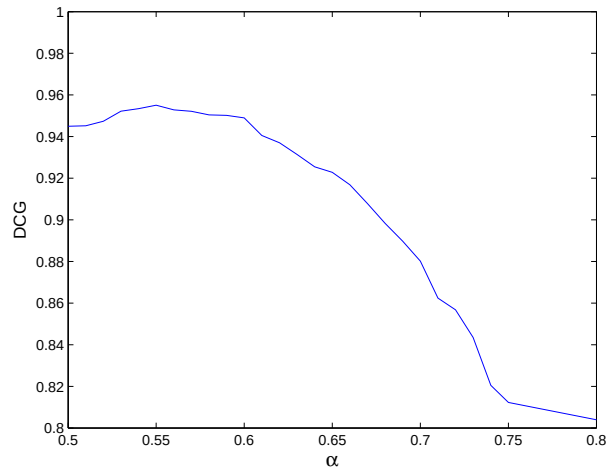


FIG. 3.12: Courbes de l'indice DCG en fonction de α . L'indice DCG mesure la performance de la classification. Nous voyons sur cette figure que le maximum est atteint à $\alpha = 0.55$.

son travail. Si nous considérons les peintures, les fabricants de toiles montées sur cadre proposent depuis longtemps une gamme de ratios définis (carré, double carré, paysage, marine, figure, ...). Même en supposant que les artistes adoptent toujours ces formats (ce qui est faux), est-il légitime de ne pas mettre en correspondance deux œuvres horizontales de ratios proches (comme marine et paysage) si elles ont la même OSC ? Inversement, une œuvre panoramique peut elle être mise en correspondance avec une fresque verticale quelle que soit son OSC ? Une solution pourrait être d'introduire une fonction de pénalisation de la similarité qui serait décroissante avec la différence de ratios.

Dans les expériences nous avons choisi plus simplement de séparer les formats verticaux et horizontaux. Les formats purement carrés apparaissent dans les deux catégories. Cette séparation assez stricte permet néanmoins de simplifier la méthode sans introduire de nouveaux paramètres difficiles à fixer étant donné la nature de ce que nous étudions. Elle est par ailleurs cohérente avec les distributions bimodales des ratios observés sur la Figure 3.13.

Notons que le ratio des images est un problème récurrent en recherche d'images qui ne se limitent pas aux œuvres d'arts. Ce problème est à notre connaissance éliminé en choisissant des descripteurs invariants par changement de format, en normalisant les descripteurs, ou en ramenant toutes les images à un même format. Ces solutions ne sont bien entendu pas adaptées à la comparaison selon l'OSC.

Remarquons aussi que l'EMD n'ayant aucune contrainte sur la taille des descripteurs d'images, nous pourrions directement comparer des imagelettes verticales et horizontales, de taille différente ou non [Rubner et al., 2000]. Pour nous permettre de rester dans le cadre du problème d'affectation (et non du problème général de transport), nous avons décidé

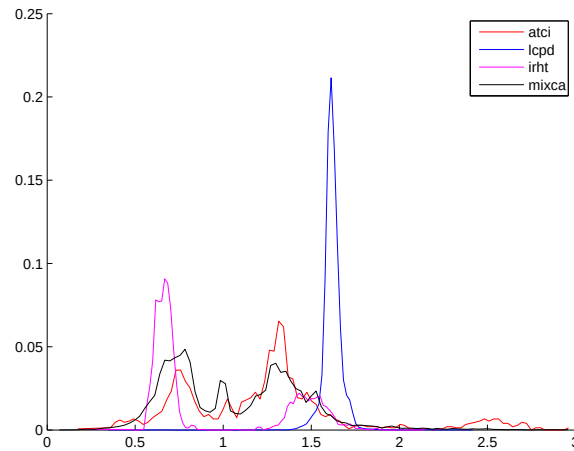


FIG. 3.13: Histogrammes normalisés des ratios entre la hauteur et la largeur de l'image pour les quatre premières bases utilisées. Nous voyons que ces distributions sont continues et présentent un creux au niveau du format carré. La base LCPD est uniquement composée de photographies verticales. La base CLIC n'a pas été incluse car elle n'est constituée que de deux ratios ($4/3$ et $3/4$).

de ramener toutes les images verticales à des imageries de ratio $3/4$, et toutes les images horizontales à des imageries de ratio $4/3$. Ceci induit donc une certaine déformation pour les images éloignées de ces deux ratios.

Ces ratios pourraient être fixés selon la base. Si ils sont bien adaptés à la base ATCI par exemple (voir Figure 3.13), nous voyons que les imageries de la base LCPD seraient moins déformées avec un ratio proche de 1.6.

3.3.4 Taille des imageries

Dans cette section, nous étudions l'influence du choix de la taille n des imageries. Nous avons vu à Section 3.2.2 que conserver la taille originale des images n'était pas concevable compte tenu des coûts calculatoires de l'algorithme de Kuhn-Munkres pour résoudre un problème d'affectation. Nous avons effectué plusieurs requêtes sur la base IRHT avec des valeurs de $n = 150, 600, 2400$. Ces valeurs correspondent à des imageries de taille 10×15 , 20×30 et 40×60 pour les imageries horizontales.

Trois résultats sont visibles sur les Figure 3.14 à Figure 3.16. Nous pouvons observer qu'une augmentation de n a un très faible impact sur ceux ci. Ceci illustre le fait qu'un ordre de grandeur de quelques centaines pour n est suffisant pour décrire l'OSC d'une image. Pour ces expériences nous avons utilisé un Pentium IV 4.3GHz et une programmation en C. Les temps de calculs sur une base de 1000 images sont d'environ 10 secondes

pour $n = 150$, puis 7 minutes pour $n = 600$, et 5 heures pour $n = 2400$. Ces temps sont cohérents avec la complexité en $O(n^3)$ de la méthode Hongroise. Pour les expériences présentées par la suite dans cette thèse, nous utiliserons toujours $n = 150$.

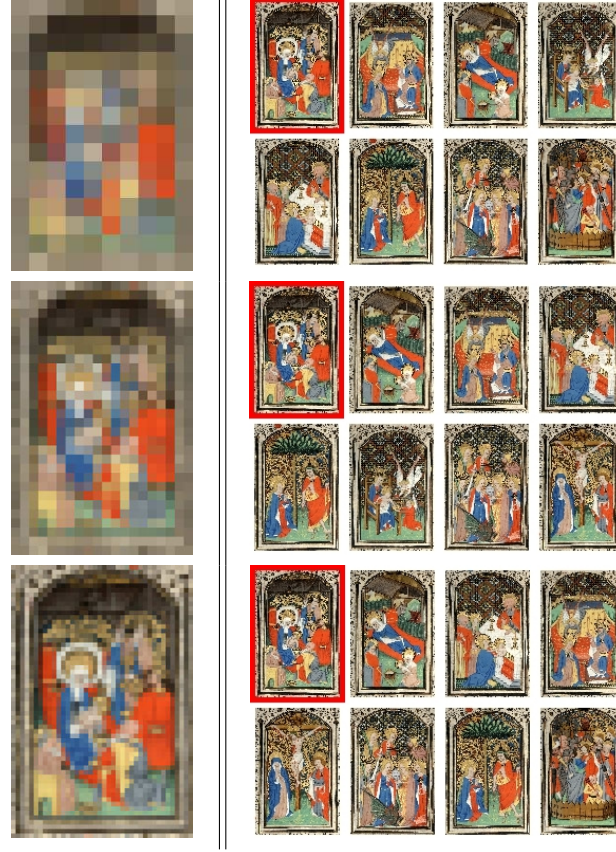


FIG. 3.14: Trois résultats de recherche par le contenu avec des imagerie de tailles différentes. Colonne de gauche, les imagerie de haut en bas ont une taille $n = 150, 600, 2400$. Colonne de droite, la requête est encadrée en rouge suivie, par ordre de distance croissante, des 7 résultats les plus proches. Les résultats et leur ordre d'apparition varient peu en fonction de n .

3.3.5 Comparaison avec l'utilisation d'une méthode de segmentation

Dans cette section nous comparons notre choix des imagerie en tant que descripteur de l'OSC par rapport à celui qui consiste à utiliser une méthode de segmentation automatique avant l'utilisation de la distance EMD comme c'est le cas dans les références [Greenspan et al., 2000, Dvir et al., 2002, Jing et al., 2004, Liu et al., 2005]. Pour cela, nous avons choisi la méthode de segmentation classique proposée par Malik et al. appelée BlobWorld⁵

⁵Code disponible à l'adresse : <http://elib.cs.berkeley.edu/blobworld/>

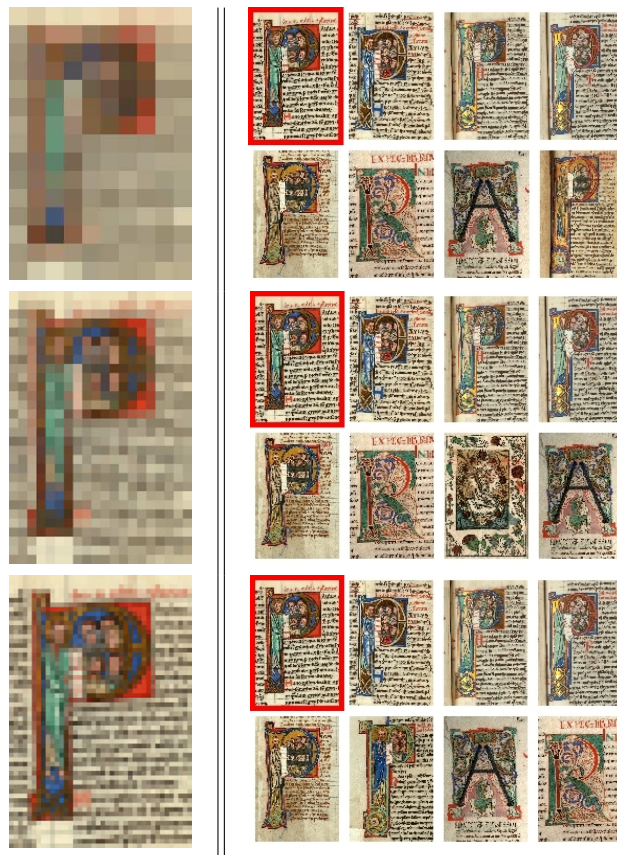


FIG. 3.15: Trois résultats de recherche par le contenu avec des imagerie de tailles différentes. Même organisation que sur la Figure 3.14



FIG. 3.16: Trois résultats de recherche par le contenu avec des imagerie de tailles différentes. Même organisation que sur la Figure 3.14

[Carson et al., 2002] que nous avons décrite à la Section 3.1.5. La méthode originale utilise les couleurs, textures et positions spatiales pour extraire les blobs. Pour que la méthode soit directement comparable à la notre dans le cadre de la distance d_e exponentielle nous avons utilisé la méthode BlobWorld uniquement en tenant compte des couleurs et positions spatiales. Nous utilisons l'EMD pour mesurer la distance entre deux ensembles de blobs. Les descripteurs issus de chaque blob sont alors le centre de gravité et la couleur moyenne du blob, pondéré par la surface du blob. Cette approche utilisant Blobworld suivie de l'EMD est aussi celle choisie par [Greenspan et al., 2000, Dvir et al., 2002].

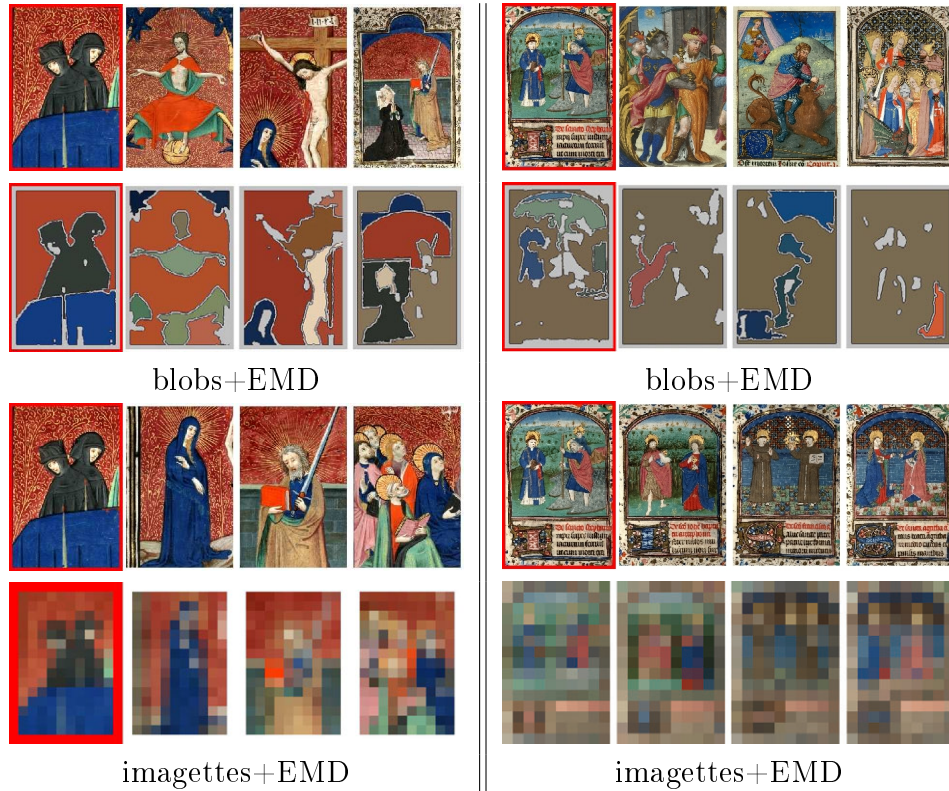


FIG. 3.17: Comparaison entre l'utilisation d'une segmentation BlobWorld (haut) ou d'imagettes (bas) en conjonction avec l'EMD. Une image requête (encadrée en rouge) est suivie des 3 résultats les plus proches sur la première ligne de chaque approche. Les blobs ainsi que les imagettes (10×15) sont visibles sur les secondes lignes de chaque approche. Nous voyons à gauche que les deux méthodes donnent des résultats de qualité assez similaire lorsque l'image représente des objets dont l'échelle des détails est faible. A droite en revanche, la méthode Blobworld sous-évalue le nombre de régions à segmenter, ce qui conduit à fusionner plusieurs régions ensembles. Ces erreurs se répercutent sur la qualité des résultats.

Quatre résultats sont visibles sur les Figures 3.17 et 3.18. Sur l'ensemble de nos tests, nous avons observé que la méthode *blobs+EMD* fonctionnent bien lorsque l'image représente un objet relativement homogène sur un fond lui aussi homogène. Dans ce cas, les résultats sont similaires à la méthode utilisant des imagettes. Remarquons que dans le

3.3. Expériences

cas *blobs+EMD*, les blobs n'ont pas tous le même poids. L'EMD permet donc de pouvoir séparer ces régions en plusieurs sous-régions pour les déplacer vers les cibles. Une telle situation est visible sur la Figure 3.17 à gauche où la région bleue peut être envoyée vers des régions non connexes.

Des erreurs de segmentation surviennent lorsque l'image est trop complexe ou texturée. Le nombre de blobs à extraire est généralement sous évalué. Ceci conduit à une fusion de plusieurs régions ayant des couleurs parfois très différentes, et crée par conséquent une seule région ayant une couleur moyenne non fidèle à l'image originale. Ces erreurs affectent énormément les résultats car de fausses mises en correspondance de grandes régions de teinte moyenne assez neutre sont alors facilement provoquées. Ce type d'erreurs est visible sur la Figure 3.17 à droite et sur la Figure 3.18.

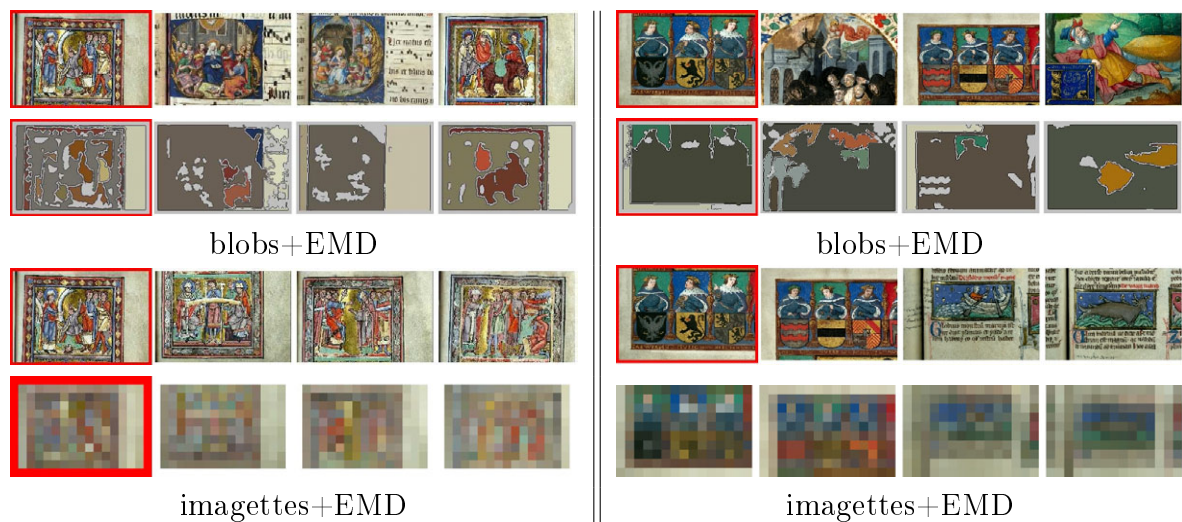


FIG. 3.18: Comparaison entre l'utilisation d'une segmentation (a) ou d'imagelets (b) en conjonction avec l'EMD. Même règles de présentation que sur la Figure 3.17. Ici la méthode Blobworld sous évalue le nombre de régions à segmenter, ce qui conduit à fusionner plusieurs régions ensemble. Ces erreurs se répercutent fortement sur la qualité des résultats.

Discussion

Lorsque l'EMD est utilisée avec une modélisation par régions obtenues par segmentation, chaque région est en général pondérée par sa surface normalisée par la surface totale de l'image [Liu et al., 2005, Jing et al., 2004, Dvir et al., 2002]. Cette segmentation peut sembler inutile car lors de la résolution du problème linéaire, aucune contrainte n'interdit à l'EMD de diviser les régions pour les envoyer vers plusieurs régions cibles si cela peut permettre de diminuer le coût de transport global. C'est pourquoi nous pensons dans cette thèse qu'utiliser cette distance en conjonction avec une méthode de segmentation, généralement coûteuse, est au mieux inutile et au pire dangereux.

Les méthodes de segmentation sont en effet bien souvent supervisées et difficiles à paramétrer. Elles sont toujours imparfaites. Et nous avons vu dans les expériences de cette section que les erreurs de segmentation se répercutent facilement sur les résultats d'indexation.

L'approche que nous proposons à base d'imagettes est plus simple car hormis la valeur de n (Section 3.3.4), elle ne nécessite pas le choix de paramètres de segmentation. Elle est plus robuste en réalisant une forme de sur-segmentation. Enfin, elle tire partie pleinement de l'EMD pour mesurer la similarité entre deux images. Notre méthode sursegmente l'image en petites régions uniformément réparties dans l'image, et nous laissons l'EMD résoudre le problème de la comparaison entre deux images.

3.3.6 Apport des imagettes bicolores

Pour comparer l'utilisation d'imagettes à une seule couleur par pixel ou bicolores, deux exemples sont présentés sur les Figure 3.19 et 3.20. Les exemples montrés sont représentatifs des résultats où une différence apparaît.

L'apport des imagettes bicolores, est sensiblement le même que celui qui consiste à augmenter n . Bien que nous ayons montré que le fait d'augmenter n avait un impact faible sur les bases que nous avons utilisé, nous pouvons imaginer que cela soit utile sur des bases d'images très texturées. Sur les images très texturées, ce type de modélisation est en effet plus fidèle au contenu des images originales. Le risque lorsque nous utilisons une valeur de n trop faible (ou des imagettes à une seule couleur par pixel) est de mettre en association des imagettes qui ont la même couleur moyenne en chaque pixel, mais que cette couleur moyenne soit issue d'une texture composée d'un mélange de deux couleurs différentes. Sur l'exemple de la Figure 3.19, la requête est une image très texturée. Nous voyons sur la première ligne qu'elle est mise en correspondance au deuxième et troisième résultat avec des images dont les couleurs sont plus homogènes et dont les moyenne sont assez proche de celles de la requête.

L'approche proposée par imagettes bicolores permet d'augmenter artificiellement n d'un facteur 2, sans atteindre la complexité calculatoire d'une imagette à une seule couleur de taille $2n$. Les temps de calculs ne sont que très légèrement supérieur. L'augmentation est d'environ 30% avec les imagettes bicolores, contre 800% en passant de n à $2n$. L'effet sur n et les résultats ne sont pas exactement les mêmes car les pixels bicolores imposent une contrainte spatiale aux couples de couleurs qu'ils représentent. En effet ces deux couleurs ne peuvent pas être séparées et envoyées vers des pixels différents comme ce serait le cas avec une augmentation directe de n sur des imagettes à une seule couleur par pixel. Nous voyons toutefois, comme à la Section 3.3.4 que le gain est limité lorsque n est déjà de l'ordre de la centaine de pixels. C'est pourquoi dans le reste des expériences, nous n'utiliserons que les imagettes à une seule couleur par pixel.

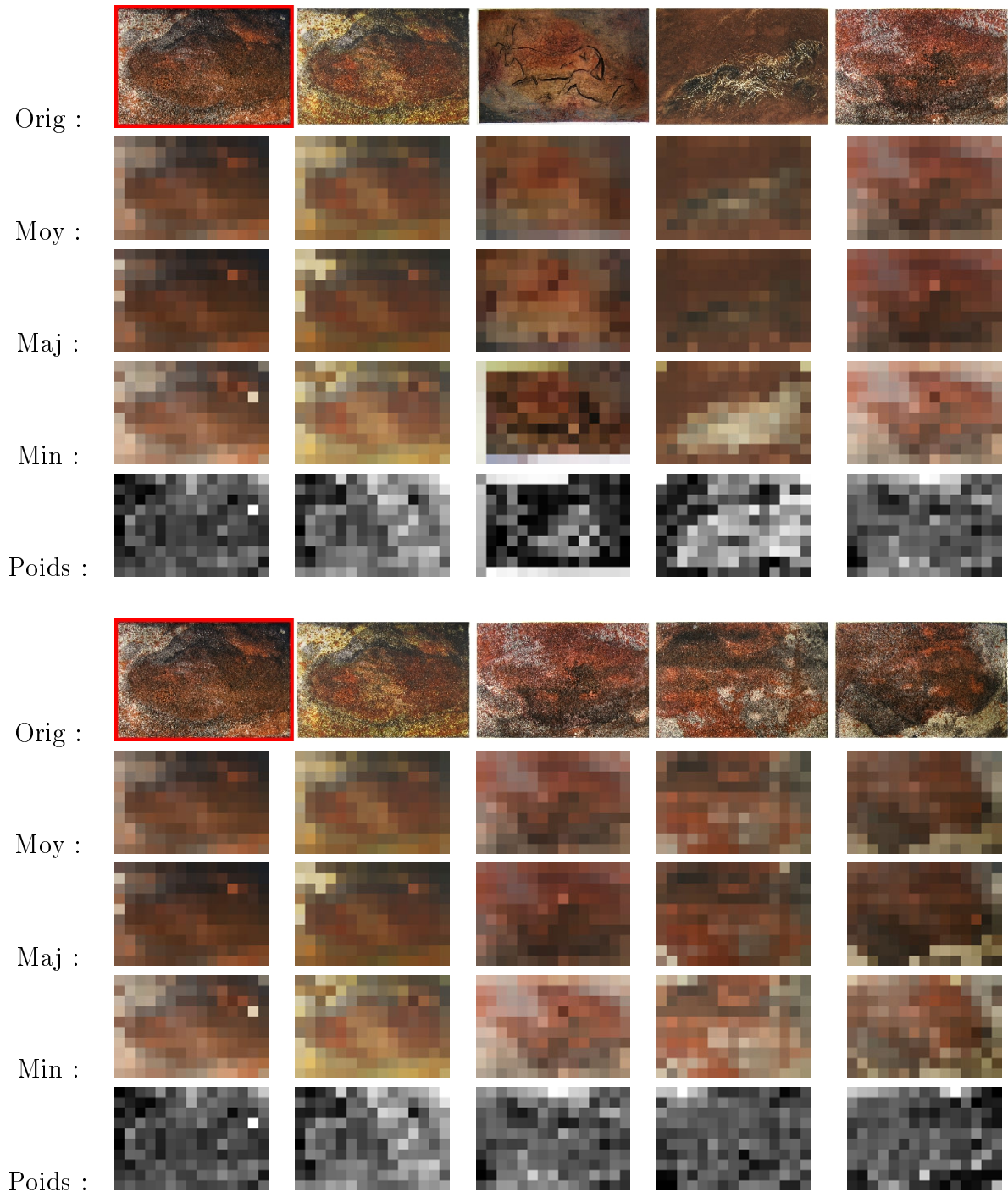


FIG. 3.19: Base ATCI et imagerie bicolore. Première ligne, images originales d'un résultat de requête utilisant les imagerie à une seule couleur par pixel. Les quatre lignes suivantes rappellent les imagerie à une seule couleur utilisées, et pour information, les imagerie bicolores des couleurs majoritaires, minoritaires, et leur poids (normalisés sur $[0, 255]$). Les cinq dernières lignes présentent le résultat en utilisant les imagerie bicolores.

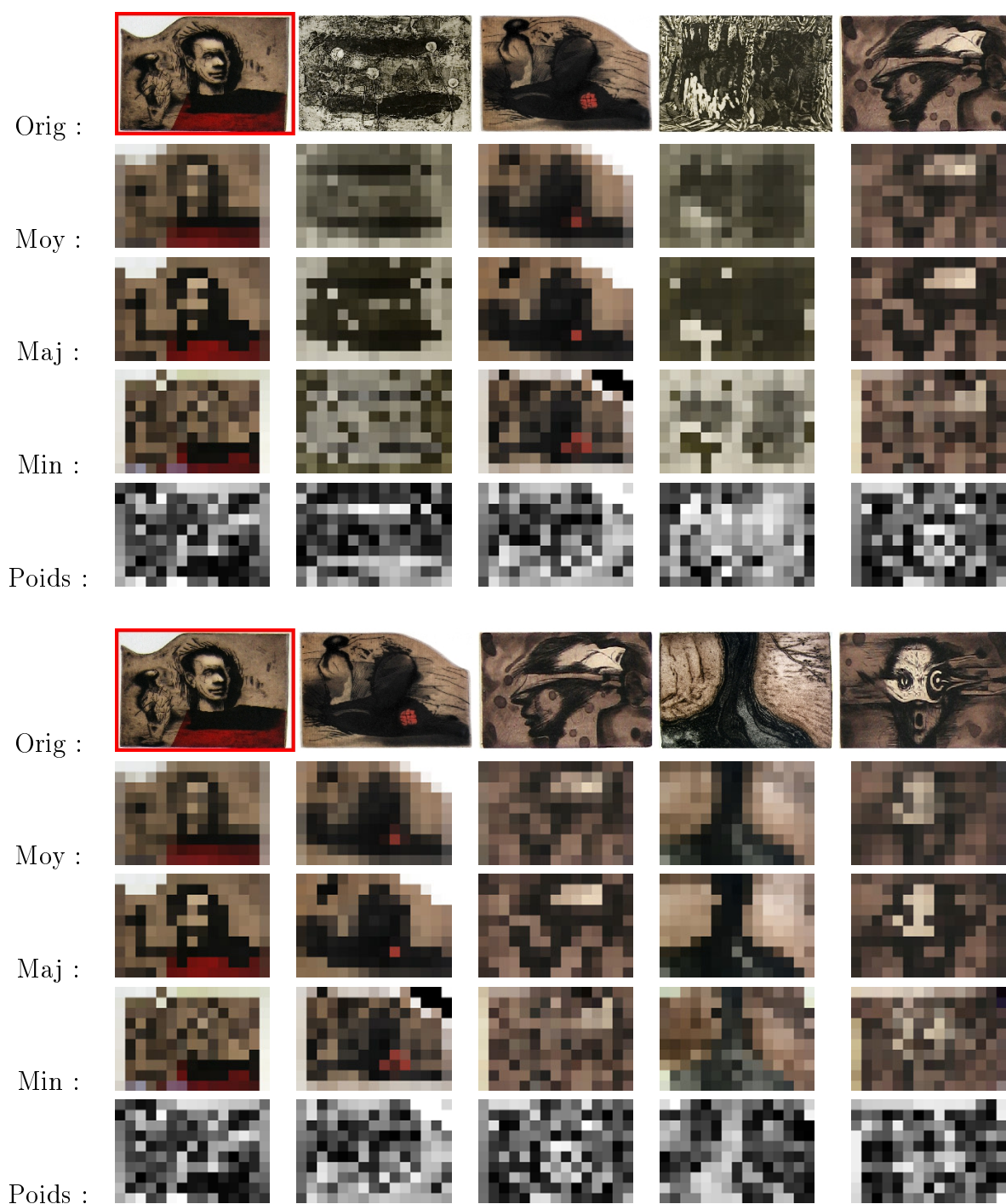


FIG. 3.20: Base ATCI et imagerie bicolore. Même organisation que sur la Figure 3.19

3.3.7 Base d'œuvres en noir et blanc

La base LCPD est uniquement composée de photographies en noir et blanc. Dans cette section nous montrons que pour ce type particulier de base d'images, il peut être bénéfique de réajuster le paramètre σ_c de la distance d_e . En effet pour les bases couleurs, $\sigma_c = 15$ dans l'espace *Lab* correspond à un changement de couleurs assez bien visible. La distance $\sigma_c = 1$ est théoriquement la différence minimale perceptible entre deux couleurs. Pour les images en noir et blanc, la distribution des couleurs se situent le long de l'axe des luminance L . Et notre oeil est moins sensible à un écart important entre deux gris par exemple, qu'entre un bleu et un rouge. Dans un cas nous interprétons souvent cela comme une variation de luminance ou d'éclairage, alors que dans l'autre, nous identifions plus clairement un changement de teinte intrinsèque.

C'est pourquoi dans le cadre des bases en noir et blanc, il semble plus naturel de choisir une valeur plus élevée de σ_c pour être moins sévère sur les variations de luminosité. Sur la Figure 3.21 sont présentés deux exemples significatifs avec $\sigma_c = 15$ en haut, et $\sigma_c = 50$ en bas. Sur le deuxième exemple en bas à droite, nous remarquons que le contraste du premier résultat est plus faible que la requête. Néanmoins nous y sommes peu sensibles du point de vue de l'OSC.

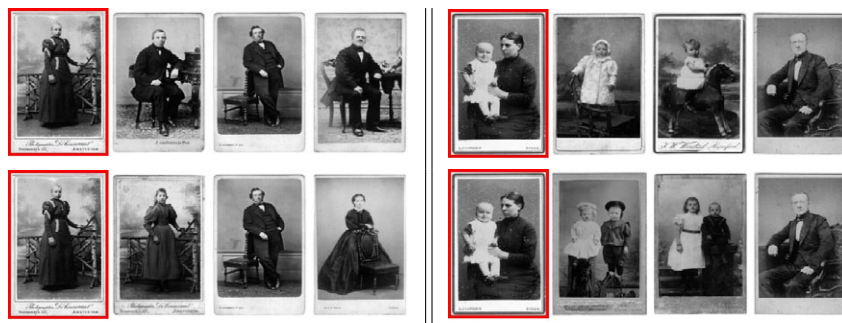


FIG. 3.21: Base LCPD. La valeur $\sigma_c = 15$ est utilisée sur la première ligne, et $\sigma_c = 50$ sur la seconde.

3.3.8 Plusieurs bases d'essais

Les Figure 3.22 à 3.32 présentées dans cette section utilisent des imagerie à une seule couleur, avec les paramètres méthodologiques suivants : $\alpha = 0.55$, σ_s égal au quart de la diagonale, $\sigma_c = 50$ pour la base LCPD et $\sigma_c = 15$ (Figure 3.31) pour toutes les autres bases et enfin $n = 150$ pixels. Sur l'ensemble des résultats, la requête est en haut à gauche encadrée en rouge suivie des résultats les plus proches par ordre croissant de la distance EMD. Comme la méthode proposée dans ce chapitre ne dispose par d'un critère d'arrêt sur l'EMD, le nombre de résultats présenté pour chaque requête varie manuellement selon

la base. Nous proposons au Chapitre suivant un seuil non supervisé pour répondre à ce problème.

3.4 Conclusion

En dehors de la section consacrée au choix de la valeur de α , nous n'avons pas utilisé de bases classées pour évaluer la méthode d'indexation proposée dans ce chapitre. L'organisation spatiale des couleurs est en effet un critère non sémantique et assez subjectif. Si nous avions voulu quantifier les performances de la méthode par des courbes précision-rappel, nous aurions eu besoin d'une base assez riche, et classée selon le critère de l'OSC. La petite base que nous avons utilisée pour l'étude de la valeur de α est constituée de 2×10 classes de séries d'images très distinctes où les OSC sont clairement distinguables. Dans une base réelle et riche, il est assez subjectif de classer l'ensemble de la base. De nombreux éléments peuvent être uniques, ou au contraire appartenir à plusieurs classes.

Notre but ici est donc de proposer plus une méthode de parcours de bases de données qu'une solution pour trouver des images strictement identiques selon un critère de type *recherche objective*. Nous nous plaçons toujours dans un angle d'étude de l'impact pictural d'une œuvre. Et nous illustrons l'efficacité d'une méthode de comparaison de l'OSC en mettant en correspondance des groupes d'œuvres similaires selon ce critère.

En effet l'OSC pouvant être indépendante du contenu descriptif, seule la présence de séries dans une base ayant à la fois la même composition et le même contenu permet de retrouver la même sémantique. Cela apparaît uniquement lorsque il y a un lien entre ces deux types de contenu. Dans les enluminures par exemple, Figure 3.24 troisième exemple à droite, le fait de représenter un personnage religieux dans un coin cerclé de vert a probablement une origine iconographique. Ces règles sémantiques de représentation se traduisent en fait de manière picturale. Il est donc logique que pour ces types d'exemples le contenu descriptif soit assez similaire. De même pour les exemples d'objets décoratifs (Figures 3.28 à 3.30), le photographe avait probablement pour consigne de toujours numériser les chaises sous un angle de $3/4$, et sur un fond blanc. Le pas de quantification n utilisé se traduit par une échelle faible de résolution, et de fait nous nous éloignons des détails qui bien souvent permettent de caractériser sémantiquement l'objet décrit lorsqu'ils appartiennent à des bases d'images au contenu hétéroclite.

Sur l'ensemble des résultats nous voyons que la similarité avec la requête selon l'OSC décroît de façon assez monotone. Néanmoins elle décroît très souvent jusqu'à un point où l'OSC devient plutôt *différente* à la requête que *similaire*. La frontière entre ces deux qualités varie d'un exemple à l'autre. Parfois, les 8 résultats montrés sont tous similaires, et nous souhaiterions voir quels sont les résultats suivants (par exemple sur l'ensemble des résultats de la Figure 3.31). Parfois nous avons l'impression que la base ne contient que très peu d'images similaires à la requête (par exemple sur la Figure 3.22, dernier résultat

3.4. Conclusion



FIG. 3.22: Base ATCI.

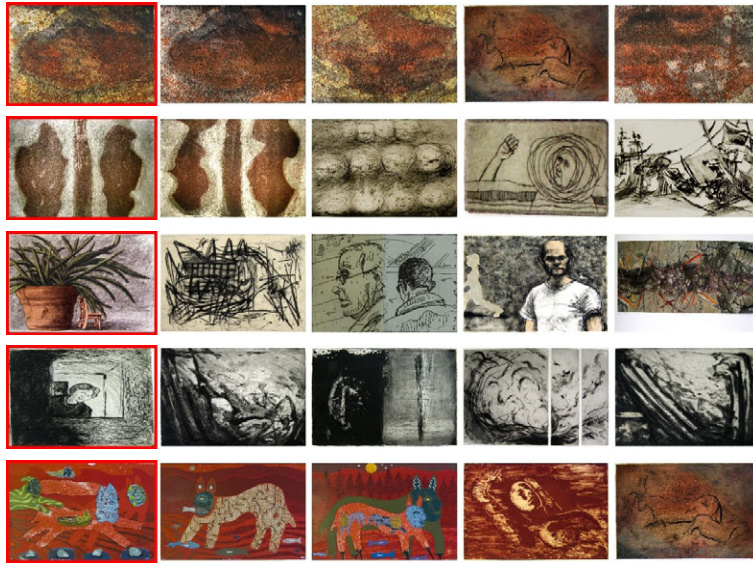


FIG. 3.23: Base ATCI.

en bas à droite). Pour répondre à ce problème de limite entre images similaires et images différentes, nous proposons dans le chapitre suivant un critère de décision non supervisé pour la mise en correspondance d'images selon leur OSC.

3.4. Conclusion



FIG. 3.24: Base IRHT.



FIG. 3.25: Base IRHT.

3.4. Conclusion



FIG. 3.26: Base IRHT.



FIG. 3.27: Base IRHT.

3.4. Conclusion

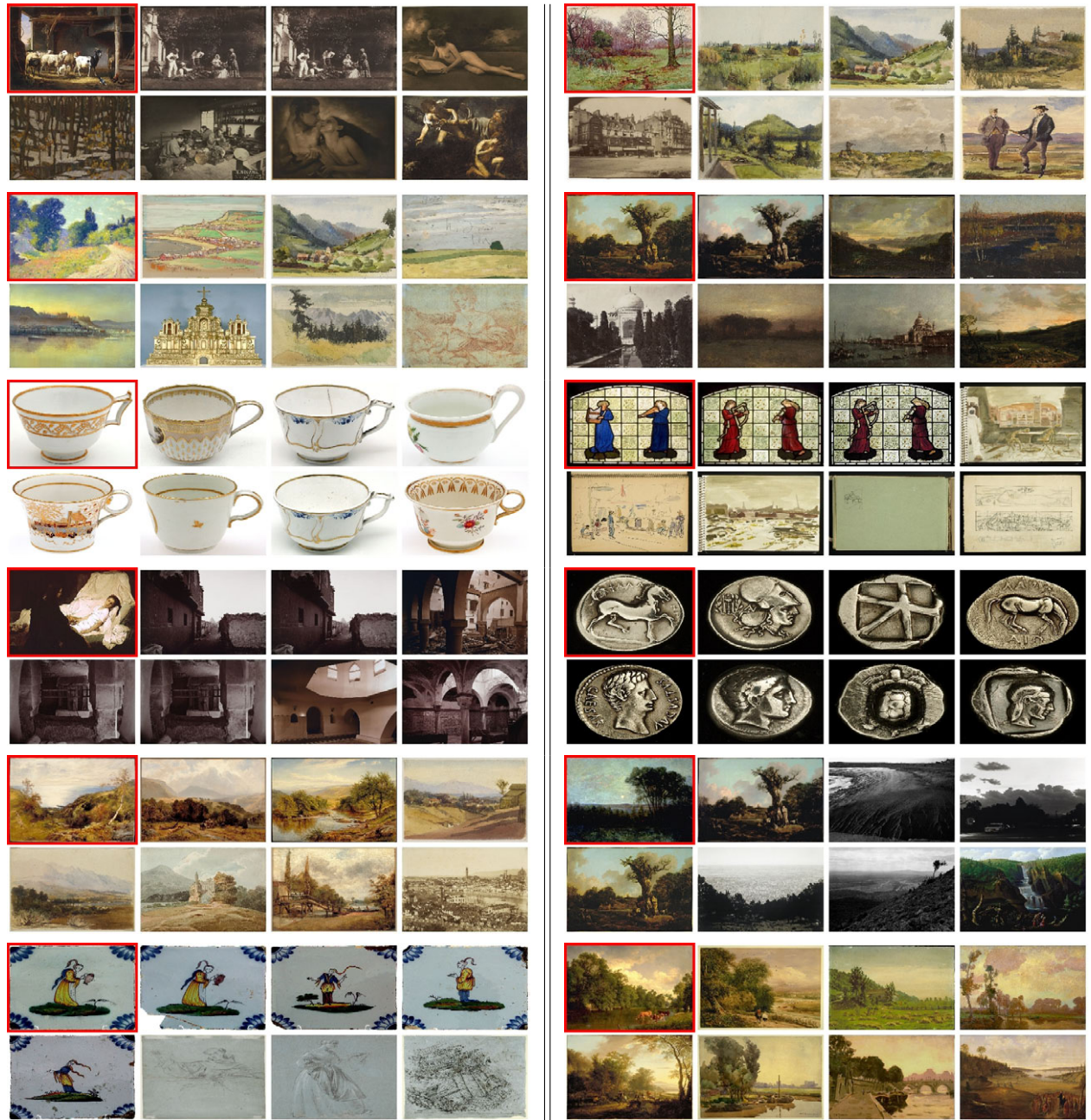


FIG. 3.28: Base MIXCA.

3.4. Conclusion

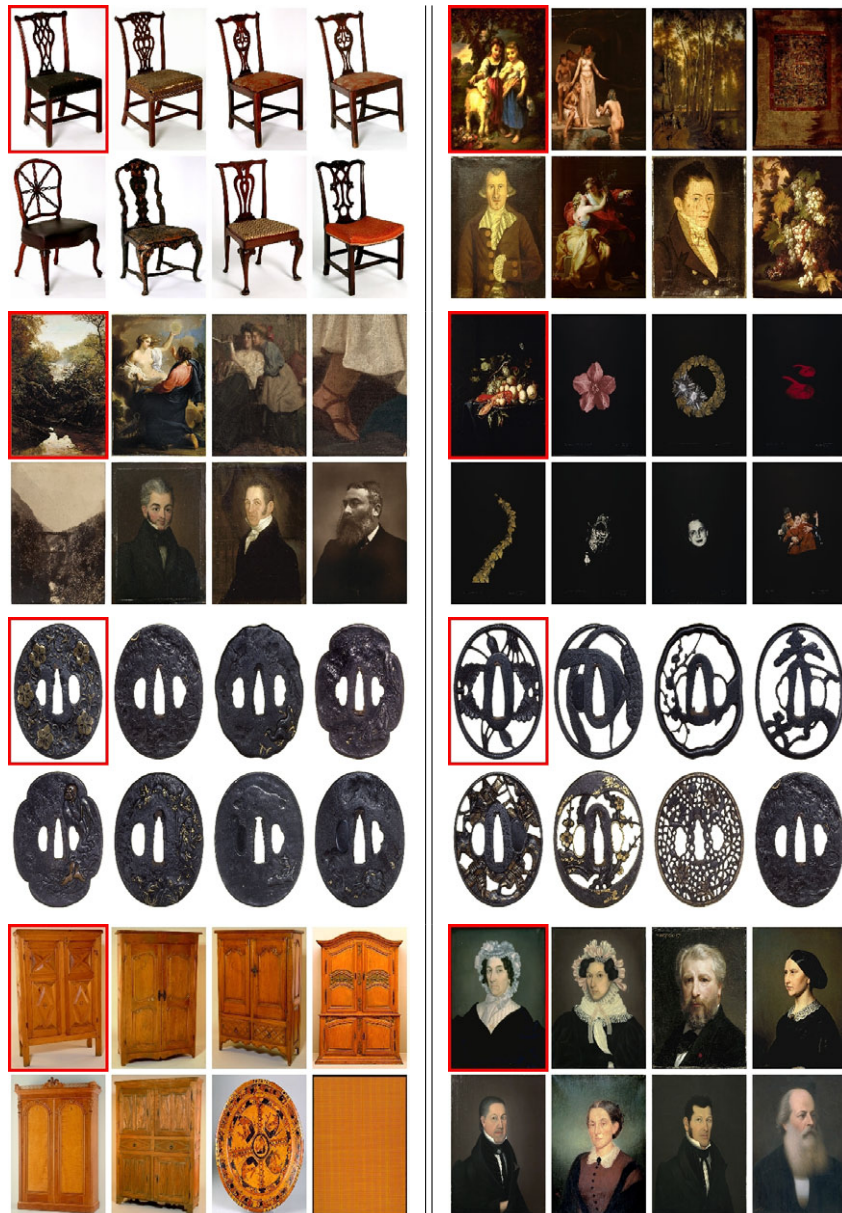


FIG. 3.30: Base MIXCA.



FIG. 3.31: Base LCPD

3.4. Conclusion

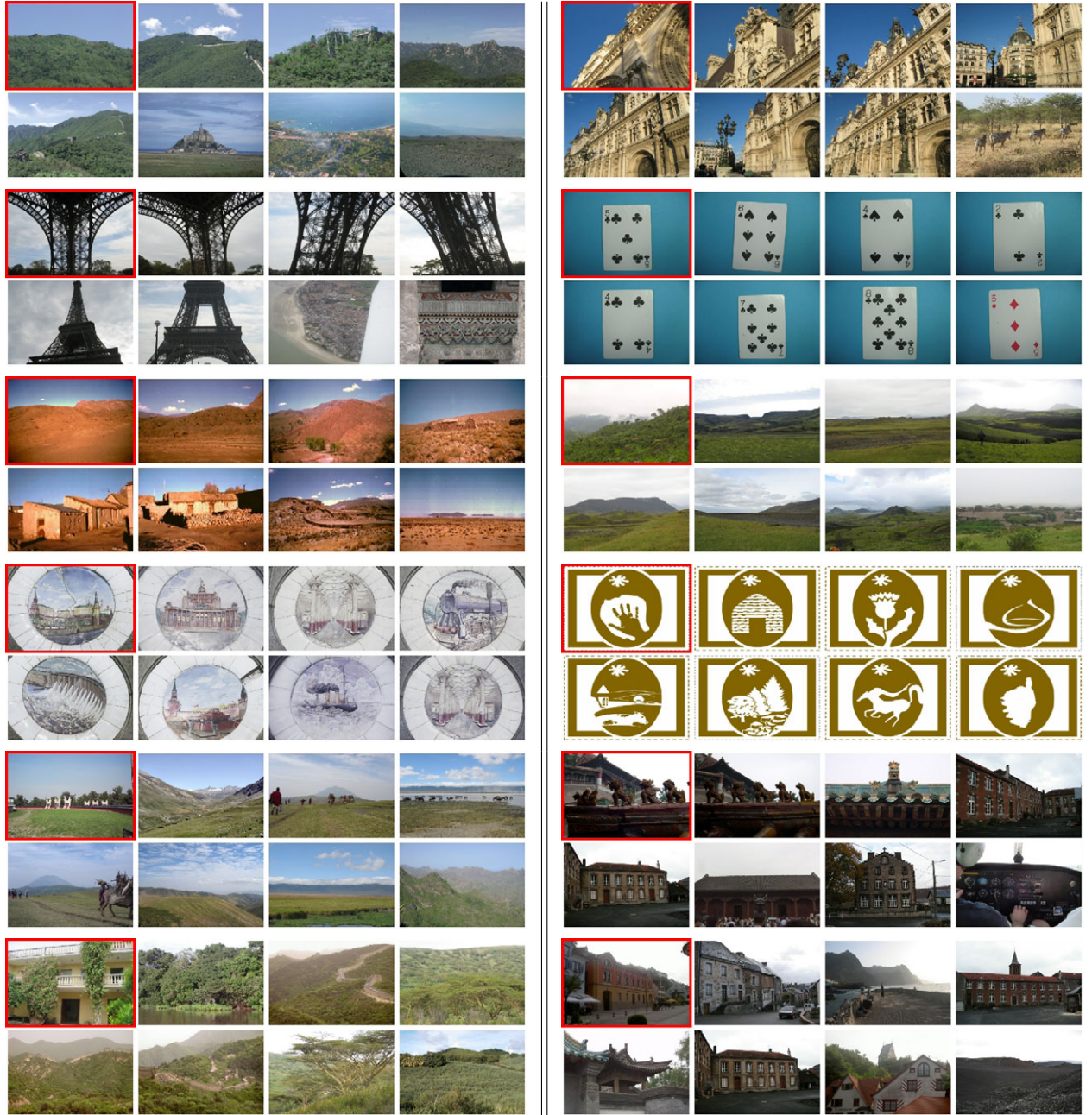


FIG. 3.32: Base CLIC.

Chapitre 4

Seuil automatique pour la recherche d'images selon l'OSC

Nous proposons dans ce chapitre une solution au problème pratique du choix du nombre d'images résultats à retourner à l'utilisateur lors de l'interrogation d'une base de données, dans le cadre du critère d'organisation spatiale des couleurs introduit au Chapitre 3. La solution la plus couramment adoptée est d'afficher un nombre fixe d'images résultats, triées par ordre décroissant de distance à la requête. Une alternative est de retourner la totalité de la base ordonnée selon la mesure de similarité, sous forme de pages à consulter. Cette approche présente un risque élevé de n'afficher qu'une partie des images similaires à la requête ou au contraire de noyer l'utilisateur sous un très grand nombre d'images non pertinentes. En pratique, le nombre d'images pertinentes selon la base consultée varie énormément d'une requête à l'autre. Une base d'images peut ne contenir aucune image similaire à une requête ou au contraire des centaines. Connaître cette information permet de sonder une base de données ce qui peut être très utile quand celle-ci atteint une taille telle qu'un parcours manuel n'est pas envisageable.

Ce chapitre propose un critère de similarité basé sur un schéma statistique *a contrario* qui s'adapte à chaque requête utilisée et à chaque base d'images consultée. La méthode présentée permet de compléter le schéma de recherche d'images par une étape de décision donnant une réponse à la question : *cette image candidate est-elle similaire à la requête* ? Il est important de distinguer ici la capacité à mesurer la distance entre deux images de celle à décider si deux images sont suffisamment proches pour pouvoir être considérée comme similaires. Après une brève présentation à la Section 4.1 de deux travaux précédents sur l'indexation de formes et d'images, nous présenterons à la Section 4.2 la méthodologie proposée qui est inspirée de ces travaux. La Section 4.3 présente plusieurs expériences permettant d'étudier et d'illustrer la méthode.

4.1 Méthodes a contrario pour la recherche dans des bases de données

4.1.1 Méthodes a contrario

[Attneave, 1954] a observé qu’aucune structure géométrique n’apparaît dans une image de bruit blanc. Une forme de réciproque de ce principe a été énoncée par [Lowe, 1985] qui suggère qu’une structure est perceptible à partir du moment où son apparition est improbable dans du bruit. Desolneux, Moisan et Morel ont formalisé ce principe de détection (appelé principe de Helmholtz) et l’ont appliqué à la définition des Gestalts, principes élémentaires de groupement qui gouvernent notre perception visuelle, [Desolneux et al., 2000, Desolneux et al., 2008]. Leur approche peut être résumée de la façon suivante. Tout d’abord, un événement (groupe) est défini. Puis les seuils de détection de cet événement sont fixés de manière à ce que l’événement soit très improbable dans le cadre d’un modèle aléatoire dit modèle de fond. Ce modèle de fond repose sur une hypothèse d’indépendance des constituants élémentaires à grouper. Les seuils de détection sont fixés de sorte que l’espérance du nombre de fausses détections (dans le cadre du modèle de fond) soit contrôlée. Leurs travaux ont depuis été appliqués à des problématiques aussi variées que la détection d’alignements ou de parallélisme [Desolneux et al., 2000], de points de fuite [Almansa et al., 2003], de contours contrastés [Desolneux et al., 2001], mais aussi l’analyse d’histogrammes [Desolneux et al., 2003], l’indexation de formes [Musé et al., 2006], d’images [Gousseau, 2003] et la mise en correspondance de descripteurs géométriques [Rabin et al., 2007]. Toutes ces méthodes portent le nom de méthodes *a contrario* car elles proposent de mesurer le degré de sûreté d’une reconnaissance par confrontation à un contexte aléatoire. Plusieurs approches connexes ont été proposées en détection de cibles [Grimson and Huttenlocher, 1991, Olson, 1998, Lindenbaum, 1997], sans toutefois proposer d’approche systématique de la détection de structures visuelles.

4.1.2 Application à la recherche dans des bases de données

L’application de telles méthodes à la recherche dans une base conduisant à l’estimation automatique d’un seuil sur une mesure de similarité a déjà été étudiée pour l’indexation de formes par [Musé et al., 2003, Musé et al., 2006], et la composition de photographies en noir et blanc par [Gousseau, 2003] (voir la Section 3.1.5 dans le Chapitre 3).

Dans ces approches la méthodologie consiste tout d’abord à définir un modèle générique aléatoire des objets étudiés (appelé modèle de fond, modèle de bruit, ou *background model* en anglais). C’est ce modèle de fond qui permet d’estimer la distance au delà de laquelle il est peu probable qu’un candidat soit similaire à la requête. L’idée générale est que le modèle de fond repose sur une hypothèse d’indépendance entre les constituants élémentaires de l’objet recherché. Les seuils de mise en correspondance sont alors calculés

de façon à ce que la mise en correspondance avec une réalisation du modèle de fond soit très improbable.

Dans le cadre de la recherche de formes, [Musé et al., 2006] choisissent un modèle de fond qui repose sur une hypothèse d'indépendance de morceaux de lignes. Les marginales de ces morceaux sont apprises sur la base. Pour l'organisation spatiale d'une image en noir et blanc, [Gousseau, 2003] utilise des réalisations du modèle *feuilles mortes* de la morphologie mathématiques qui consiste en la superposition d'objets aléatoires 2D indépendants [Matheron, 1968, Bordenave et al., 2006]. La méthode proposée dans ce chapitre s'inspire aussi de ce type de modèle.

4.2 Critère de décision non supervisé

Nous présentons dans cette section les caractéristiques de l'approche *a contrario* que nous proposons dans le cadre de l'indexation d'images selon l'organisation spatiale des couleurs (OSC). Cette approche permet de répondre à la question suivante : disposant d'une image requête, quelles sont les images d'une bases de données qui lui sont similaires ? En pratique, nous verrons que la réponse se traduira par un seuil sur la distance EMD qui permettra de retourner à l'utilisateur un nombre d'images qui varie selon la requête. La contribution de ce chapitre peut alors se résumer au fait que nous proposons une méthode qui ajuste ce seuil de manière automatique à chaque requête et à chaque base d'images.

Après avoir défini le problème sous la forme d'un test statistique à la Section 4.2.1, nous définirons la notion de mise en correspondance ε -significative à la Section 4.2.2. Le modèle générique aléatoire (*modèle de fond*) utilisé et ses caractéristiques seront précisés à la Section 4.2.3.

4.2.1 Une décision sous forme de test statistique

Pour deux imageries Q et T , la distance EMD entre elles sera notée $d(Q, T)$ (voir la Section 3.2.1 du Chapitre 3). L'objectif principal est de pouvoir calculer un seuil δ_Q sur $d(Q, \cdot)$ qui soit capable de délimiter la frontière entre images similaires et images non similaires à Q . Idéalement, on souhaiterait que δ_Q soit fixé de manière non supervisée. Nous proposons d'utiliser le même schéma général *a contrario* que dans [Musé et al., 2006] reposant sur un test statistique à deux hypothèses.

Pour une image requête Q et une image T aléatoire, nous construisons un test statistique reposant sur un seuil δ_Q sur la distance $d(Q, T)$ entre ces deux images. Les deux hypothèses H_0 et H_1 sont :

$$\begin{aligned} H_1 &: "T \text{ est similaire à } Q", \\ H_0 &: "T \text{ suit le modèle de fond }". \end{aligned}$$

En pratique, la notion “est similaire à” est donc définie par le rejet de H_0 . Le modèle de fond de l’hypothèse H_0 est un modèle générique aléatoire de l’OSC. Autrement dit, le test statistique que nous construisons considère qu’une image aléatoire T est similaire à Q lorsqu’elle n’est pas plus proche de Q que δ_Q . L’idée générale est de fixer δ_Q en contrôlant l’erreur de première espèce, à savoir $Pr_{H_0} = Pr(d(Q, T) \leq \delta | H_0)$. Cette probabilité est aussi appelée probabilité de fausses alarmes. Contrôler l’erreur de seconde espèce $Pr_{H_1} = Pr(d(Q, T) > \delta | H_1)$ nécessiterait de s’appuyer sur un modèle *a priori* de similarité. Dans le cadre de l’OSC, ceci imposerait de s’appuyer sur des modèles d’OSC généraux de type *compositions verticales, horizontales, centrées, ...*. Cette solution n’est pas envisageable car l’OSC ne peut être réduite à un nombre fini de modèles. C’est également pour cette raison que nous n’utilisons pas une approche Bayésienne.

La définition de la similarité entre images se fait *a contrario* dans le sens où elle repose sur le rejet d’un modèle de fond que nous noterons par la suite $\mathcal{M}$. Nous définirons ce modèle de fond $\mathcal{M}$ pour l’OSC à la Section 4.2.3.

4.2.2 Mise en correspondance ε -significative

Dans toute la suite, nous notons Q une imagerie requête et $\mathcal{B} = \{T_1, \dots, T_{m_B}\}$ une base d’images que nous souhaitons consulter, composée de m_B images candidats.

Définition : Nous dirons qu’une image T_i forme une mise en correspondance ε -significative avec Q si $d(Q, T_i) \leq \delta_Q$ où δ_Q est défini par :

$$\delta_Q = \sup \left\{ \delta : Pr_{H_0} (d(Q, T) \leq \delta) \leq \frac{\varepsilon}{m_B} \right\}. \quad (4.1)$$

Dans cette formule, Pr_{H_0} désigne le fait que l’image aléatoire T est distribuée selon $\mathcal{M}$. Nous allons maintenant expliquer le choix de cette définition. La croissance de Pr_{H_0} avec δ conduit au fait qu’en fixant une borne supérieure à Pr_{H_0} , il est possible de trouver une borne supérieure à la distance $d(Q, T_i)$. Alors, une distance observée sous le seuil δ_Q nous permet d’avoir l’assurance que la probabilité d’une fausse alarme est bornée par $\frac{\varepsilon}{m_B}$.

La borne $\frac{\varepsilon}{m_B}$ a été choisie car elle permet d’interpréter les mise en correspondances ε -significatives d’une manière relativement intuitive. Si toutes les images T_i de la base de taille m_B suivent le modèle de fond $\mathcal{M}$, alors le nombre moyen de mises en correspondance ε -significatives est plus petit que ε . En effet, si on suppose que les $T_1, \dots, T_{m_B}$ suivent $\mathcal{M}$, alors par linéarité de l’espérance mathématique E :

$$\begin{aligned}
E(\text{du nombre de } T_i \text{ tel que } d(Q, T_i) \leq \delta_Q) \\
&\leq \sum_{1 \leq i \leq m_{\mathcal{B}}} Pr_{H_0}(d(Q, T_i) \leq \delta_Q) \\
&\leq m_{\mathcal{B}}\varepsilon/m_{\mathcal{B}} = \varepsilon.
\end{aligned}$$

Remarque : autrement dit, la notion de mise en correspondance ε -significative sur $\mathcal{B}$ est définie de telle manière qu'en moyenne, il y a moins de ε imagerie similaires à Q dans une base de $m_{\mathcal{B}}$ imagerie aléatoires distribuées selon $\mathcal{M}$. Pour avoir un contrôle effectif du nombre de fausses alarmes, il est donc très important que le modèle $\mathcal{M}$ rende compte du contenu de la base.

En pratique et comme dans toute cette thèse, la valeur $\varepsilon = 1$ sera toujours utilisée. Dans toute la suite une mise en correspondance d'imagerie 1-significatives sera simplement appelée une mise en correspondance significative.

4.2.3 Modèle de fond

Modèles non structurés

Il reste maintenant à définir précisément le modèle de fond $\mathcal{M}$ utilisé pour une imagerie. Suivant le principe général des approches *a contrario*, le modèle $\mathcal{M}$ repose sur une hypothèse d'indépendance des constituants de la structure d'intérêt. La solution la plus simple et directe est d'utiliser un modèle de bruit blanc sur les couleurs des pixels. Dans ce modèle les couleurs des pixels d'une imagerie 10×15 sont choisies de manière aléatoire, indépendantes et uniformément distribuées dans l'espace couleur *Lab*.

Dans les tests que nous avons effectué, cette approche donnait des résultats très insatisfaisants. Le seuil δ_Q calculé était beaucoup trop permissif. Une première cause est que la distribution des couleurs sur l'ensemble d'une base $\mathcal{B}$ d'imagerie réelle est rarement homogène. Si le modèle de fond $\mathcal{M}$ ne respecte pas la même distribution des couleurs que celle de $\mathcal{B}$, alors la probabilité de fausse alarme est sous-évaluée. En s'inspirant de [Desolneux et al., 2001, Musé et al., 2006], nous avons donc ensuite testé un modèle de fond où les couleur des pixels sont toujours choisies de manière indépendante mais distribuées selon les même marginales de couleurs que pour la base $\mathcal{B}$.

Cette approche donne des seuils encore trop permissifs. Notre interprétation est la suivante. Même lorsqu'elles sont perceptuellement différentes selon l'OSC, les imagerie peuvent avoir une certaine structure commune. En effet, quelque soit l'OSC, les imagerie

ont généralement un contenu spatial structuré en régions homogènes, et non pas sous la forme de pixels indépendants. Ceci a pour conséquence que deux imagerie réelles dont l'OSC est différente vont souvent être plus proches entre elles qu'elles ne le seraient d'un bruit blanc.

Un superposition aléatoire de disques

C'est pourquoi nous avons choisi d'utiliser un modèle de fond d'imagerie spatialement structuré, comme dans [Gousseau, 2003]. Le modèle choisi est un modèle feuilles mortes [Matheron, 1968, Gousseau and Roueff, 2006]. Ce modèle consiste en la superposition d'objets aléatoires indépendamment et identiquement distribués sur l'imagerie et dont la couleur est uniformément distribuée selon les marginales de couleur de $\mathcal{B}$. Dans le cadre de cette thèse nous avons choisi d'utiliser un modèle d'objets simples et isotropes : des disques de couleur homogène.

Il a été démontré expérimentalement que ces modèles étaient bien adaptés à la structure des images naturelles lorsqu'on utilisait une loi de distribution en puissance pour fixer la taille des objets [Alvarez et al., 1999, Lee et al., 2001]. Les rayons des disques superposés sur l'imagerie sont choisis en suivant les valeurs de variables aléatoires distribuées de manière indépendante et identiques avec une densité $f(r) \propto r^{-\gamma}$. Trois paramètres fixent alors ce modèle : le paramètre d'échelle γ , et les rayons minimaux et maximaux des disques¹ r_0 et r_1 .

Dans les tests nous choisirons $\gamma = 3$ qui est une valeur typique pour les images naturelles. Pour les rayons, il est assez intuitifs de fixer $r_0 = 1$ et r_1 égal à la plus grande dimension de l'image pour permettre l'apparition rare mais possible, d'imagerie complètement uniforme. Chaque disque est coloré indépendamment de sa taille et des autres disques. Ceci implique que la distribution marginale des couleurs du modèle de fond suit celle des couleurs des disques. Cette distribution doit aussi être la même que celle de base $\mathcal{B}$ comme nous l'avons explicité plus haut. Cette distribution sera donc apprise d'abord sur la base $\mathcal{B}$, puis réutilisée pour fixer celle du modèle de fond. Des exemples d'imagerie produites par le modèle de fond que nous venons de décrire sont présentées sur la Figure 4.1.

¹Il est possible de fixer automatiquement le paramètre r_0 en considérant un modèle limite où ce paramètre d'échelle est lié à l'échelle des petites structures dans les images naturelles [Gousseau and Roueff, 2006]. Néanmoins, cette approche est inutilement trop complexe pour être justifiée dans le cadre de l'OSC.



FIG. 4.1: Exemples d'imagettes produits par le modèle de fond $\mathcal{M}$. Les couleurs des disques superposés suivent les marginales des distributions des couleurs de la base d'enluminures IRHT.

4.3 Expériences

Dans cette section, nous précisons tout d'abord comment le seuil automatique est concrètement estimé numériquement (Section 4.3.1). Puis nous effectuons une étude de la variabilité du seuil (Section 4.3.3). Enfin nous montrons un grand nombre de résultats sur les différentes bases de données déjà rencontrées au chapitre précédent (Section 4.3.5).

4.3.1 Estimation numérique de δ_Q

Nous avons introduit et proposé à la Section 4.2 une méthode permettant le calcul d'un seuil automatique sur la mesure de similarité $d(\cdot, \cdot)$ proposée au Chapitre 3 et fondé sur une distance de transport. Nous nous plaçons dans le contexte d'une recherche d'images par l'exemple, et en pratique nous souhaitons donc interroger une base d'images $\mathcal{B}$ de taille $m_{\mathcal{B}}$ avec une imagette requête Q . Le seuil δ_Q nous permet de décider la mise en correspondance significative d'une imagette T avec Q dès que $d(Q, T) \leq \delta_Q$. Nous avons vu à la Section 4.2 que δ_Q dépend à la fois de Q et d'un paramètre de tolérance ε (voir Equation (4.1)). Les calculs permettant d'obtenir cette probabilité reposent sur une étude géométrique du modèle feuilles-mortes et sont particulièrement complexes. Cette probabilité est donc estimée à partir de simulations numériques en observant la fréquence d'événements $d(Q, T) \leq \delta$, pour $\delta > 0$ sur un grand nombre de réalisations du modèle de fond. En pratique, nous construisons tout d'abord une base de $n_{bg} \times m_{\mathcal{B}}$ imagettes M_j suivant le modèle de fond $\mathcal{M}$, où n_{bg} est un entier. Puis nous approximations δ_Q par le quantile d'ordre $\varepsilon/m_{\mathcal{B}}$ de l'ensemble des distances $\{d(Q, M_j)\}_{j=1, \dots, n_{bg} \times m_{\mathcal{B}}}$. En notant M_Q la n_{bg}^{eme} plus proche imagette de Q parmi $\{M_1, \dots, M_{n_{bg} \times m_{\mathcal{B}}}\}$, on obtient $\delta_Q \approx d(Q, M_Q)$.

4.3.2 Influence de la taille de la base d'imagettes bruitées

Les seuils de mise en correspondance δ_Q étant appris sur un nombre $n_{bg} \times m_{\mathcal{B}}$ de réalisations de $\mathcal{M}$, nous devons nous poser la question de l'influence de n_{bg} , et de sa valeur pour permettre une estimation stable de δ_Q . Pour répondre à cette question nous

avons fait l'expérience suivante. Nous avons construit des bases de données d'imagettes bruitées de taille $n_{bg} \times m_B$ avec $n_{bg} = 1, \dots, 9$. Nous avons simulé 10 bases différentes pour chaque n_{bg} . Pour une imagette requête Q , le nombre d'images significativement similaires dans B selon δ_Q est noté r_Q .

La Figure 4.2 à gauche montre l'écart-type de r_Q normalisé par sa valeur en $n_{bg} = 1$ en fonction de n_{bg} sur les 10 bases de taille $n_{bg} \times m_B$ pour différentes requêtes. Pour chacune de ces imagettes requêtes, nous avons donc fait 100 tests, en faisant varier $n_{bg} = 1, \dots, 9$ pour chaque base de bruit. A droite de la Figure 4.2 est visible l'écart type normalisé du seuil δ_Q . La Figure 4.2 montre cette expérience faite avec la base IRHT (ligne du haut) et la base MIXCA qui est 15 fois plus grande (ligne du bas). Les différentes bases sont présentées en détails dans l'Annexe A.

Selon ces résultats, $n_{bg} = 3$ est un compromis acceptable entre stabilité des résultats et coût calculatoire. Par ailleurs, nous avons également mesuré que l'écart type sur le rang est d'environ 15% en moyenne selon les requêtes. Ceci indique qu'avec $n_{bg} = 3$, pour une requête ayant environ 20 résultats, la variabilité du nombre d'images retournées est de l'ordre de 3.

Sur la Figure 4.3, nous avons tracé les moyennes sur les différentes bases de bruit des rangs et seuils, normalisés par leur valeur en $n_{bg} = 1$ (pour ajuster toutes les courbes au même niveau). Nous voyons sur cette figure que le rang et le seuil décroissent avec n_{bg} . Ceci signifie qu'en augmentant n_{bg} il existe une imagette de bruit plus proche de la requête, et ce, bien que nous prenions l'imagette correspondant au quantile d'ordre ε/m_B . Nous observons néanmoins que cette décroissance se stabilise assez rapidement. A nouveau, notre choix de $n_{bg} = 3$ repose sur un compromis avec les coûts calculatoires.

4.3.3 Variabilité du seuil

Dans cette section nous étudions la variabilité du seuil δ_Q et du nombre d'imagettes similaire r_Q sur l'ensemble des bases utilisées. Nous utilisons $n_{bg} = 3$ dans toute cette section. L'histogramme bidimensionnel des couples (δ_Q, r_Q) est présenté sur la Figure 4.4 pour les base ATCI, IRHT, MIXCA et LCPD.

Cette expérience nous permet d'illustrer la très grande variabilité des couples (δ_Q, r_Q) qui s'adaptent à chaque requête. Le seuil automatique proposé est donc très différent d'un seuil fixe sur la mesure de similarité ou plus directement sur le nombre d'images retournées. Ces histogrammes montrent aussi la faible corrélation entre δ_Q et r_Q . Une valeur élevée de r_Q indique la présence d'une grande série d'images similaires dans la base, comme celles visibles sur les Figures 4.13 et 3.31. Nous voyons sur cet histogramme que ceci peut survenir avec des valeurs très variables de δ_Q puisque cela dépend essentiellement du contenu de la base B . Nous observons aussi la présence de très grandes séries d'image similaires dans la base LCPD. Cette base est composée uniquement de portraits en noir

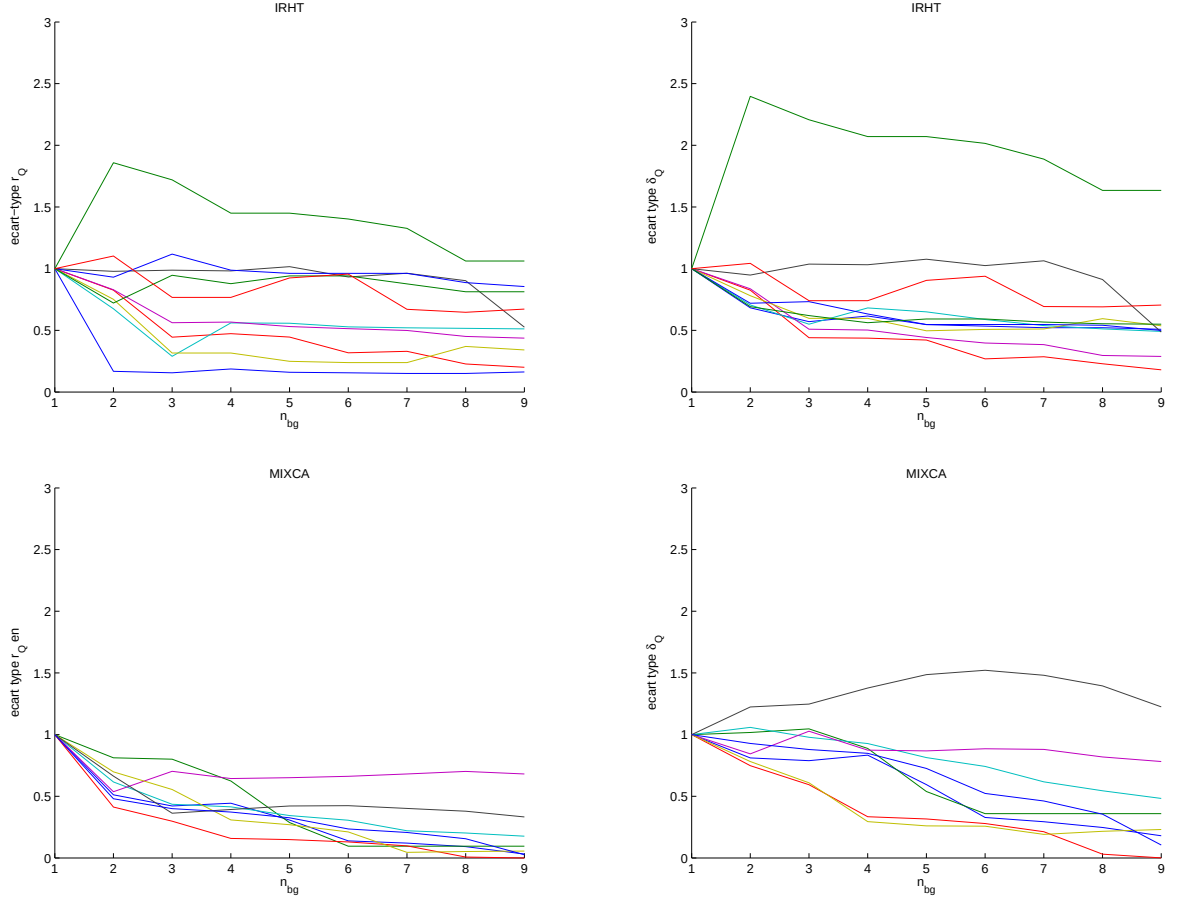


FIG. 4.2: Ecart-type du rang r_Q et du seuil δ_Q sur les bases IRHT (haut) et MIXCA (bas). A gauche, chaque courbe correspond à une imagerie requête Q au sein d'une base de taille m_B . L'écart type du rang r_Q de la dernière imagerie significativement similaire selon δ_Q normalisé par sa valeur en $n_{bg} = 1$ est calculé pour un ensemble de 10 bases de bruits de taille $m_B \times n_{bg}$, pour chaque $n_{bg} = 1, \dots, 9$. A droite, même courbes mais pour l'écart type du seuil δ_Q .

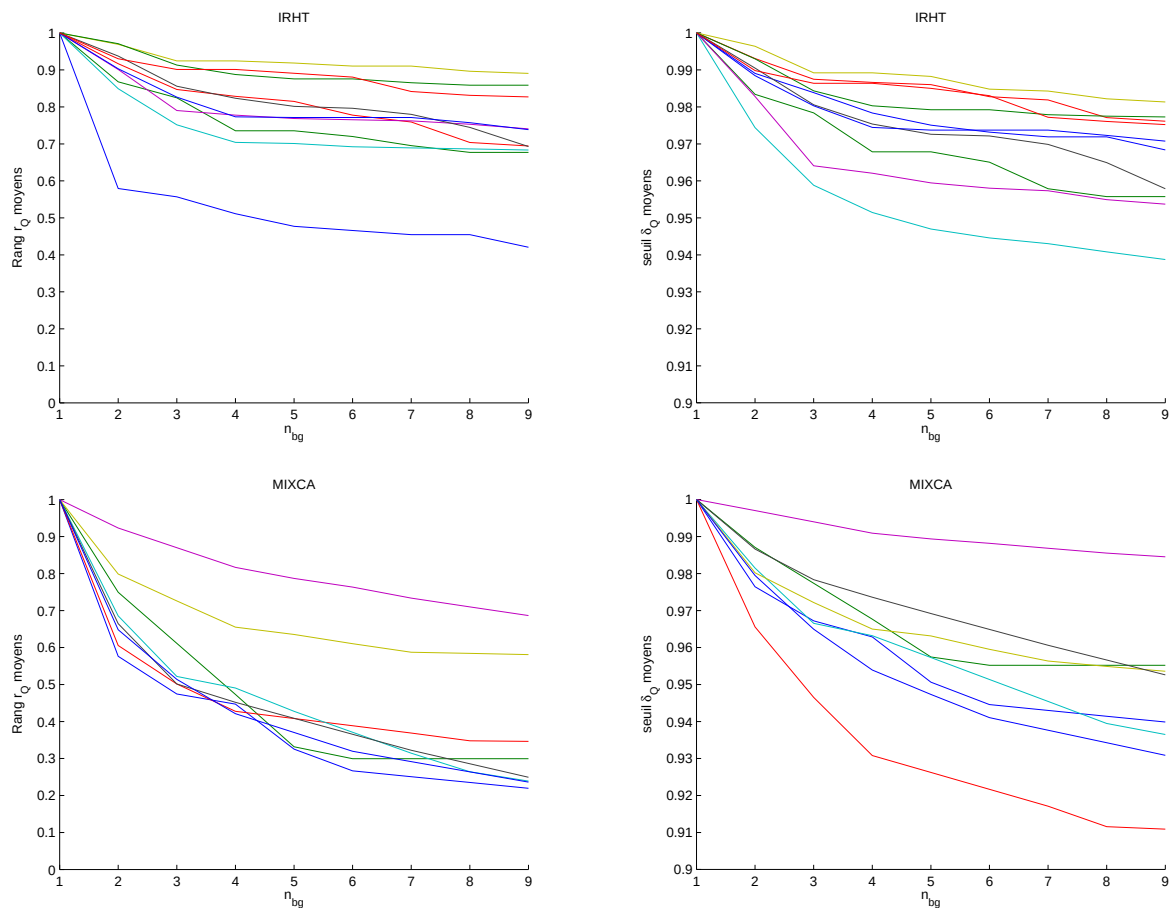


FIG. 4.3: Moyenne du rang r_Q et du seuil δ_Q sur la base IRHT (haut) et MIXCA (bas). A gauche, chaque courbe correspond à une imagerie requête Q au sein d'une base de taille m_B . La moyenne du rang r_Q de la dernière imagerie significativement similaire selon δ_Q normalisé par sa valeur en $n_{bg} = 1$ est calculé pour un ensemble de 10 bases de bruits de taille $m_B \times n_{bg}$, pour chaque $n_{bg} = 1, \dots, 9$. A droite, les mêmes courbes sont tracées pour le seuil δ_Q . Nous voyons que le seuil et le rang décroissent quand n_{bg} augmente.

et blanc selon certaines compositions choisies par le photographe.

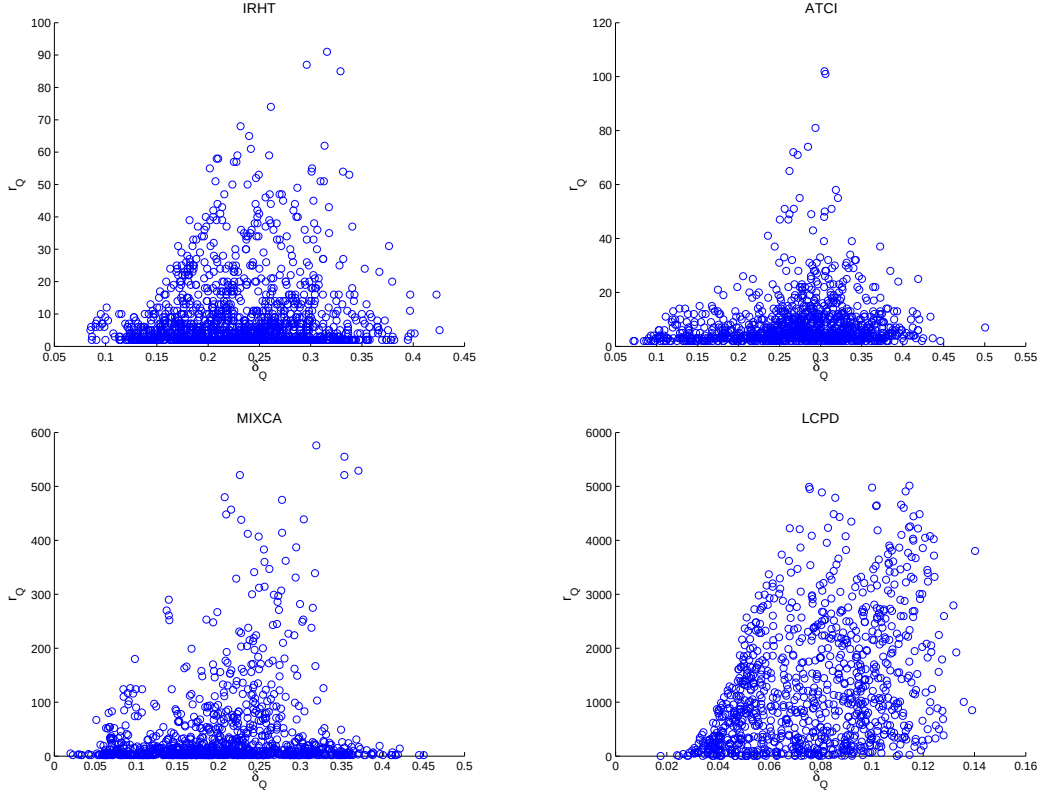


FIG. 4.4: Histogramme bidimensionnel des couples (δ_Q, r_Q) pour les 1500 requêtes de la base IRHT (en haut à gauche), pour les 1000 requêtes de la base ATCI (en haut à droite), 1000 requêtes choisies aléatoirement dans la base MIXCA (en bas à gauche) et 1000 requêtes choisies aléatoirement dans la base LCPD (en bas à droite).

4.3.4 Optimisation de l'algorithme de calcul de l'EMD

Rappelons que la résolution classique du problème d'affectation pour calculer la distance EMD $d(\mathcal{F}^1, \mathcal{F}^2)$ entre deux imagerie $\mathcal{F}^1$ et $\mathcal{F}^2$ (Section 3.2.1) est réalisée en deux étapes. Premièrement il faut calculer la matrice des coûts $(c_{ij})_{i,j} = \{d_e(f_i^1, f_j^2)_{1 \leq i,j \leq n}\}$. Puis la méthode de résolution de Kuhn-Munkres trouve le chemin d'assignement de coût global minimal dans cette matrice par permutation, et soustraction des lignes et colonnes (pour plus de détails sur cette seconde étape voir [Kuhn, 1955]).

La méthode la plus directe consisterait à calculer un total de $(n_{bg} + 1)m_{\mathcal{B}}$ distances EMD, correspondant à l'ensemble des distances EMD entre Q et l'union de la base $\mathcal{B}$ avec celle des réalisations du modèle $\mathcal{M}$ de taille $n_{bg}m_{\mathcal{B}}$. Le seuil automatique δ_Q présenté dans

ce chapitre permet de réduire ce coûts calculatoire. En effet une fois que la matrice des coûts a été calculée, un minorant de $d(\mathcal{F}^1, \mathcal{F}^2)$ peut être facilement trouvé :

$$d(\mathcal{F}^1, \mathcal{F}^2) \geq \sum_{i=1}^n \min_{1 \leq j \leq n} \{c_{ij}\}. \quad (4.2)$$

Si ce minorant est supérieur à δ_Q , il est inutile de continuer le calcul de l'EMD jusqu'à la résolution du problème linéaire par la méthode Kuhn-Munkres (sauf si l'on souhaite connaître les distances avec les imagerie non similaires à la requête...). Le calcul du minorant peut aussi se faire au fur et à mesure du calcul de la matrice des coûts. Nous pouvons en arrêter le calcul dès qu'il existe k tel que :

$$\sum_{i=1}^k \min_{1 \leq j \leq n} \{c_{ij}\} > \delta_Q. \quad (4.3)$$

Par ailleurs, pour estimer préliminairement δ_Q , nous devons aussi calculer un très grand nombre de distances EMD entre la requête et la base suivant le modèle de fond de taille $n_{bg} \times m_B$. Nous avons vu que δ_Q correspond au quantile d'ordre ε/m_B de l'ensemble de ces distances (Section 4.3.1). On peut donc optimiser le calcul de l'estimation de δ_Q au fur et à mesure du calcul des distances entre la requête et la base de bruit en utilisant ce quantile comme seuil dans l'Equation (4.3) pour choisir de résoudre le problème linéaire jusqu'au bout.

Ces deux optimisations sont résumées dans le pseudo-code de l'Algorithme 2. Précisons qu'elles n'introduisent aucune approximation ou imprécisions sur le calcul des distances EMD entre la requête et les imagerie similaires. Elles permettent en pratique de réduire le temps d'une requête en moyenne de 40%. Sur la base IRHT, seules environ 10% des distances EMD doivent être calculées complètement. Le temps de calcul d'une requête est d'en moyenne 15 secondes pour $m_B = 1000$ images en utilisant des imagerie de taille $n = 150$ pixels et $n_{bg} = 3$, avec un Pentium IV 4.3GHz.

4.3.5 Exemples de requêtes par l'exemple

Des résultats de requêtes par l'exemple sont présentés sur les Figure 4.5 à 4.16. Sur l'ensemble des résultats nous voyons que le nombre d'images retournées, r_Q , varie de 0 (Figure 4.11) à plus de 100 (Figure 4.13 et 4.14). Certains résultats sur la base LCPD conduisent à des valeurs de r_Q égales à plusieurs centaines, mais pour des raisons d'espace et de lisibilité nous ne les montrons pas dans ce manuscrit.

La valeur $r_Q = 0$ induite par le seuil automatique indique que la base ne contient

Algorithme 2 Calcul des distances EMD pour une requête avec seuil automatique

Calcul du seuil δ_Q :

$\delta_t \leftarrow$ initialisation

pour toutes les imageries dans $\mathcal{M}$ **faire**

$k \leftarrow 1$

tant que $\sum_{i=1}^k \min_{1 \leq j \leq n} \{c_{ij}\} < \delta_t$ ET $k \leq n$ **alors** $k \leftarrow k + 1$

si $k == n + 1$ **alors**

 calculer la distance EMD jusqu'à la résolution du problème linéaire

 mettre à jour δ_t si le quantile a été modifié

fin si

fin pour

$\delta_Q \leftarrow \delta_t$

Calcul des distances EMD entre Q et les imageries de $\mathcal{B}$:

pour toutes les imageries dans $\mathcal{B}$ **faire**

$k \leftarrow 1$

tant que $\sum_{i=1}^k \min_{1 \leq j \leq n} \{c_{ij}\} < \delta_Q$ ET $k \leq n$ **alors** $k \leftarrow k + 1$

si $k == n + 1$ **alors**

 calculer la distance EMD jusqu'à la résolution du problème linéaire

fin si

fin pour

aucune image similaire à la requête selon l'OSC. Cette situation peut apparaître sur toutes les bases indépendamment de leurs tailles. Inversement, une grande valeur de r_Q nous indique la présence d'une grande série d'images similaires dans la base. L'exemple du bas de la Figure 4.13 peut être vu comme une erreur de la méthode. Cette longue série de dessins au crayon a été faite sur le même papier. L'empreinte picturale du crayon est légère et couvre peu la surface du dessin. Le contraste faible de ces dessins et la couleur constante du fond fait qu'elles sont toutes très proches les unes des autres selon l'EMD. Le seuil n'est pas assez sévère par rapport à ces subtilités, et pour lui toutes ces œuvres sont effectivement très similaires comparativement au modèle de fond. Ce type d'erreur apparaît aussi dans la base LCPD. De nombreuses photographies en noir et blanc présentent en moyenne un contraste peu marqué.

La base MIXCA contient aussi de nombreux objets décoratifs hérités de la collection d'arts décoratifs du Musée des Beaux arts de Montréal : statues, chaises, cuillères, pièces de monnaies, boîtes à encens... Ces séries d'objets dont nous n'avons pas mesuré l'ampleur au moment de constituer la base MIXCA sont rapidement ressorties dans les résultats (par ex. Figure 4.13 en haut). De même pour la base CLIC, nous avons ainsi vu *a posteriori* la quantité assez grande de paysages répondant à l'OSC horizontale assez classique d'un ciel bleu et d'un sol vert. En dehors de la base LCPD, de telles séries sont peu fréquentes en peintures par exemple, même si il arrive souvent de trouver des séries de 10 à 20 peintures similaires.

Ceci illustre bien la capacité et l'utilité du seuil à pouvoir sonder le contenu d'une base trop grande pour pouvoir être explorée manuellement. Même si des bases comme ATCI ou IRHT peuvent être explorées manuellement en étant patient, il est très difficile d'envisager une telle tâche pour une base de taille supérieure à quelques milliers.

Nous voyons sur l'ensemble des résultats que la frontière entre images similaires et images différentes indiquée par le seuil automatique n'est pas toujours parfaite. Au contraire, un utilisateur placerait souvent le seuil quelques rangs en avant ou en arrière selon les exemples. Ceci est souvent frappant lorsque nous visualisons des résultats à r_Q faible où nous avons rapidement une bonne vue d'ensemble des images retournées. Néanmoins nous voyons que le seuil donne généralement un bon ordre de grandeur et permet de savoir si il y a 0, 10, 50, 100 ... ou beaucoup plus d'images semblables dans la base. Cette information est d'un intérêt certain pour *interroger* une base et connaître son contenu.

4.4 Conclusion

Nous avons proposé dans ce chapitre une approche *a contrario* pour répondre au problème de mise en correspondance entre deux images selon l'OSC. Un important avantage de la méthode proposée est qu'elle fixe automatiquement un seuil pour chaque requête et chaque base consultée. Tout d'abord, le critère devient plus sévère avec la taille de la base, ce qui peut-être primordial lorsqu'on interroge des bases très volumineuses. De plus, la base de bruit $\mathcal{M}$ utilise les marginales de la base consultée $\mathcal{B}$. Par conséquent, le seuil s'adapte aussi à la fréquence d'apparition de ces couleurs dans la base. Si par exemple la base est constituée exclusivement d'enluminures rouges sur un fond blanc, le seuil sera très sensible au respect du rouge, si le rouge de la requête est assez particulier. Cet aspect peut être très utile dans le cadre de bases spécialisées. Le seul paramètre dont dépende cette méthode est le paramètre ε qui correspond au nombre de fausses alarmes lorsqu'on soumet une requête. Ce paramètre est plus intuitif à fixer qu'une borne fixe sur la mesure de similarité.

Enfin nous avons vu dans les résultats généraux que ce seuil automatique donne effectivement souvent une bonne idée de nombre d'image similaires à la requête dans la base. La délimitation précise qu'il crée entre images similaires et différentes est souvent discutable. Quelques images déclarées comme différentes peuvent sembler similaire selon l'OSC et selon l'utilisateur qui regarde les résultats. Inversement il arrive assez souvent que des images qui nous semblent différentes sont déclarées similaires. Ces dernières sont néanmoins généralement placées dans les derniers résultats. Bien que le rang exact au delà duquel les images sont déclarés comme différentes soit souvent imparfait, nous voyons que la méthode se trompe rarement sur l'ordre de grandeur du nombre d'images similaires à la requête contenu dans la base. Cet ordre de grandeur varie de 0 à plusieurs centaines selon les bases et les requêtes. Cette information est intéressante car elle nous renseigne sur le contenu de la base en fonction d'une requête.



FIG. 4.5: Base ATCI.

4.4. Conclusion

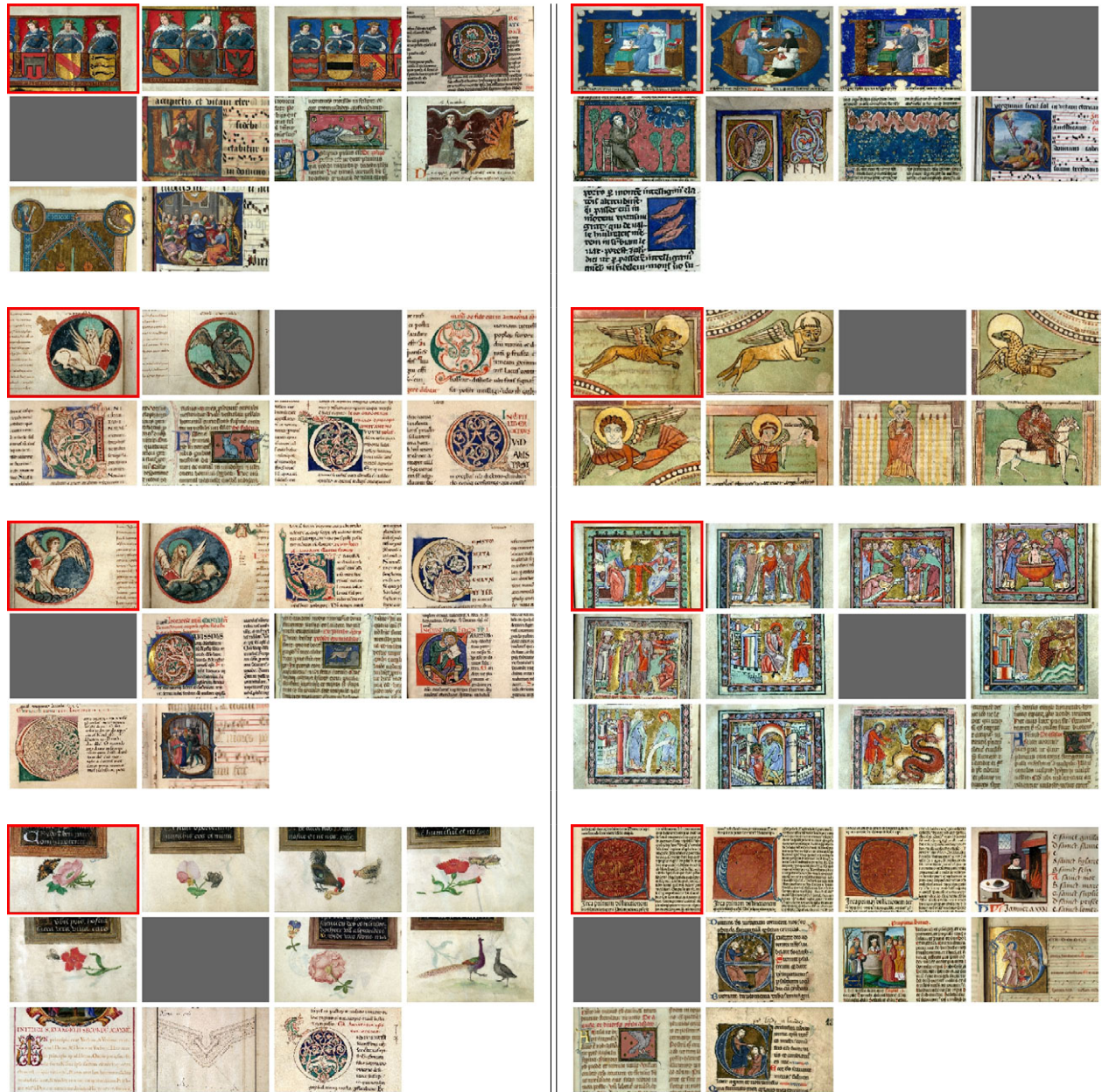


FIG. 4.6: Base IRHT.

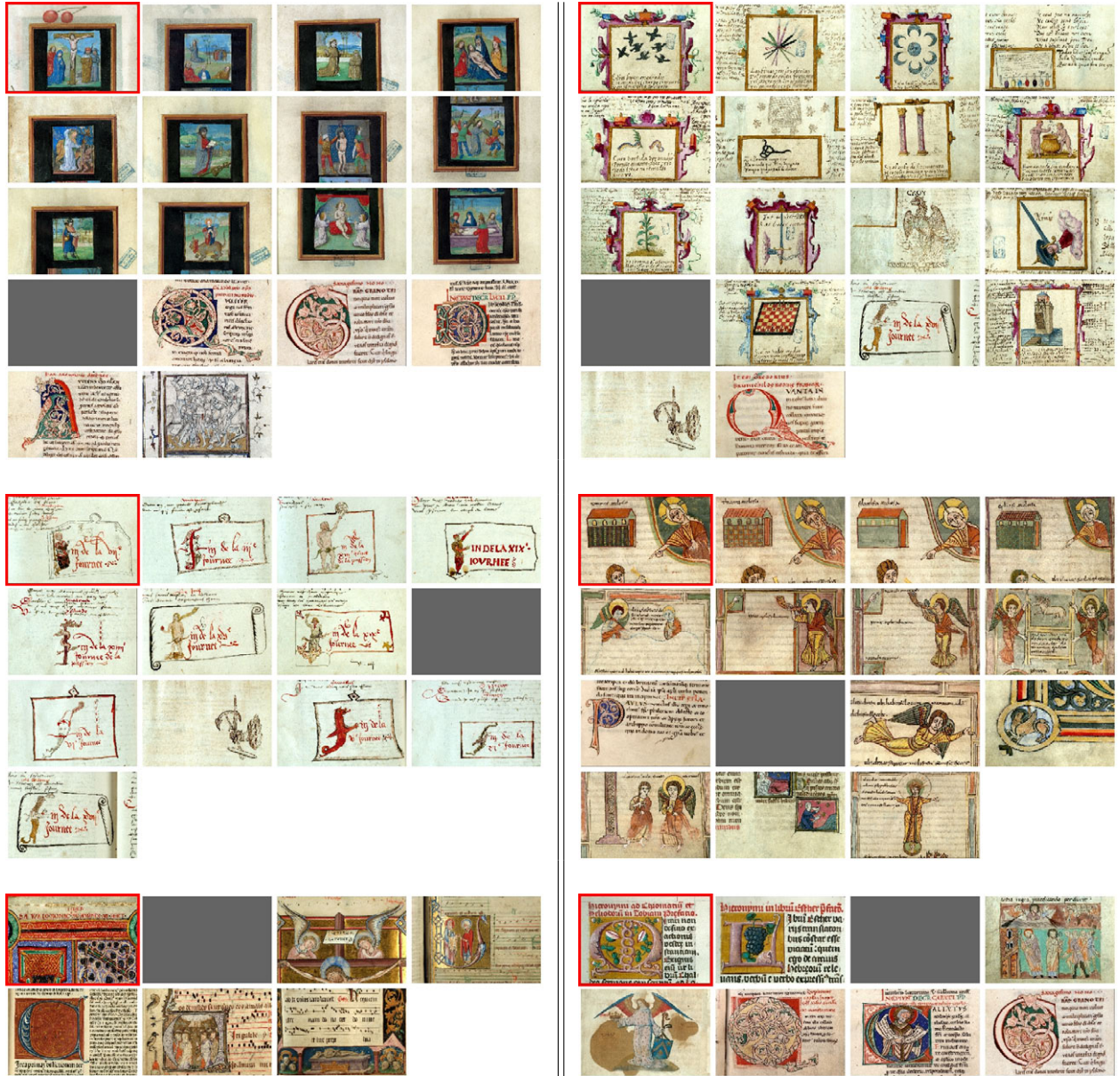


FIG. 4.7: Base IRHT.

4.4. Conclusion



FIG. 4.8: Base IRHT.

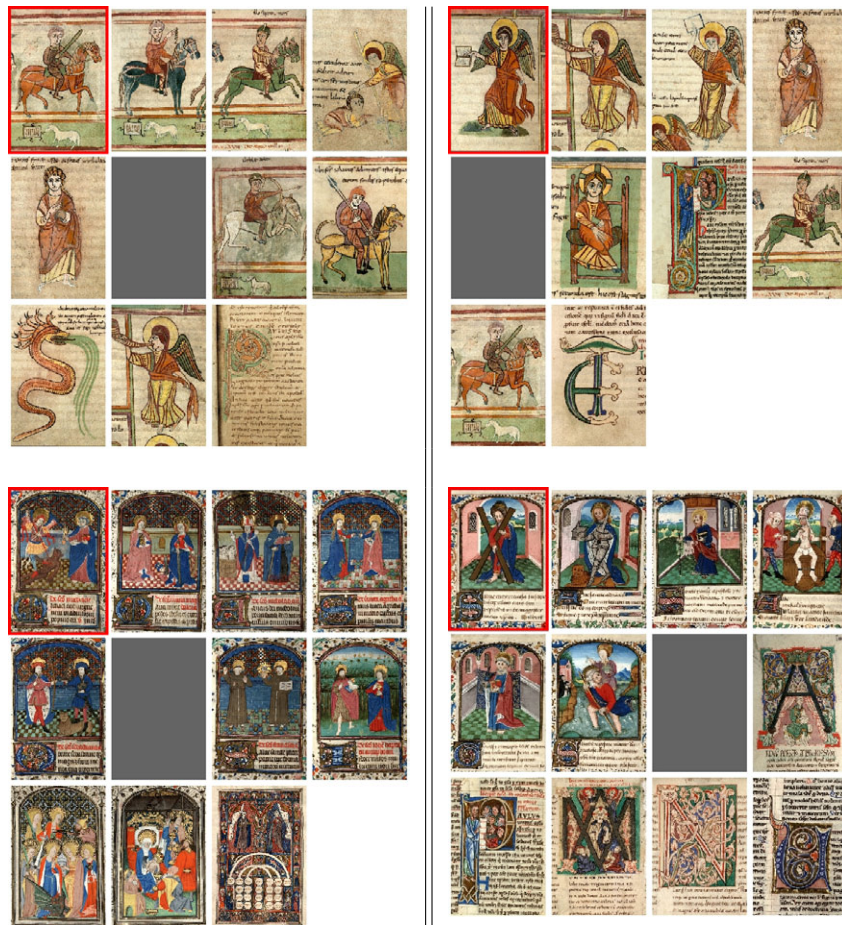


FIG. 4.9: Base IRHT.

4.4. Conclusion

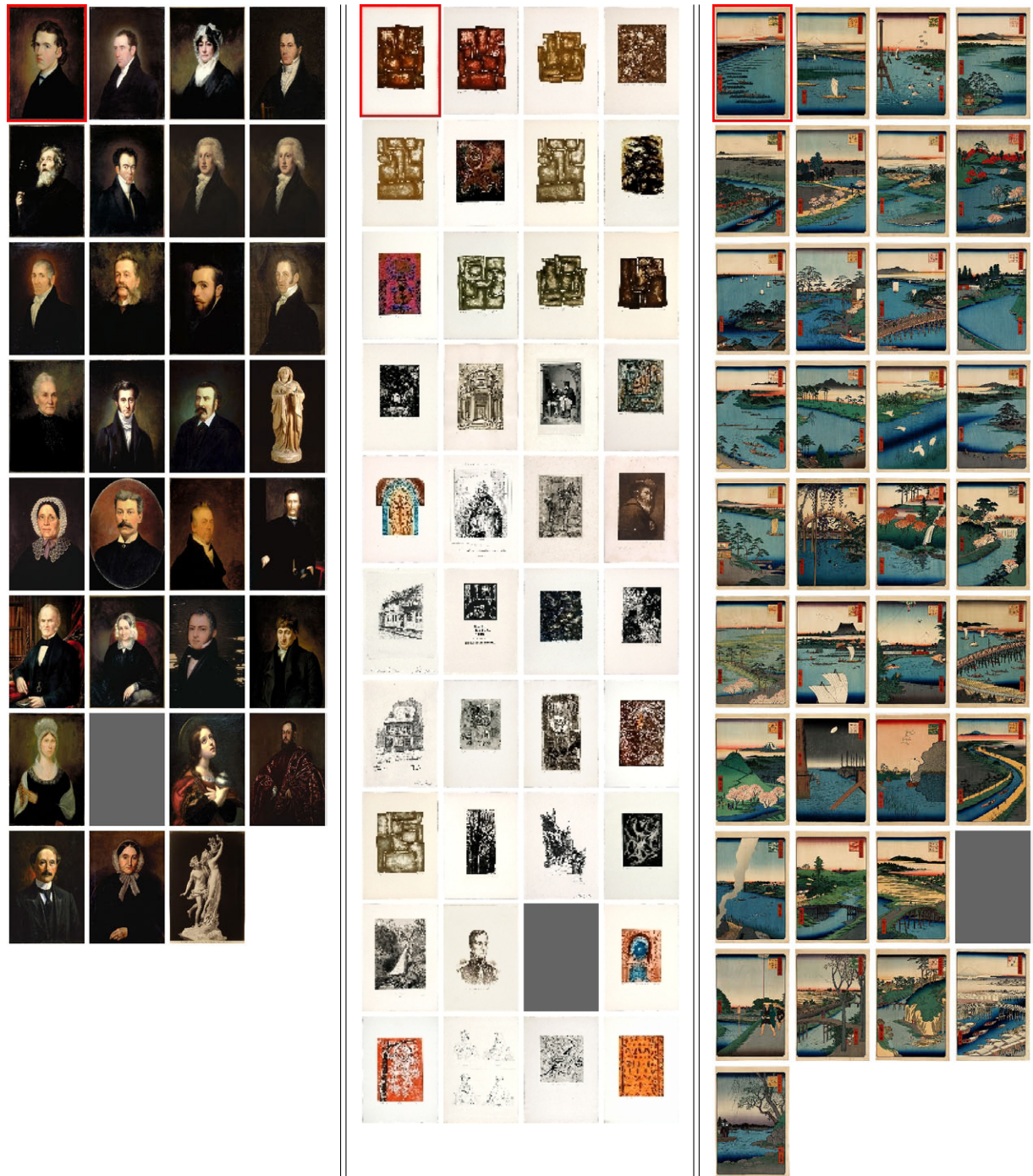


FIG. 4.10: Base MIXCA.

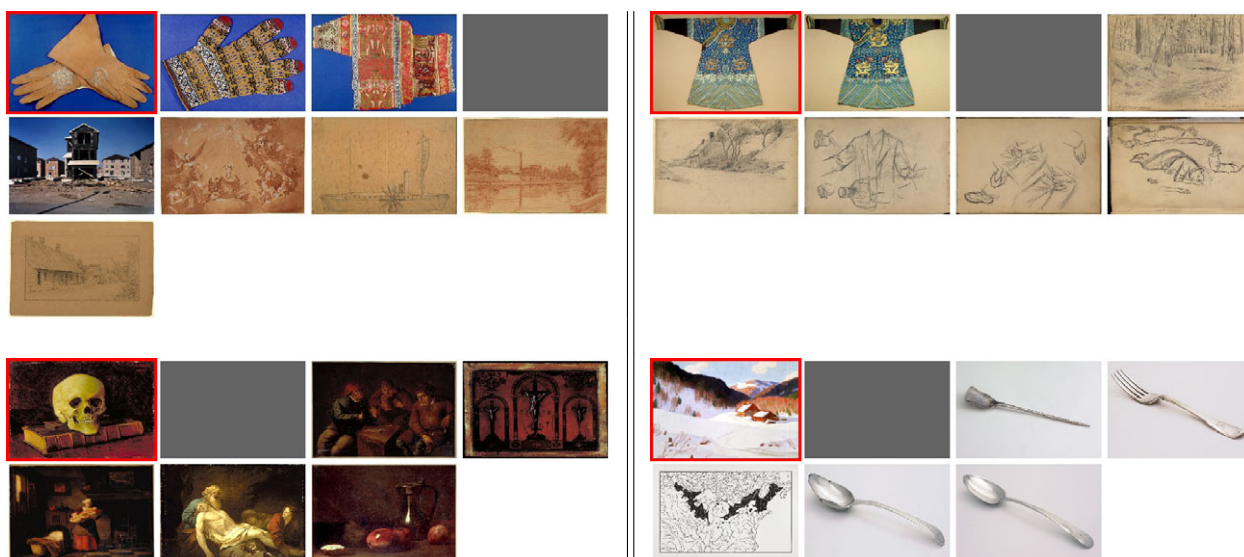


FIG. 4.11: Base MIXCA.

4.4. Conclusion

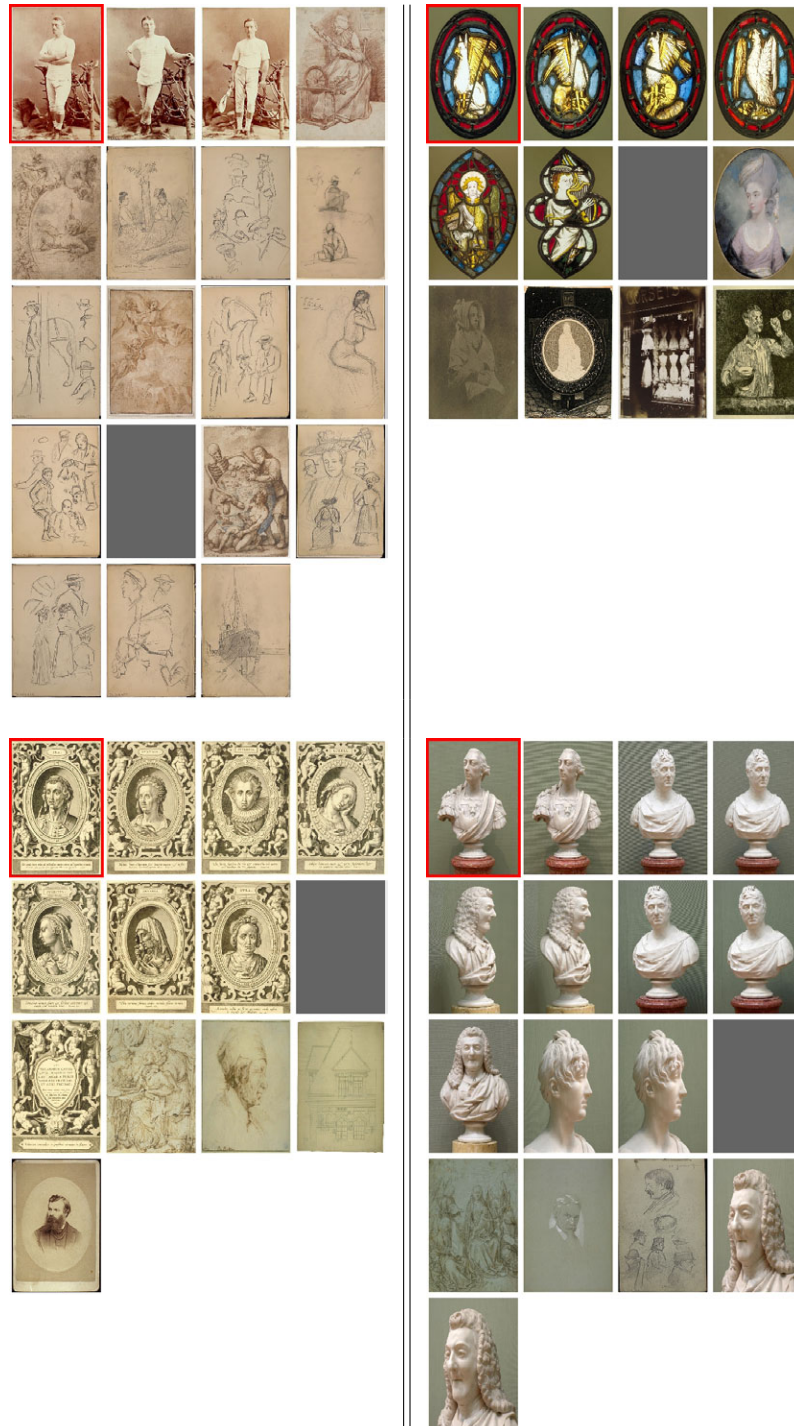


FIG. 4.12: Base MIXCA.



FIG. 4.13: Base MIXCA.

4.4. Conclusion



FIG. 4.14: Base LCPD.

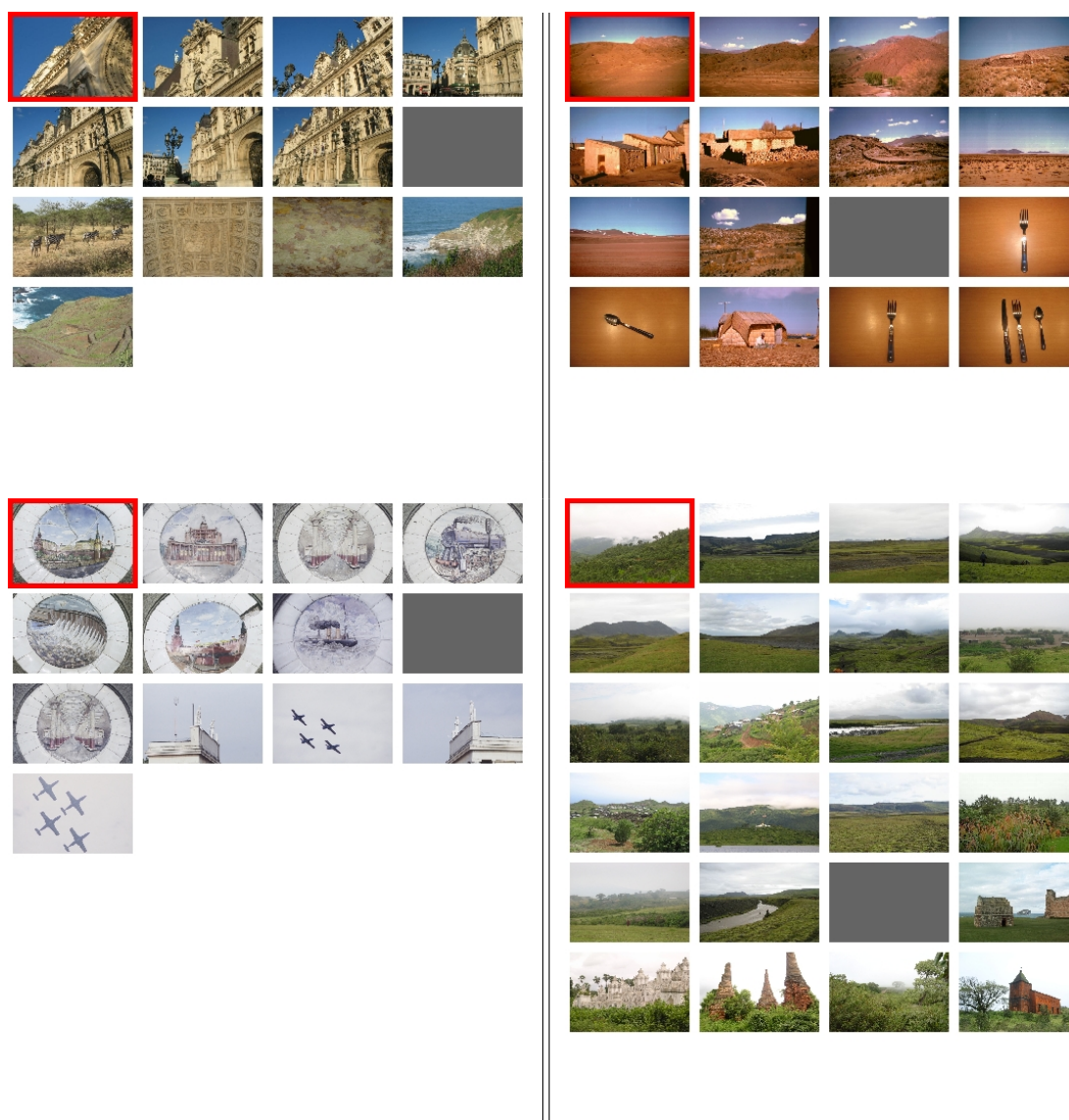


FIG. 4.15: Base CLIC.

4.4. Conclusion

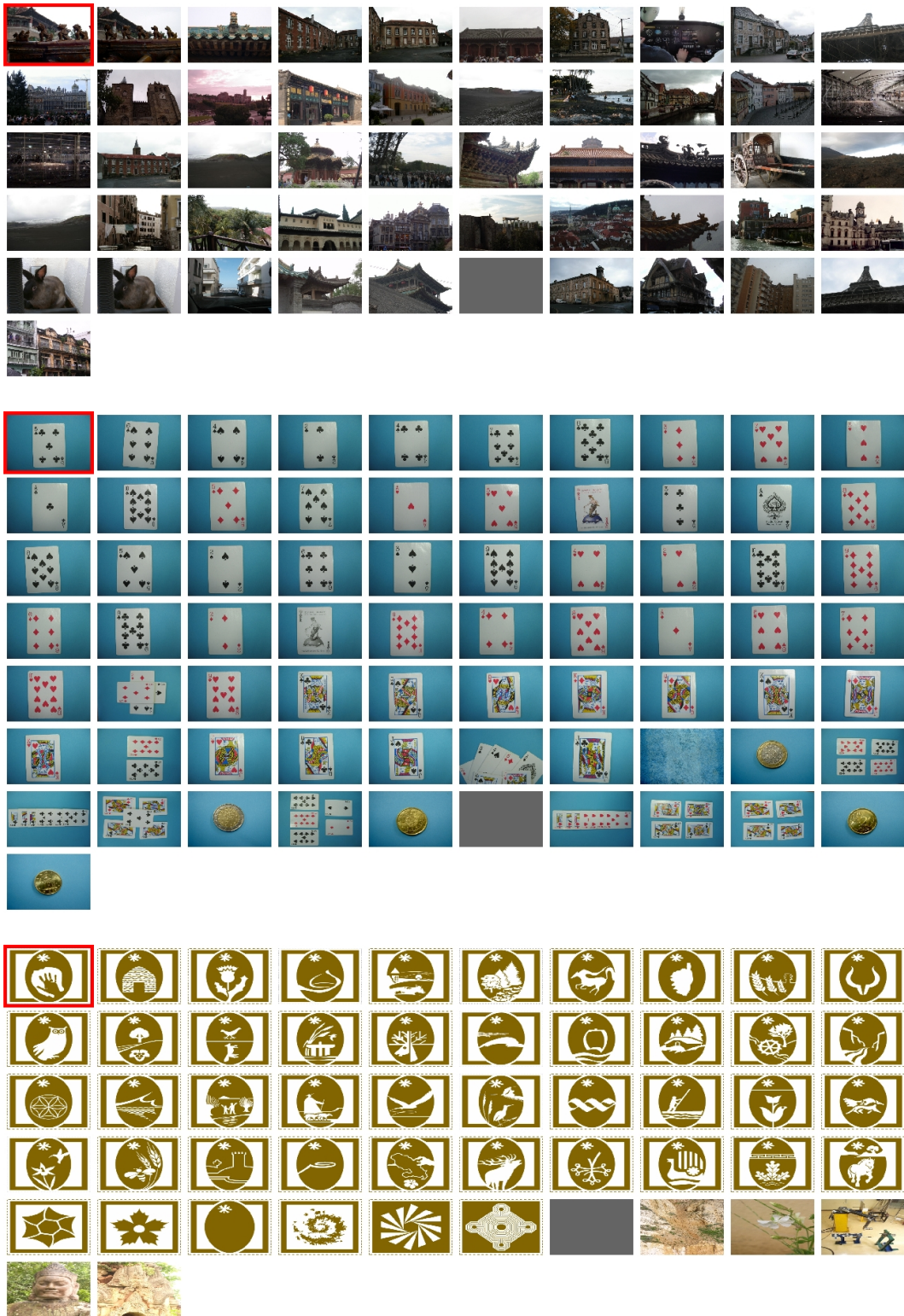


FIG. 4.16: Base CLIC.

Deuxième partie

Geométrie des lignes

Chapitre 5

Extraction des contours des traits



FIG. 5.1: Exemples d'œuvres uniquement à base de traits. De gauche à droite, des dessins de Matisse, Ellsworth Kelly, et Picasso (Annexe A).

Nous nous intéressons dans cette partie aux images qui contiennent uniquement des traits monodimensionnels (Figure 5.1). Les images d'écritures manuscrites font partie de cette catégorie. En ce qui concerne les œuvres d'arts, on retrouve ce type d'image en particulier dans les dessins réels et par ordinateur, croquis, estampes ... De ce fait, nous nous plaçons volontairement dans un cadre d'étude limité, mais cohérent avec la classification du contenu proposée par [Willats, 1997] et utilisée par [Grabli, 2005] en synthèse de dessins à bases de lignes. Notre but est d'étudier l'impact pictural des primitives 1D dans les œuvres (Chapitre 6).

Nous présentons dans ce chapitre la méthode que nous utiliserons pour extraire les contours des traits dans les œuvres. Après un état de l'art des approches possibles pour extraire les contours d'une image (Section 5.1) et de deux applications connexes, nous présenterons une approche *a contrario* proposée par [Desolneux et al., 2001] pour extraire

les contours significatifs (Section 5.2). Cette approche donne des résultats satisfaisants sur des images générales, mais elle ne prend pas en compte les spécificités des images de traits et les contraintes qui y sont liées. C'est pourquoi nous proposons à la Section 5.3 une amélioration. Plusieurs expériences permettront à la Section 5.4 d'illustrer les propriétés de la méthode proposée.

5.1 Travaux connexes sur les méthodes d'extraction des contours

Les méthodes existantes pour extraire les contours des objets dans une image peuvent s'organiser en trois groupes. Les méthodes locales sont les plus classiques (Section 5.1.1). Les contours actifs sont souvent rencontrés également (Section 5.1.2). Enfin la morphologie mathématique a aussi abordé cette problématique (Section 5.1.3). La reconnaissance d'écriture cursive (Section 5.1.4) et la vectorisation de dessins d'ingénierie (Section 5.1.5) sont deux problématiques assez similaires que nous présenterons brièvement.

5.1.1 Méthodes locales

La détection de contours est un problème qui a reçu une grande attention depuis les débuts de la recherche en analyse automatique d'images. Les approches dites classiques reposent sur des détecteurs fondés des masques [Hueckel, 1971], ou le passage par zéro du Laplacien [Marr and Hildreth, 1980, Haralick, 1984], ou le gradient des niveaux de gris [Torre and Poggio, 1986]. Des approches dites analytiques ont été proposées dont les plus connues sont celles de Canny [Canny, 1986] et Deriche [Deriche, 1987]. Elles possèdent les inconvénients de nécessiter l'ajustement de plusieurs paramètres et d'être relativement sensibles au bruit. Ces visions du contour sont purement locales et elles n'incorporent pas de principe de regroupement comme ceux de la Gestalt [Kanizsa, 1996]. Néanmoins ces types de détecteurs restent parmi les plus utilisés aujourd'hui, car ils sont simples et peu coûteux en calculs.

L'ensemble de ces détecteurs fournissent des contours ouverts nécessitant souvent une étape de fermeture avant une éventuelle étape d'analyse ou de reconnaissance de formes. Les principales méthodes pour fermer un contour décrit localement sont celles basées sur la recherche dans des graphes en fonction d'un critère de coût fixé [Martelli, 1972, Montanari, 1971]. Ces méthodes sont les plus efficaces mais aussi les plus coûteuses en temps de calcul. Un autre groupe de méthodes de fermeture concerne les automates. Beaucoup plus rapides et peu coûteuses à mettre en œuvre, mais donnant des résultats sous optimaux, ces méthodes proposent de déterminer en chaque point du contour le point suivant, en fonction de critères locaux [Giraudon, 1987].

5.1.2 Contours actifs

Un autre type d'approche très différent concerne les contours actifs ou *snakes* [Kass et al., 1988] [Blake and Zisserman, 1998]. Il s'agit de méthodes semi-automatiques dans lesquelles l'opérateur place dans l'image un contour initial au voisinage de la forme à détecter. Ce contour est amené à se déformer sous l'action de plusieurs forces externes (attaches aux données) et internes (régularisation). Une autre représentation variationnelle des contours est constituée par les méthodes par ensembles de niveaux, dont l'avantage est de pouvoir changer de topologie si les contours l'imposent [Osher and Sethian, 1988] [Sethian, 1996]. Ici le terme d'ensembles de niveaux se rapporte au fait que le contour est défini comme le niveau zéro d'une surface en dimension trois. Enfin les approches par contours actifs peuvent être aussi reliées à la famille des modèles géodésiques actifs [Caselles et al., 1997]. Ces trois dernières approches ont l'inconvénient d'être très sensibles au choix des paramètres et d'avoir une forte dépendance au contour initial.

5.1.3 Utilisation de la carte topographique et des lignes de niveaux

La carte topographique

La morphologie mathématique propose d'extraire les formes dans une image en utilisant les frontières des ensembles de niveau [Serra, 1982]. En notant u une image et $\lambda \geq 0$, nous pouvons définir une ligne de niveau $\mathcal{L}_\lambda$ comme étant une composante connexe des frontières topologiques de l'ensemble de niveau $\chi_\lambda(u) = \{x \in \mathbb{R}^2, u(x) \leq \lambda\}$. Selon la topologie de $\chi_\lambda(u)$, il peut exister plusieurs lignes de niveaux pour chaque niveau $\lambda \geq 0$. L'ensemble des lignes de niveaux dans une image est appelée carte topographique. Monasse a proposé une méthode d'extraction rapide de la carte topographique, et un regroupement hiérarchique de l'ensemble de ces lignes dans un arbre d'inclusion [Monasse, 2000]. La carte topographique présente les propriétés intéressantes suivantes :

1. Elle contient toute l'information de l'image. L'image peut être complètement reconstruite à partir de la donnée de ces lignes de niveaux.
2. Le calcul de la carte topographique peut être fait de manière efficace [Monasse and Guichard, 2000]. La transformée rapide d'extraction de la carte topographique est appelée **FLLT**¹.
3. Cette représentation est invariante par changement de contraste local.
4. Les lignes de niveaux sont des lignes fermées, hormis celles rencontrant le bord de l'image. Mais ces dernières peuvent être fermées en ajoutant une bordure d'un pixel autour de l'image de valeur -1 par exemple (si les niveaux de gris sont quantifiés sur $[0, 255]$).

¹Fast Level Lines Transform

5. Enfin, les contours des objets coïncident localement avec des lignes de niveaux [Serra, 1982, Salembier and Garrido, 2000, Lisani et al., 2003].

La réciproque de la dernière propriété est par contre fausse. La carte topographique contient aussi tous les détails des textures et du bruit. La carte topographique est également redondante : de nombreux morceaux de lignes codent le même bord d'objet. C'est pourquoi il est indispensable de filtrer la carte topographique pour extraire les lignes de niveaux représentatives du contenu dans l'image.

Filtrage d'arbre

La principale contribution que nous apporterons dans ce chapitre à la Section 5.3 et consiste en une méthode de filtrage d'une partie de la carte topographique en utilisant sa structure hiérarchique. Les méthodes existantes de filtrage de la carte topographique [Monasse, 2000] ou d'autres représentations morphologiques sous forme d'arbres [Salembier et al., 1998] utilisent des attributs associés aux noeuds de l'arbre. On peut distinguer les attributs de type morphologique qui ne dépendent pas des valeurs des pixels dans $\chi_\lambda(u)$ (par exemple : surface couverte par la région délimitée par la ligne, entropie, rapport entre surface et périmètre ...) des attributs non morphologiques (contraste minimal le long de la ligne, valeur de λ , ...).

Certains de ces attributs comme la surface décroissent avec la descendance dans l'arbre. Filtrer un arbre lorsque l'attribut est décroissant est assez simple car nous pouvons supprimer toute la descendance dès que l'attribut d'un noeud ne vérifie pas une condition sur un seuil. Pour les attributs qui ne sont pas décroissants, plusieurs stratégies existent. Ces méthodes peuvent se regrouper en deux familles.

- Les méthodes de suppression noeud par noeud :
 - *Directe* : tous les noeuds qui ne respectent pas le seuil sont supprimés. Dans le cadre d'un filtrage des ensembles de niveau (par ex. pour segmenter l'image), la valeur du niveau de gris des pixels liés au noeud supprimé est ajusté à la valeur du premier parent non supprimé, par ex. [Monasse, 2000, Luo et al., 2007].
 - *Par soustraction* : identique à la méthode directe mais les pixels correspondant aux noeuds enfants ont leur valeur de niveau gris diminuée de la différence de niveaux de gris entre le noeud supprimé et le noeud parent [Urbach, 2002].
- Les méthodes d'élagage ou *pruning* en anglais, qui enlèvent directement une partie de la descendance [Salembier et al., 1998, Monasse, 2000] :
 - *Min* : le noeud est enlevé si il ne vérifie pas la condition liée au seuil ou si un de ses parents est aussi enlevé, par ex. [Caselles and Monasse, 2002].
 - *Max* : le noeud est enlevé si il ne vérifie pas la condition liée au seuil et si tous ses enfants ont été enlevés, par ex. [Caselles and Monasse, 2002].
 - *Viterbi* : cette méthode consiste à associer un coût à chaque possibilité de suppression ou de conservation d'un noeud, puis à trouver le chemin de coût minimal

dans le graphe des décisions possibles pour l'ensemble de l'arbre.

Le choix de l'une de ces stratégies dépend de l'attribut utilisé et du niveau de complexité toléré. [Desolneux et al., 2001] ont proposé une méthode que l'on peut qualifier de type directe adaptative pour sélectionner les lignes de niveaux les plus significatives dans la carte topographique. Cette méthode a été comparée à l'utilisation de modèles déformables [Cao et al., 2005]. L'attribut utilisé par Desolneux et al. est un attribut statistique calculé à l'aide d'une méthode de type *a contrario*. La méthode proposée dans ce chapitre utilisera aussi cet attribut. Nous en rappellerons les détails à la Section 5.2.1. [Desolneux et al., 2001] complètent aussi leur stratégie de filtrage par un deuxième filtrage reposant sur un *principe de maximalité*. Nous verrons que ce deuxième filtrage n'est pas suffisamment sélectif pour être appliqué à notre problématique. La contribution principale de ce chapitre est de proposer un nouveau principe de maximalité adapté aux dessins au trait et présenté à la Section 5.3.1.

5.1.4 Reconnaissance automatique d'écriture cursive

L'écriture cursive, quelque soit l'alphabet utilisé, présente de nombreux points communs avec les dessins au trait [Georges, 1989]. Les outils sont très diversifiées et proches de ceux des dessins. Les gestes et effets sont très similaires aussi. Un dessin réalisé au trait tout comme une phrase écrite de manière cursive présente des croisements, jonctions, points de rebroussement, levés et posés de crayon etc. (Figure 5.2).



FIG. 5.2: Les 26 éléments de l'alphabet étrusque. Ecriture et dessins au trait ont de nombreux points de similitudes.

En analyse automatique de documents écrits plusieurs problématiques variées ont émergées auxquelles différentes méthodologies ont été appliquées. Sans vouloir couvrir le domaine entièrement, nous nous restreindrons au cas de l'analyse d'écriture hors ligne [Plamondon and Srihari, 2000, Mullot, 2006], c'est à dire sans avoir l'information temporelle du tracé. Ceci est en opposition avec l'écriture en ligne où le scripteur utilise généralement un système de stylet et de tablette graphique.

Les recherches en ce domaine ont commencé il y a environ quarante ans [Earnest, 1962, Frishkopf and Harmon, 1961, Lindgren, 1965]. Le problème de l'écriture cursive reste cependant encore irrésolu. La grande variabilité des styles et modèles d'écriture compliquent la tâche des méthodes automatiques. Ces différents modèles peuvent être mélangés par le scripteur dans un même mot, et chaque scripteur a par ailleurs aussi son propre style lié au geste physique. Cette double variabilité délibérée [Mullot, 2006] s'ajoute à des variations de qualité et de lisibilité qui peuvent être plus accidentelles.

La reconnaissance automatique de l'écriture poursuit un but sémantique très clair : décoder le message écrit indépendamment du style d'écriture propre au scripteur. Il est alors naturel de vouloir extraire le tracé du manuscrit. Les liens entre contenu et contenant (style et sujet) évoqués à la Section 2.1.5 sont visibles ici aussi.

La squelettisation est une approche classique pour aborder la reconnaissance d'écriture car elle permet de s'approcher du geste graphique initiale et se libérer de la morphologie et du style générale. Il existe plus de 300 méthodes pour extraire le squelette d'une image [Lee et al., 1991] car la notion de squelette ne conduit pas à une définition mathématique unique. Deux grandes familles de méthodes se distinguent : les méthodes d'érosions itératives, et les méthodes de calcul d'axe médian (voir [Lam et al., 1992] pour un article de synthèse). L'ensemble de ces méthodes utilisent une image binaire. Binariser une image est une tâche assez difficile qui conduit toujours à une dégradation du tracé [Pavlidis, 1993].

Garain et al. ont proposé une méthode de binarisation adaptative dans le cadre d'images fortement dégradées et particulièrement adaptée aux documents sur papier [Garain et al., 2006]. Après un prétraitement visant à éliminer toutes les marges, une étape d'étiquetage des composantes connexes est réalisée sur l'image dans l'espace couleur HSV quantifié sur des voisinages 9×9 . Le pas de quantification est estimé automatiquement sur chaque image. Pour chaque composante connexe, la taille, valeur moyenne dans HSV, le barycentre et la plus petite boîte englobante sont calculés. Ces informations permettent de classer l'arrière plan et l'avant plan en fonction de critères moyens itératifs et hiérarchiques. Les performances de cette binarisation sont évaluées sur plusieurs bases de documents écrits anciens et comparés avec la binarisation DjVu [Bottou et al., 1998].

Malheureusement même dans le cas idéal où l'image est facilement binarisable sans perte, il a été montré qu'à résolution égale une image binaire contient environ 10 fois moins d'informations utiles qu'une image en niveau de gris pour un rapport d'espace mémoire égal à 3.5 environ [Pettier and Camillerapp, 1993]. C'est pourquoi Pettier et Camillerapp proposèrent une méthode de représentation à partir de squelettes généralisées [Pettier et al., 1993] utilisant des images en niveaux de gris. Ils utilisent une approche mixte mêlant segmentation de régions et suivi de contours. Le volet segmentation utilise une détection par laplacien et gradient. Par ailleurs leur méthode est hiérarchique car l'objectif d'un squelette global n'est pas poursuivi. Les zones non squelettisables sont traitées à part et viennent enrichir une représentation hétérogène du contenu en vue d'une reconnaissance des caractères. Cette mixité des approches permet selon les auteurs d'améliorer la robustesse de leur méthode. [Goto and Aso, 1999] ont proposé une méthode d'extraction des lignes de texte en utilisant un critère de linéarité pour être plus robuste aux variations de l'épaisseur du tracé. Nous justifierons expérimentalement le fait que nous n'ayons pas choisi une méthode de squelettisation dans les expériences à la Section 5.4.1.

5.1.5 Reconstruction et analyse de dessins d'ingénierie

Une problématique assez proche de l'extraction des traits dans une œuvre est la reconstruction de dessins techniques en mécanique, électronique, architecture, etc. Dans un souci d'économie d'espace mémoire et de facilité de manipulation, des méthodes ont été proposées pour vectoriser de tels dessins à partir d'images scannées. Pour une article de synthèse de ce domaine, ses enjeux et perspectives, voir [Tombre, 1997]. Ces méthodes peuvent être regroupées en deux classes.

Premièrement de nombreuses méthodes utilisent la détection des points de cassures. Les points de cassures, ou points dominants ou points à fortes courbures, sont tout d'abord extraits dans le dessin [Mokhtarian and Mackworth, 1992, Fayolle et al., 2000, Wu, 2003, Gu and Tjahjadi, 2000]. Nous reverrons plus en détail ces méthodes à la section 6.1.3 dans un état de l'art sur l'utilisation de la courbure et des points dominants. Les traits de dessin entre ces points sont ensuite estimés par minimisation d'une mesure de distance à l'image binaire.

Deuxièmement des méthodes de reconstruction le long des traits utilisent des modèles rectilignes, circulaires, elliptiques ou curvilignes. Hormis les méthodes classiques de squelettisations topologiques qui ont été proposés en dehors de ce domaine [Gonzales and Woods, 2002], Han et Fan ont développé en 1994 une squelettisation qui utilise un détecteur de contours par masques et une mise en correspondance de ces contours à l'aide de critère de proximité et parallélisme [Han and Fan, 1994]. Cette méthode se base sur l'hypothèse de dessins relativement clairs, en majorité composés de lignes droites et arc de cercles, et où il y a peu de zones d'ambiguïtés tels que des recouvrements de traits ou des régions à forte densité de jonctions.

Chen et al. ont proposé en 2003 une approche mixte qui utilise à la fois une segmentation par masque des lignes en éléments simples entre deux cassures, et l'utilisation d'informations au niveau des points de cassures au sein d'un algorithme génétique [Chen et al., 2003]. Ici encore, la méthode fonctionne tant que les zones d'ambiguïtés sont simples à résoudre.

Les dessins réels présentent en général de nombreuses zones d'ambiguïtés. Par ailleurs nous verrons dans le Chapitre 6 que notre problématique ne nécessite pas une reconstruction ou vectorisation totale des traits de dessins.

5.2 Détection a contrario des contours dans une image

Dans cette section, nous rappelons brièvement la méthode *a contrario* proposée par [Desolneux et al., 2001] pour sélectionner les lignes de niveaux les plus significatives de la carte topographique. L'ensemble des lignes significatives contient de nombreuses redon-

dances. [Desolneux et al., 2001] ont donc proposé un principe de filtrage pour en éliminer une partie, que nous discuterons à la Section 5.2.2.

5.2.1 Lignes de niveaux significatives

Schéma probabiliste

Pour sélectionner les lignes de niveaux les plus significatives [Desolneux et al., 2001] ont proposé une approche *a contrario* (voir Section 4.1). L'événement défini dans le cadre de la sélection de lignes de niveaux est la présence d'une ligne de niveau suffisamment longue et suffisamment contrastée. Si nous notons u une image, nous cherchons les lignes de niveaux le long desquelles le contraste atteint une valeur minimale qui reste significativement élevée par rapport à un contexte aléatoire. Les valeurs du contraste le long des lignes de niveaux sont comparées à la distribution des contrastes non nuls dans l'image étudiée. Nous supposons que le contraste de l'image est distribué comme la variable aléatoire X :

$$\forall \mu > 0, P(X > \mu) = \frac{\#\{x, |\nabla u(x)| > \mu\}}{\#\{x, |\nabla u(x)| > 0\}}, \quad (5.1)$$

où $\#$ dénote le cardinal de l'ensemble, et $|\nabla u(x)|$ est la norme du gradient calculée par différence finie sur un voisinage de taille 2×2 . Nous notons $H(\mu)$ cette probabilité estimée empiriquement. Remarquons que $H(\cdot)$ est monotone décroissante. Notons N_u le nombre de lignes de niveaux dans u extraites avec un pas de quantification égal à 1 sur les niveaux de gris. [Desolneux et al., 2001] définissent une ligne de niveau $\mathcal{L}$ comme étant ε -significative si :

$$NFA(\mathcal{L}) = N_u \times \left[H\left(\min_{x \in \mathcal{L}} |\nabla u(x)|\right) \right]^l < \varepsilon, \quad (5.2)$$

où l est la longueur de $\mathcal{L}$. Ce nombre est appelé nombre de fausses alarmes (NFA). Nous voyons que le NFA d'une ligne de niveau ne dépend ici que de sa longueur et de son contraste minimal. Une valeur faible du NFA indique une significativité élevée de la frontière délimitée par $\mathcal{L}$. Nous voyons avec cette définition, que la valeur du NFA décroît (et donc la significativité croît) avec la longueur l et la valeur du contraste minimal. Cette définition a été analysée et une variante locale a été proposée par [Cao et al., 2005]. Les auteurs ont également comparé cette approche aux contours actifs.

Comme dans le chapitre 4, cette méthode de décision ne dépend que du paramètre ε . [Desolneux et al., 2001] montrent que la dépendance à ce paramètre est très robuste. Le nombre de lignes significatives varient logarithmiquement avec ε . Nous choisirons toujours $\varepsilon = 1$. Dans toute la suite, les lignes de niveaux 1-significative seront simplement appelées lignes de niveaux significatives.

Signe du contraste

Les lignes de niveaux qui ne rencontrent pas la bordure de l'image sont naturellement fermées sur elle mêmes. Pour fermer celles qui rencontrent les bords de l'image, une bordure de 1 pixel de large et de valeur artificielle -1 est construite autour de l'image. La première ligne de niveau qui englobe toutes les autres est la frontière du support de l'image. Nous pouvons donc considérer toutes les lignes de niveaux comme étant fermées. Lorsque le vecteur gradient est orienté de l'extérieur de la ligne de niveau vers l'intérieur, nous dirons que la ligne est positive. Nous dirons qu'une ligne de niveau est négative quand le gradient est orienté de l'intérieur vers l'extérieur (Figure 5.3).

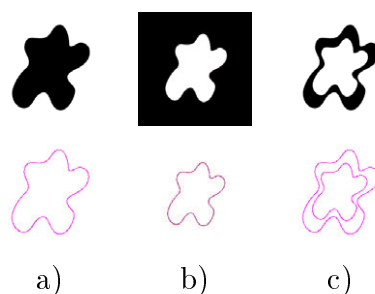


FIG. 5.3: Signe des lignes de niveaux. Trois images synthétiques sont présentées en haut. Les lignes de niveaux permettant de décrire leurs contours sont visibles en bas. a) lorsque le vecteur gradient est orienté de l'extérieur de la ligne de niveau vers l'intérieur, nous dirons que la ligne est positive. b) ici le contraste est négatif. c) deux lignes de niveaux de signes opposés délimitent un trait courbe fermée.

Arbre hiérarchique des lignes significatives

La structure hiérarchique de la carte topographique est conservée lorsque les lignes de niveaux significatives sont sélectionnées. Deux lignes de niveaux sont soit disjointes soit liées par inclusion. Nous pouvons ainsi représenter l'ensemble des lignes significatives sous la forme d'un arbre où chaque noeud contient une ligne de niveau et son type. Un exemple de dessin est présenté sur la Figure 5.4a). Un ensemble des ligne de niveaux manuellement sélectionnées pour représenter au mieux les contours des traits est présenté sur la Figure 5.4b). L'arbre correspondant est présenté sur la Figure 5.4c). Une ligne de niveau positive (resp. négative) est symbolisée par un noeud blanc (resp. noire). L'inclusion est représentée de bas en haut. Une ligne enfant dans l'arbre est incluse dans son parent au niveau supérieur.

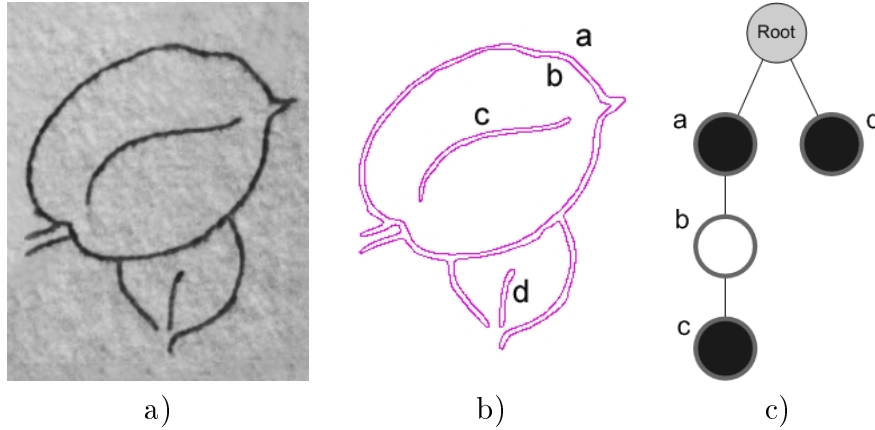


FIG. 5.4: Un exemple simple. a) un dessin de deux feuilles de plante scanné sur 256 niveaux de gris. b) ensemble de lignes de niveaux manuellement choisies pour décrire au mieux les contours des traits c) l'arbre hiérarchique associé à b). Une ligne de niveau positive (resp. négative) est symbolisée par un noeud blanc (resp. noire). La ligne positive décrit ici le contour intérieur de la plus grosse des deux feuilles. La racine *root* est la ligne qui passe par la bordure de l'image, obtenue en ajoutant une marge d'un pixel de valeur -1 .

5.2.2 Lignes redondantes et principe de maximalité

Bien que l'ensemble des lignes significatives contienne beaucoup moins de lignes que la carte topographique complète, il contient généralement de nombreuses lignes redondantes dès qu'il s'agit d'images réelles. Les images réelles présentent généralement des contours légèrement flous et ayant une certaine épaisseur. C'est pourquoi de nombreuses lignes de niveaux siègent le long de ces contours (Figure 5.5).

Pour éliminer les lignes redondantes, Desolneux et al. utilisent les branches de l'arbre dites *maximales monotones*, qui avait été proposées par [Monasse, 2000]. Une branche maximale monotone est une branche de n noeuds consécutifs $\{\mathcal{N}_i\}_{i=1\dots n}$ de même type (**monotone**), où les noeuds $\mathcal{N}_i, i \geq 2$ n'ont qu'un seul parent, et qui ne puisse pas être incluse strictement dans une autre branche monotone monoparentale (**maximale**). Sur la Figure 5.6 est présenté le début de l'arbre des lignes significatives associées à la Figure 5.5a. Chaque groupe vertical de noeuds symbolise une branche maximale monotone notée B_1 à B_{26} .

Une ligne significative est dite maximale si elle a un NFA minimal sur la branche maximale monotone à laquelle elle appartient. Pour chaque branche maximale monotone de l'arbre des lignes significatives, seule la ligne significative maximale est conservée. Tous les autres noeuds de la branche maximale monotone sont éliminés. Cette méthode de sélection est illustrée sur la Figure 5.6. Sur cette figure, chaque série de noeuds représente une branche maximale monotone, dont le groupement est symbolisé dans le dessin associé



FIG. 5.5: Lignes de niveaux significatives. Cette figure présente l'ensemble des lignes de niveaux significatives du dessin de la Figure 5.4. L'arbre hiérarchique associé contient 717 lignes de niveaux. La carte topographique totale extraite de l'image codée sur 8 bits avec un pas de quantification de 1 contient plus de 26000 lignes de niveaux. A droite, une région du dessin encadrée en rouge est détaillée.

par une ligne noire épaisse.

Le principe de maximalité proposé par Desolneux et al. n'élimine pas toutes les redondances dans l'arbre. Par exemple, la ligne significative maximale choisie par ce principe de maximalité dans la branche B_1 sur la Figure 5.6a) va induire une redondance visible sur la Figure 5.7c). La ligne $\mathcal{L}_3$ résultant de cette sélection est visible sur la Figure 5.7c). Les lignes maximales $\mathcal{L}_1$ et $\mathcal{L}_2$ visibles sur la Figure 5.7c) sont celles sélectionnées dans les branches B_2 et B_3 visibles sur la Figure 5.6b).

Remarque : ce principe de maximalité peut être vu comme une méthode de type directe (voir Section 5.1.3) mais adaptative, où le seuil est recalculé à chaque branche monotone maximale et est égal au NFA minimal.

5.3 Extraction des contours de traits dans les dessins

Dans toute la suite, un trait monodimensionnel fait par l'artiste sera simplement appelé un **trait**. Dans ce chapitre, notre but est de proposer une méthode qui permet d'extraire le contour d'un trait en utilisant les lignes de niveaux. Un groupe de traits qui se croisent sera appelé un **agrégat**.

Nous présentons dans cette section la méthode utilisée pour extraire les contours des traits et agrégats dans les œuvres à bases de traits. Sur le dessin de la Figure 5.4a) par exemple, ces contours correspondent à ceux de la Figure 5.4b). Nous utilisons le schéma

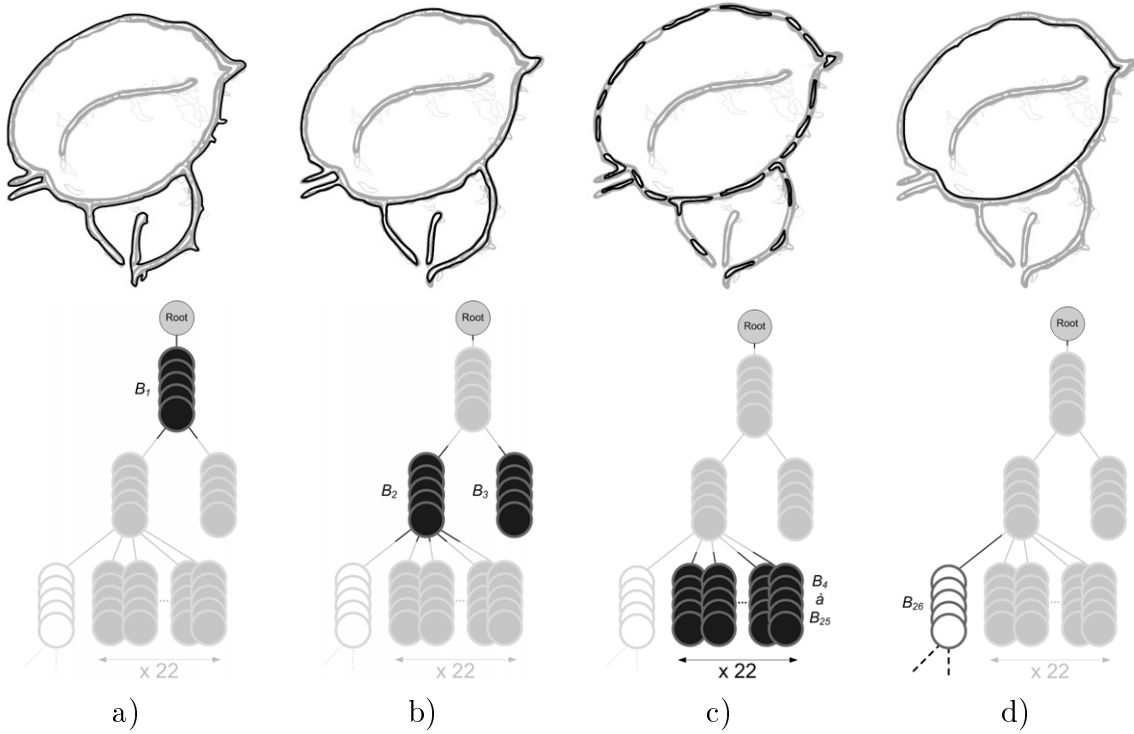


FIG. 5.6: Structure hiérarchique de l'ensemble des lignes significatives. Chaque série de noeuds représente une branche maximale monotone, notée B_i pour $i \in [1, 26]$ sur cette figure, et dont le groupement est symbolisé dans le dessin associé par une ligne noire épaisse. a) la branche maximale B_1 monotone englobe tous les traits du dessin. b) le dernier noeud de B_1 englobe deux traits qui sont physiquement disjoints. Le plus petit trait isolé est décrit par la branche B_3 au second niveau de l'arbre. c) la branche B_2 englobe les 22 branches négatives B_4 à B_{25} dues aux variations de contraste le long des traits, et la branche positive B_{26} décrivant l'intérieur de la grande feuille de la plante. d) la branche maximale monotone B_{26} composée de lignes de niveaux positives décrit le contour intérieur de la grande feuille. Pour plus de clarté, le reste de l'arbre (l'intérieur de la grande feuille) n'est pas détaillé sur cette figure.

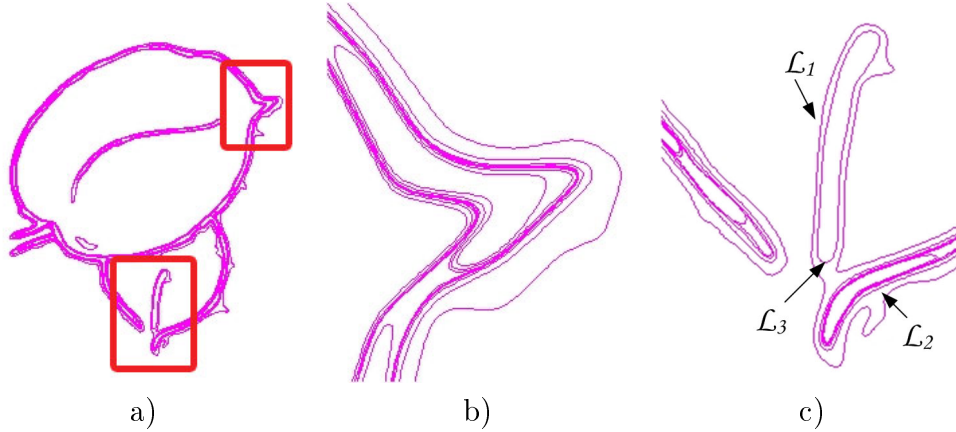


FIG. 5.7: Redondances. a) Exemple d'utilisation du principe de maximalité de Desolneux et al. L'ensemble résultant contient 54 lignes de niveaux. b) et c) Détails de deux régions du dessin encadrées en rouge. De nombreuses redondances persistent. A droite, les redondances liées aux trois premières branches B_1 , B_2 et B_3 de l'arbre visible sur la Figure 5.6.

probabiliste de Desolneux et al. introduit à la Section 5.2.1 pour extraire l'arbre des lignes significatives en éliminant directement celles dont le NFA est supérieur à $\varepsilon = 1$. Puis nous définissons un nouveau principe de maximalité appliqué aux arbres monotones maximaux que nous proposerons à la Section 5.3.1. Cette méthode de filtrage de l'arbre des lignes significatives est spécifiquement adaptée à l'extraction des contours dans une image de traits. Enfin, nous introduirons à la Section 5.3.2 la méthode utilisée pour modéliser l'ensemble des contours d'un agrégat.

Remarque : Nous considérons ici uniquement des œuvres où les traits sont plus foncés que le fond. Si en pratique nous désirons aussi considérer des dessins réalisés avec des traits blancs sur fond noir par exemple, ce type de dessins pourrait être assez facilement détectés en analysant la distribution des niveaux de gris. Nous pourrions alors inverser les niveaux de gris pour retrouver le cadre d'étude décrit ici.

5.3.1 Principe de maximalité non supervisé appliqué aux arbres monotones maximaux

Dans toute la suite, nous appellerons arbre monotone maximal (AMM), tout sous-arbre de l'arbre des lignes significatives qui contient uniquement des noeuds de même type, et qui ne peut pas être inclu strictement dans un autre AMM. Sur la Figure 5.6 par exemple, le premier AMM négatif correspondrait à l'union des 25 branches maximales négatives B_1 à B_{25} . Donc à la différence des branches maximales monotones utilisées par [Desolneux

et al., 2001], dans un arbre monotone maximal, nous autorisons la possibilité pour un noeud d'avoir plusieurs enfants.

Remarques préliminaires

Nous avons vu dans les sections précédentes que l'arbre des lignes significatives contient de nombreuses lignes de niveaux redondantes décrivant chaque contours de trait présent dans une image. L'analyse des dessins au trait que nous souhaitons réaliser dans le chapitre suivant utilise les contours des traits et leur géométrie locale. Il est donc indispensable d'extraire une représentation unique de chaque contour. Pour atteindre ce but, nous proposons un nouvel algorithme de sélection qui s'appuie sur la structure d'inclusion particulière des lignes correspondant aux traits d'un dessin.

Il est peu probable qu'un changement de contraste positif perceptible apparaisse à l'intérieur d'un trait d'artiste. En effet, cela serait le reflet d'un artefact blanc significatif à l'intérieur du trait sombre. Grâce au profil physique classique des outils artistique et le moyennage longitudinal introduit lors du geste, ce type d'artefact est improbable. Le signe d'une ligne de niveau dans l'arbre des lignes significatives change si par exemple, une ligne de niveau enfant décrit l'intérieur d'une boucle (voir Figure 5.4). Par conséquent, on peut raisonnablement considérer que toutes les redondances associées à un contour soient groupées dans un seul et même AMM. Cependant, plusieurs contours différents et leurs redondances peuvent être groupés dans un seul AMM. Cette situation survient avec le premier AMM négatif de la Figure 5.6. Le contour extérieur de la grande feuille et le trait de la nervure de la petite feuille sont incluses dans les branches B_2 et B_3 du premier AMM.

Pour étudier et éliminer les redondances d'un contour de trait, nous pouvons donc nous restreindre à l'étude de chaque AMM pris séparément. La présence de plusieurs traits possibles dans un AMM nous impose néanmoins d'envisager la possibilité de devoir sélectionner plusieurs lignes de niveaux dans un AMM, autant qu'il y a de traits contenus dans la région située entre la racine du AMM et ses feuilles.

Pour une ligne de niveau $\mathcal{L}$ délimitant un contour de trait de manière satisfaisante, nous pouvons éliminer tous ses enfants appartenant au même AMM, car ces lignes appartiennent au même trait. De même, les noeuds parents de $\mathcal{L}$ appartenant au même AMM que $\mathcal{L}$, soit décrivent le contour du même trait, soit englobe un groupe de plusieurs traits couverts par le AMM. Ils peuvent donc être éliminés.

Nouveau principe de maximalité

En suivant les remarques faites dans les deux paragraphes précédents, nous proposons un nouveau principe de maximalité. Pour chaque AMM de l'arbre des lignes de niveaux significatives, nous cherchons la ligne de niveau $\mathcal{L}$ la plus significative (i.e. celle dont le

NFA est le plus faible). Puis nous supprimons de l'arbre des lignes de niveaux significatives toutes les lignes qui sont parents ou enfants de $\mathcal{L}$ dans le AMM. Ce principe est appliqué à l'AMM tant que celui n'est pas complètement vidé. Ce nouveau principe de maximalité est non supervisé, et permet de ne conserver que les lignes de niveaux les plus significatives tout en s'assurant que chaque contour ne sera approximé que par une seule et unique ligne de niveau.

L'algorithme 3 est résumé ci-dessous. Appliqué à l'arbre présenté sur la Figure 5.5, il permet d'obtenir exactement l'arbre présenté à la Figure 5.4c). Son application est détaillée également dans la Figure 5.8 pour le premier AMM de l'arbre. Les résultats finaux de cet algorithme et du principe de maximalité proposé par Desolneux et al. sur l'arbre complet sont présentés sur la Figure 5.9. Cet algorithme est une variante d'une version que nous avons proposé dans le cadre de la segmentation d'IRM médicales [Hurtut and Cheriet, 2007]. Dans le cadre de l'étude des dessins, nous utilisons la contrainte *a priori* de l'unicité des contours des traits. Par exemple nous excluons la possibilité de présence d'un trait noir à l'intérieur d'un trait gris plus large.

Algorithme 3 Nouveau principe de maximalité proposé

Entree: Arbre des lignes significatives

Sortie: Arbre filtré par le nouveau principe de maximalité

Fonction principale

pour tous les sous arbres monotones maximaux S **faire**

filtrer(S)

fin pour

Fonction **filtrer**(S)

tant que S n'est pas vide **faire**

- $\mathcal{L} \leftarrow$ la ligne la plus significative de S
- enlever toutes les lignes de S qui soient parents ou enfants de $\mathcal{L}$

fin tant que

Remarque : pour chaque AMM, ce nouveau principe de maximalité peut être vu comme une méthode de type directe (voir Section 5.1.3) adaptative, où le seuil est recalculé à chaque itération de l'algorithme et est égal au NFA minimal du AMM.

5.3.2 Ensemble d'adhérence

Dans cette section nous cherchons à regrouper les lignes de niveaux significatives maximales (issues de l'Algorithme 3) en ensembles décrivant les contours des agrégats. Rappelons qu'un agrégat est un groupe de traits connectés par croisements entre eux. Ces ensembles seront appelés *ensembles d'adhérence*. Le dessin sur la Figure 5.4 sera par exemple décrit par trois ensembles d'adhérences. Deux de ces ensembles permettent de

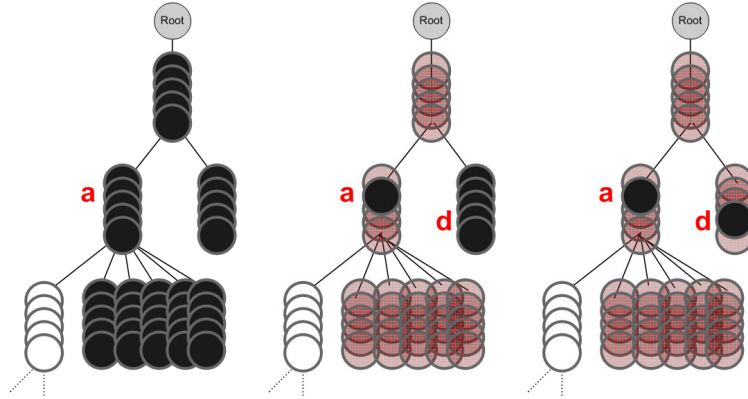


FIG. 5.8: Nouveau principe de maximalité appliqué au premier AMM de la Figure 5.6 A gauche, la ligne de niveau la plus significative est trouvée. L'étiquette *a* correspond à l'étiquette de la Figure 5.4. Au milieu, les lignes de niveaux qui sont parents ou enfants de cette ligne de niveau sont supprimées du AMM. A droite, la ligne la plus significative suivante est à nouveau trouvée (étiquette *d*). A la fin de cette étape, le AMM est vide, et l'algorithme passe au AMM suivant.

décrire les traits isolés *c* et *d* à l'intérieur et à l'extérieur de la grande feuille. Ceux ci ne contiennent chacun qu'une seule ligne de niveau (décrivant le contour extérieur du trait). Le troisième ensemble décrit le contour extérieur *a* (ligne négative) et intérieur *b* (ligne positive) de la grande feuille.

Un trait isolé qui ne se croise pas lui-même sera décrit par un ensemble d'adhérence ne contenant qu'une seule ligne négative. Un trait qui se croise une fois, sera décrit par un ensemble d'adhérence contenant une ligne négative et une ligne positive. De façon plus générale, un agrégat de traits sera décrit par une ligne négative (le contour extérieur de l'agrégat) et un nombre $n \geq 0$ de lignes positives. Ces n lignes de niveaux sont nécessairement positives et enfants directs de la ligne négative dans l'arbre des lignes significatives maximales. Comme la structure hiérarchique a été préservée jusqu'ici, construire les ensembles d'adhérence à partir de l'arbre des lignes significatives maximales est immédiat. Il y a en effet autant d'ensembles d'adhérences (et idéalement d'agrégats sous-jacents) que de lignes négatives. Chaque agrégat est constitué d'une ligne négative et de ses éventuels enfants directs de type positif. La Figure 5.10 résumé le schéma général d'extraction des contours des traits dans les dessins au trait.

5.4 Expériences

Nous comparons tout d'abord la méthode proposée à une méthode de squelettisation (Section 5.4.1). Puis nous comparons l'utilisation du principe de maximalité de Desolneux et al. avec celui que nous proposons dans ce chapitre (Section 5.4.2). Ensuite nous utilisons

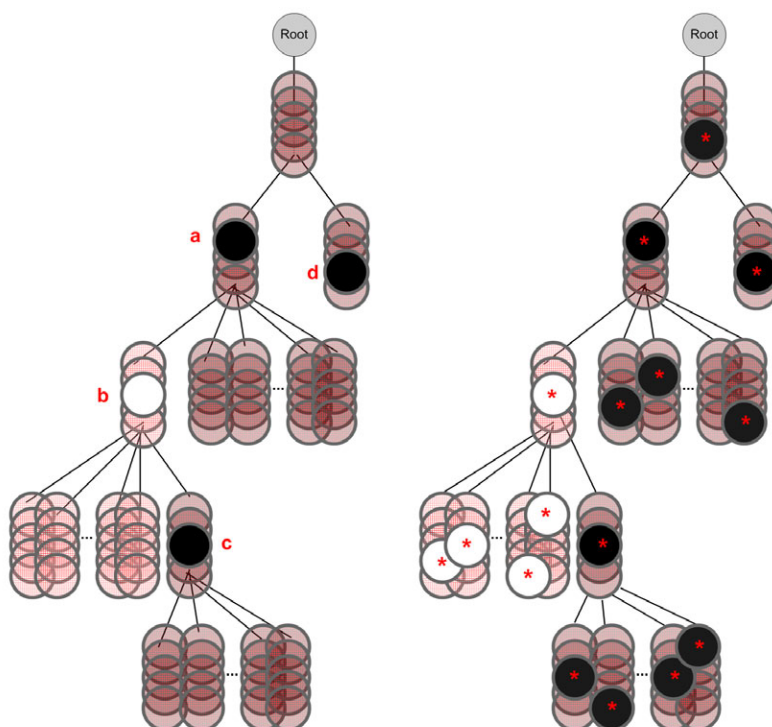


FIG. 5.9: Résultats finaux sur l'arbre complet. Sur cette figure est illustré l'application des deux différents principes de maximalité introduits dans ce chapitre. A gauche, le principe de maximalité proposé à la section 5.3.1 ne retient qu'un nombre très restreint de lignes, en suivant l'algorithme 3. A droite, le principe de maximalité de Desolneux et al. retient une ligne par branche maximale monotone (symbolisé par une astérisque rouge). Les étiquettes *a*, *b*, *c* et *d* correspondent aux étiquettes de la Figure 5.4c.

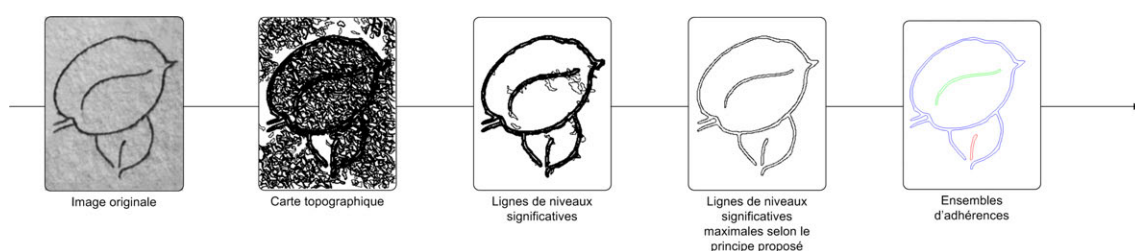


FIG. 5.10: Schéma global de la méthode proposée d'extraction des contours de traits. Tout d'abord, l'image est représentée par sa carte topographique. Le modèle probabiliste rappelé à la Section 5.2.1 est ensuite appliqué pour sélectionner les lignes significatives. Le principe de maximalité que nous avons proposé à la Section 5.3.1 élimine ensuite les lignes redondantes. Chaque agrégat de traits est enfin décrit par un ensemble d'adhérence constitué des lignes de niveaux significatives maximales décrivant ses contours. Dans ce chapitre, nos contributions à ce schéma global de détection sont les deux dernières étapes.

une image de synthèse (Section 5.4.3) pour illustrer le comportement des lignes de niveaux le long des traits. Nous présentons enfin à la Section 5.4.4 plusieurs exemples sur des images réelles issues des bases présentées dans l'Annexe A. Ne disposant pas de bases de référence pour l'extraction des traits dans une œuvre, notre évaluation se fondera sur des exemples.

5.4.1 Comparaison avec une méthode de squelettisation

Nous comparons dans cette section la méthode proposée avec une méthode de squelettisation appliquée à une image binaire. Nous avons binarisé les images en optimisant manuellement le seuil pour qu'il n'y ait aucun trait discontinu. La méthode utilisée est celle qui consiste à amincir séquentiellement la forme binaire par une série d'éléments structurants obtenus par rotation [Serra, 1982]. Cette méthode classique amincit la forme binaire jusqu'à idempotence.

Deux exemples sont visibles sur les Figure 5.11 et 5.12. L'étape de binarisation constitue une première difficulté des méthodes de squelettisation. L'ajustement du seuil peut être délicat, surtout dans le cas de dessins au tracé léger. La binarisation introduit toujours des irrégularités locales le long des traits qui sont présentes ensuite sur le squelette. Le masque d'érosion utilisé introduit lui aussi une certaine imprécision sur le trait extrait. Dans le chapitre suivant, nous utiliserons abondamment l'information de courbure le long des traits pour l'analyse et l'indexation des dessins. Nous voyons que les méthodes de squelettisation créent une certaine dégradation vis à vis de cette information.

Les croisements de traits ou rapprochements locaux et parallèles de traits introduisent aussi des erreurs importantes sur le squelette (Figure 5.11 en bas à droite) qui ne sont pas facilement supprimables *a posteriori*. Dans le chapitre suivant, nous utiliserons également beaucoup la présence de croisements. Les extrémités de traits sont des sources du même type d'erreurs.

5.4.2 Comparaison avec le principe de maximalité de Desolneux et al.

Dans cette section nous comparons l'utilisation du principe de maximalité de [Desolneux et al., 2001] présenté à la Section 5.2.2 avec celui que nous avons proposé à la Section 5.3.1. Sur l'ensemble des dessins des bases Matisse-Kelly et Picasso, le nombre final de lignes de niveaux est réduit de 88% en utilisant le principe de maximalité que nous proposons par rapport à l'utilisation de celui de Desolneux et al. Un résultat d'extraction illustrant l'application de ces différents principes sur un détail de dessin réel est visible sur la Figure 5.13. Rappelons que la méthode proposée par [Desolneux et al., 2001] n'est pas spécifiquement adaptée aux dessins au trait au contraire de celle proposée dans ce cha-

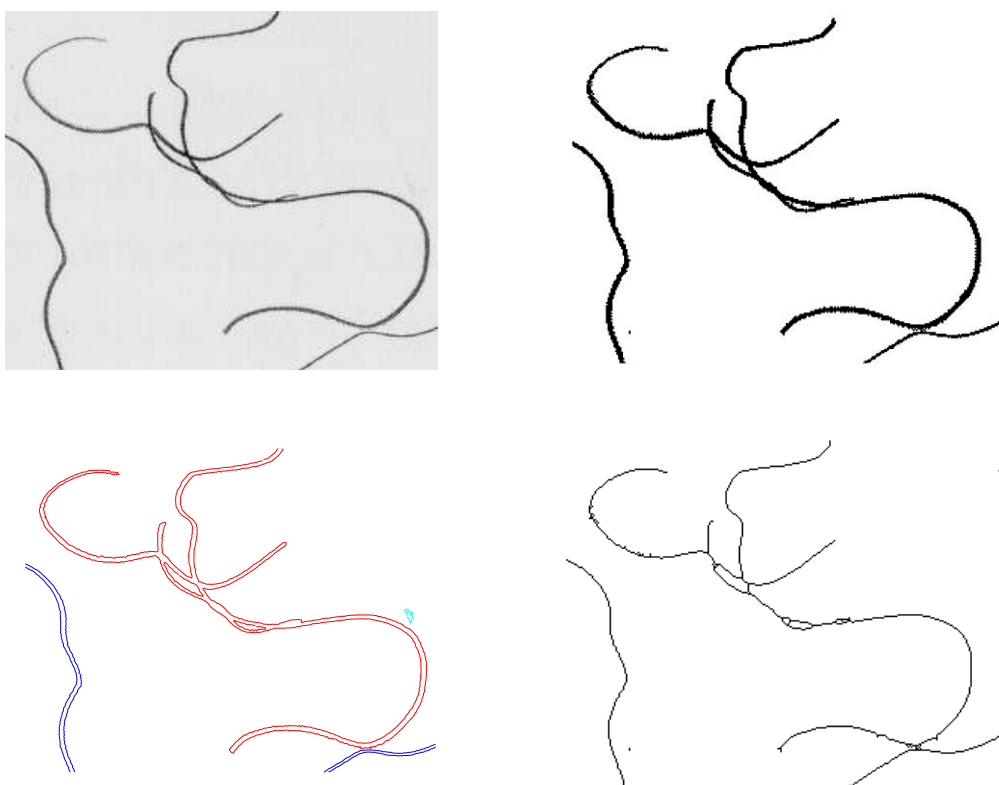


FIG. 5.11: Squelettisation. En haut l'image originale en niveau de gris, et l'image binaire. En bas à gauche, la segmentation des contours obtenue avec la méthode présentée à la Section 5.3.1 et à droite le squelette.

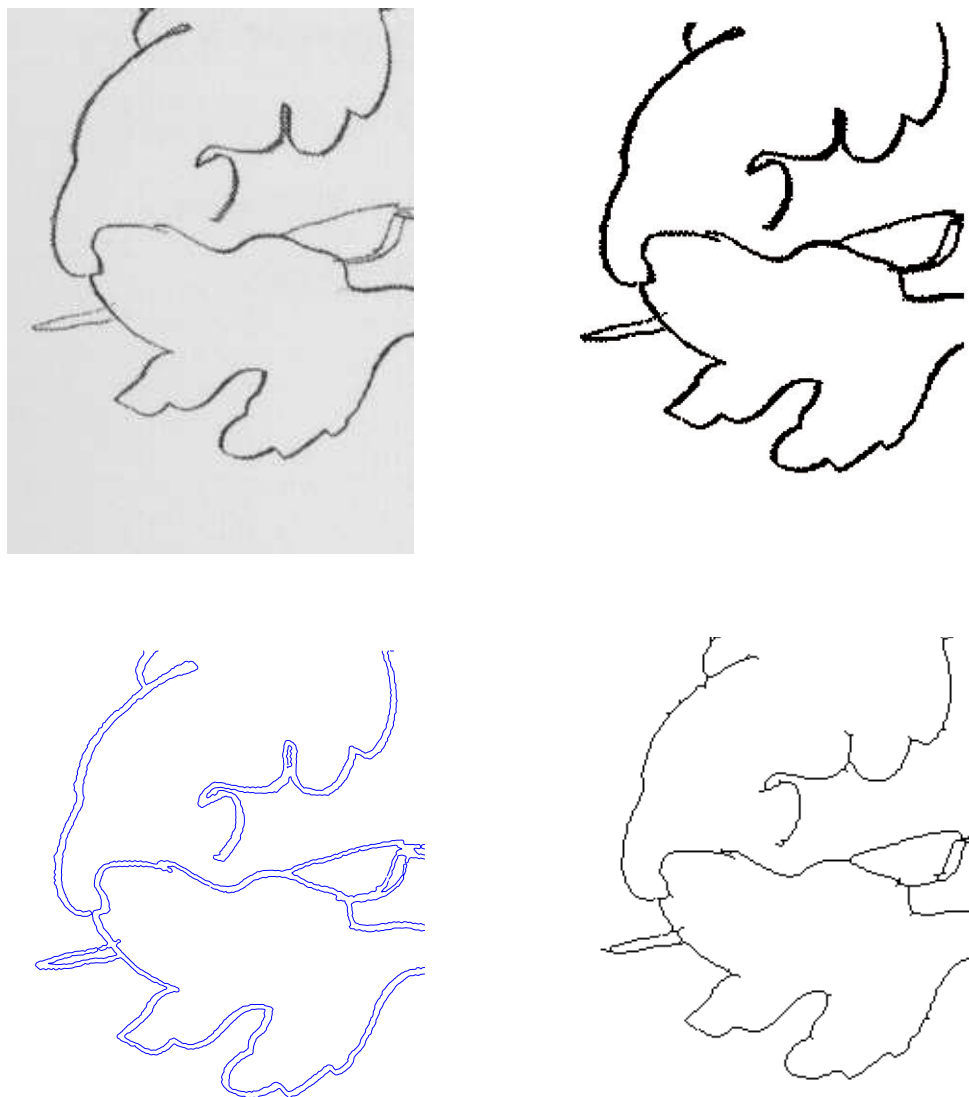


FIG. 5.12: Squelettisation. Même organisation que sur la Figure 5.11.

pitre. Cette comparaison illustre donc l'intérêt de la prise en compte de caractéristiques topologiques propres à de notre problématique.



FIG. 5.13: Principes de maximalité. A gauche, l'utilisation du principe de maximalité de [Desolneux et al., 2001] préserve un grand nombre de redondances de l'arbre des lignes de niveaux significatives. A droite, le principe proposé à la Section 5.3.1 sélectionnent les lignes les plus significatives.

5.4.3 Variations d'intensités le long des traits

Lorsque le niveau de gris varie le long d'un trait d'artiste, comment se comportent la carte topographique et la méthode proposée dans ce chapitre? Cette situation arrive fréquemment puisque nous ne binarisons pas les images et que les artistes appuient rarement avec une pression constante sur leur crayon. Sur la Figure 5.14a), plusieurs traits synthétiques d'intensité variables ont été tracés. L'image est en niveaux de gris sur 8 bits normalisée sur $[0, 255]$. Un bruit Gaussien ($\sigma = 5$) a été ajouté à l'image. Le premier trait en haut de l'image a une valeur constante égale à 0. L'intensité du dernier trait en bas de l'image croît linéairement jusqu'à la valeur 255 en son milieu puis décroît à nouveau jusqu'à 0. L'intensité des traits intermédiaires ne croît que jusqu'aux valeurs 245, 235, 225, 215 (de bas en haut). La carte topographique avec un pas de quantification égal à 15 est montrée sur la Figure 5.14b).

Sur cette carte, plusieurs lignes de niveaux entourent chaque trait. Elles contiennent par inclusion deux branches de lignes de niveaux séparées qui suivent les variations d'intensité. Nous voyons également qu'en dehors du dernier trait où l'intensité atteint celle du fond, il existe toujours au moins une ligne de niveau représentant le contour global du trait. Sur la Figure 5.14c), notre méthode réussit à extraire cette ligne de niveau. Rappelons que cette méthode ne repose sur aucun paramètre (nous avons fixé $\varepsilon = 1$ pour toutes les expériences de cette thèse). Sur la Figure 5.14d), la même image a été analysée par un filtre de Canny-Deriche, en choisissant manuellement les meilleurs seuils. La sensibilité au bruit est visible.

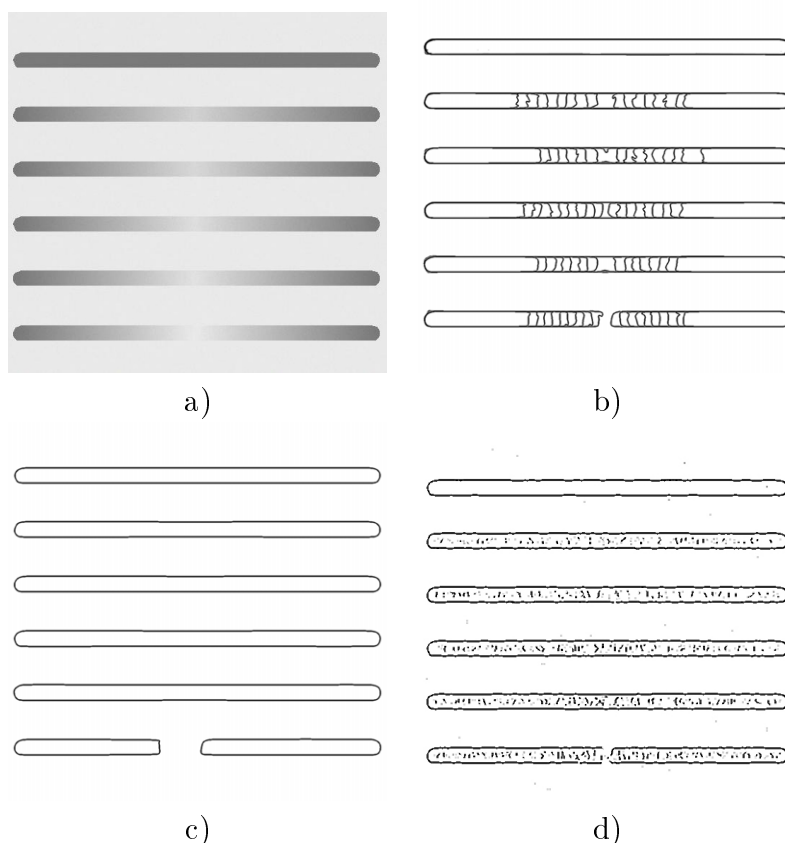


FIG. 5.14: Variation d'intensité le long d'un trait. a) plusieurs traits synthétiques ont été tracés sur un fond blanc puis bruités par un bruit Gaussien ($\sigma = 5$). Le premier trait en haut de l'image a une valeur constante égale à 0. L'intensité du dernier trait en bas de l'image croît linéairement jusqu'à la valeur 255 en son milieu puis décroît à nouveau jusqu'à 0. L'intensité des traits intermédiaires ne croît que jusqu'aux valeurs 245, 235, 225, 215 (de bas en haut). b) la carte topographique avec un pas de quantification égal à 15. Hormis le dernier trait qui atteint une intensité semblable au fond (et qu'un observateur humain identifierait comme un seul trait grâce au principe de la bonne continuation), il existe toujours une ligne de niveau englobant totalement le trait. c) notre méthode réussit à extraire la ligne la plus significative. Rappelons que cette méthode ne repose sur aucun choix de paramètre (nous avons fixé $\varepsilon = 1$ pour toutes les expériences de cette thèse). d) résultat de l'analyse par un filtre de Canny-Deriche. Les deux paramètres de la méthode ont été choisis manuellement pour obtenir le meilleur résultat.

Sur les images réelles, il arrive toutefois que les variations de contraste créent des erreurs de détection de contours. Néanmoins dans l'ensemble des résultats présentés à la section suivante, nous avons observé que la méthode est relativement robuste. Un exemple typique d'erreur est présenté sur la Figure 5.15. Ce type d'erreur survient dans des zones d'ambiguïté très fortes, où même un observateur humain ne peut pas toujours dire si l'artiste a soulevé son outil, ou fortement diminué sa pression sur le support. Dans le Chapitre 6, nous verrons que ce type d'erreur n'a d'impact que sur un seul attribut parmi les 11 proposés à la Section 6.3 pour l'indexation ou l'analyse des dessins au trait.

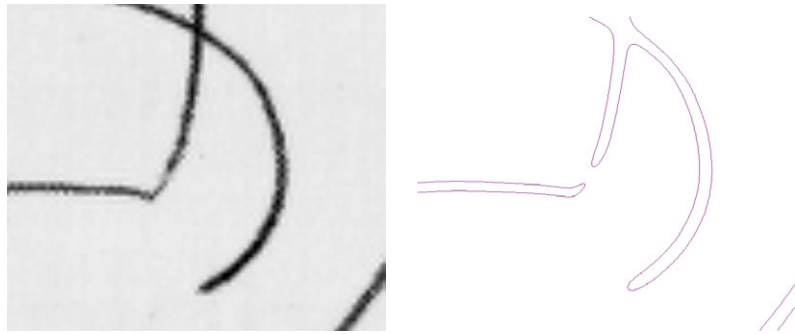


FIG. 5.15: Exemple typique d'erreur d'extraction. Ces erreurs surviennent aux endroits où le contraste minimal devient très faible et témoigne d'un point où l'artiste a fortement diminué sa pression appliquée sur le support artistique.

5.4.4 Images réelles

Nous présenterons dans cette section quelques exemples d'extractions des contours des traits sur des images réelles issues des bases Matisse-Kelly et Picasso présentées dans l'Annexe A. D'autres exemples seront visibles au Chapitre 6 à la Section 6.4 lorsque nous analyserons le contenu des images à partir de leurs ensembles d'adhérences.

Les exemples sont visibles sur les Figure 5.16 à 5.19. À gauche est présentée l'image originale, et l'ensemble des lignes de niveau sélectionnées par la méthode que nous proposons (Section 5.3.1). Chaque ensemble d'adhérence a une couleur différente. Cette couleur est choisie aléatoirement parmi 4 couleurs. Il se peut donc que deux ensembles d'adhérences aient la même couleur dans l'image.

Sur l'ensemble de ces résultats, nous voyons que l'extraction des contours est relativement bonne. L'unicité de chaque contour est toujours préservée. Des erreurs comme celle présentée à la Figure 5.15 sont visibles sur certains résultats, mais globalement elles sont assez rares. Un type d'erreur différent est dû à la compression JPEG de certaines images. Certains artefacts sont créés le long d'un trait. Les traits de dessins représentent souvent une forte discontinuité sur le fond de l'image, et les ondelettes ou la DCT utilisée par la compression peuvent créer des effets de *vagues* transverses aux traits de l'artiste, en

particulier quand la compression est forte. Certains de ces artefacts sont détectés. Ils sont généralement assez courts.

5.5 Conclusion

Dans ce chapitre, nous avons introduit une méthode automatique d'extraction des contours adaptée aux images ne contenant que des traits monodimensionnels. Nous utilisons cette méthode pour l'extraction des contours des traits dans une œuvre artistique, mais elle peut également s'appliquer de manière directe aux écritures manuscrites sur fond homogène. La méthode repose sur quatre étapes successives :

1. l'extraction de la carte topographique de l'image par FLLT [Monasse, 2000].
2. l'utilisation de l'approche *a contrario* proposée par [Desolneux et al., 2001] pour trouver les lignes de niveaux significatives dans la carte topographique.
3. l'utilisation d'une méthode de filtrage sans paramètre qui élimine les redondances dans chaque arbre monotone maximal.
4. l'utilisation de la structure hiérarchique subsistante pour grouper les lignes de niveaux en ensembles dit d'adhérences. Ces ensembles décrivent les contours d'un agrégat de traits.

Nous avons vu dans la partie expérimentale que cette méthode est assez robuste au bruit ainsi qu'aux variations géométrique et d'intensité des traits. Cette propriété est héritée du schéma *a contrario* utilisée. Par ailleurs, le nouveau principe de maximalité proposé pour filtrer l'arbre des lignes significatives permet d'obtenir un ensemble minimal de lignes de niveaux pour décrire les contours des traits. Soulignons que l'ensemble du processus d'extraction des contours ne repose que sur le paramètre ε fixé à 1 dans les chapitres 5 et 6. Les résultats sont très peu sensibles à des petites variations de ce paramètre puisque le nombre de lignes significatives varient logarithmiquement avec ε .

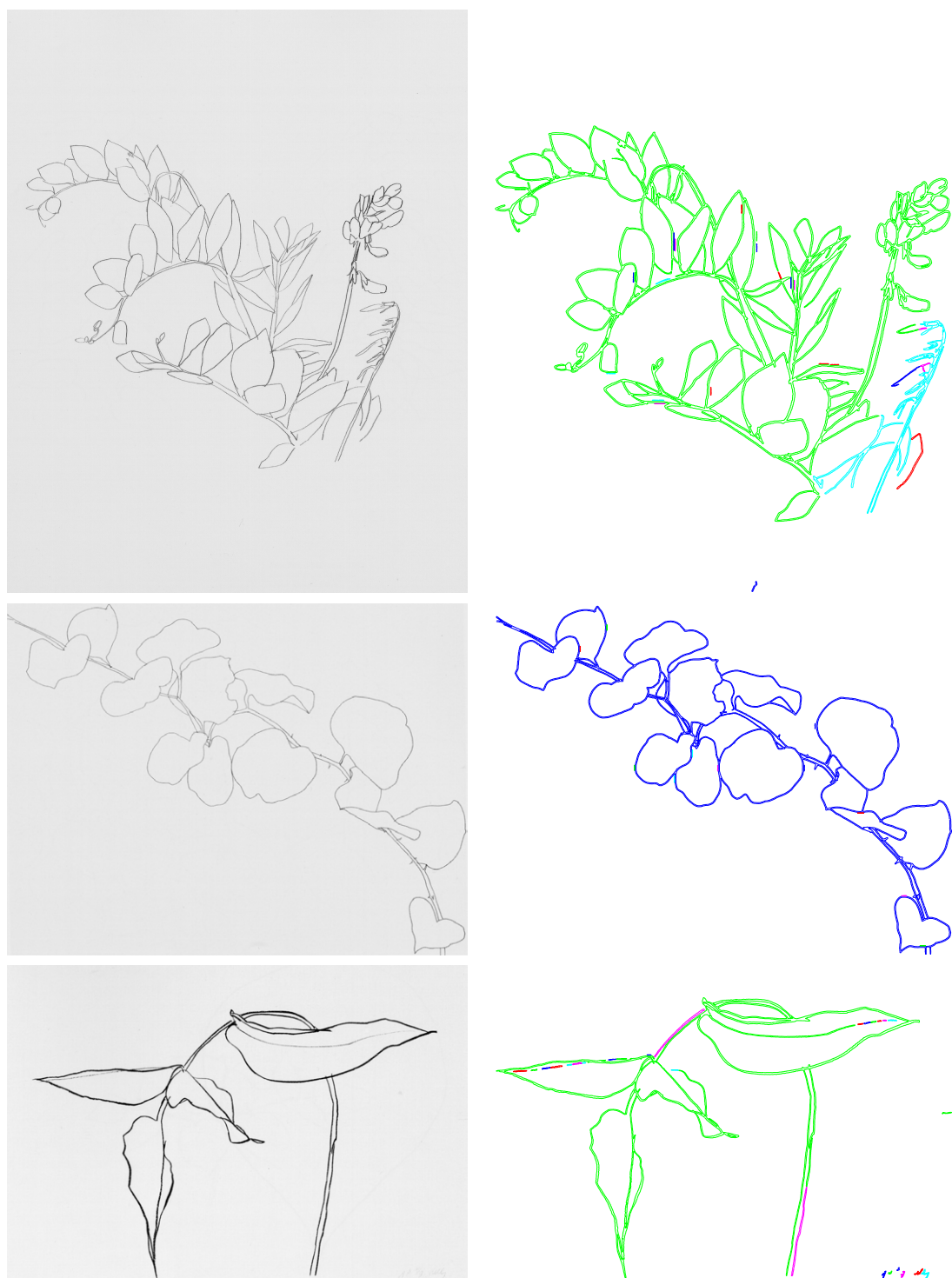


FIG. 5.16: Dessins de Ellsworth Kelly. A gauche les dessins originaux. A droite les contours extraits. Chaque couleur symbolise un ensemble d'adhérence.



FIG. 5.17: Dessins de Ellsworth Kelly. Même dispositions que sur la Figure 5.16.

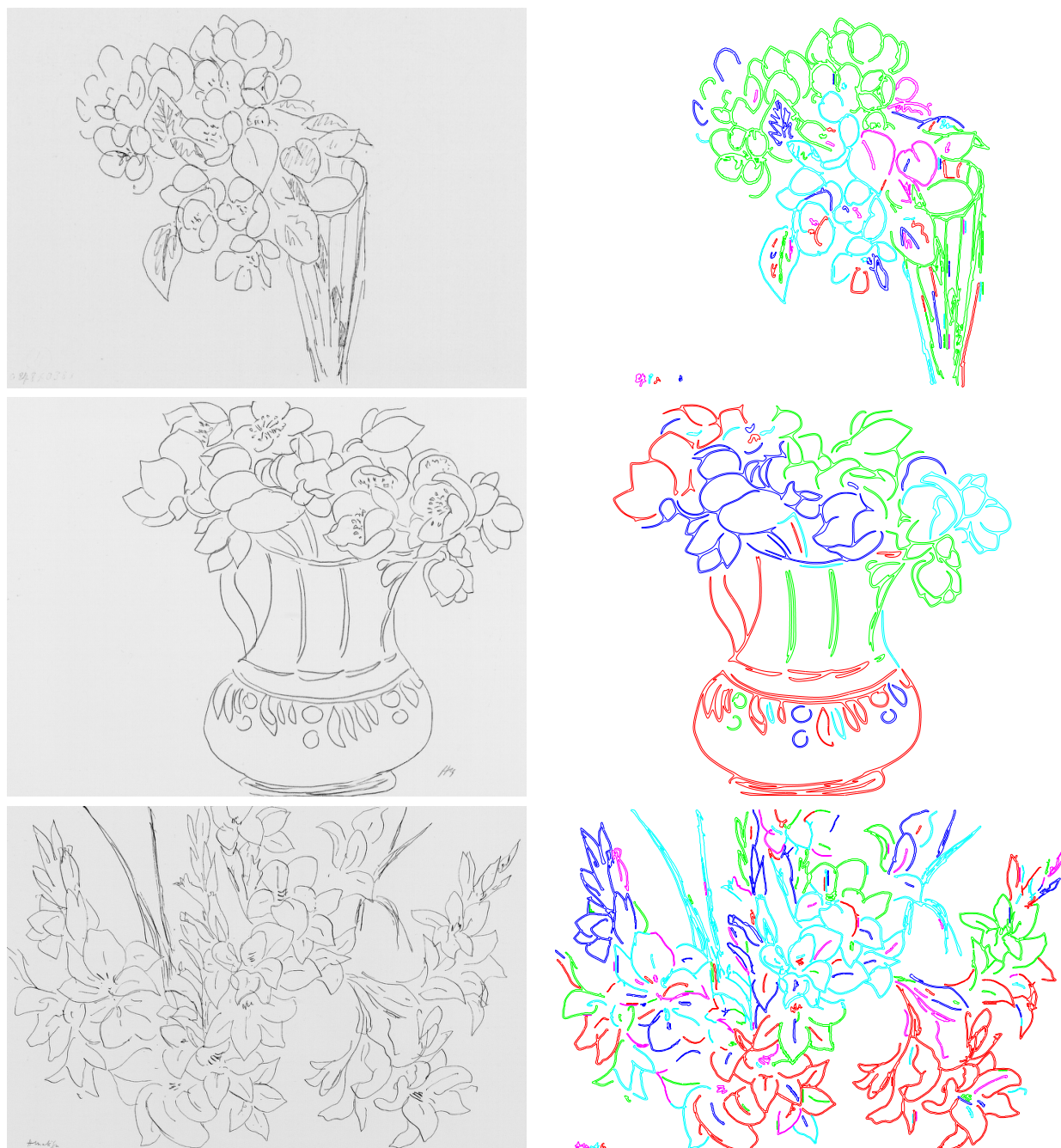


FIG. 5.18: Dessins de Matisse. Même dispositions que sur la Figure 5.16.



FIG. 5.19: Dessins de Picasso. Même dispositions que sur la Figure 5.16.

Chapitre 6

Analyse de l'impact pictural des œuvres au trait



FIG. 6.1: Exemple de regroupement d'images selon la description de l'impact pictural proposée dans ce chapitre.

L'analyse du contenu géométrique des œuvres est un aspect peu abordé dans la littérature bien qu'il recouvre une part très importante du contenu artistique [Willats, 1997]. Dans ce chapitre, nous proposons une description du contenu pictural des œuvres à partir des ensembles d'adhérences extraits à l'aide de la méthode présentée au chapitre précédent. Un exemple de regroupement d'œuvres selon la description de l'impact pictural proposée dans ce chapitre est visible sur la Figure 6.1. Après un état de l'art de thèmes connexes présenté à la Section 6.1, nous proposerons dans la Section 6.2 plusieurs méthodes pour analyser les ensembles d'adhérences et en extraire des éléments géométriques tels la courbure, les croisements de traits, les extrémités, les coins, etc. Plusieurs statistiques sur la présence de ces éléments permettent de définir un ensemble de descripteurs de l'image à la Section 6.3. Nous utiliserons ces descripteurs pour indexer des bases d'images à la Section 6.4.

6.1 Travaux connexes

Dans cette section nous justifions le choix de la courbure pour étudier les lignes de niveaux (Section 6.1.1) et présentons des méthodes pour la calculer (Section 6.1.2) ainsi que différents modes d'utilisation de cette courbure (Section 6.1.3). Nous terminons cette section par un bref état de l'art des méthodes de caractérisation du style de l'écriture (Section 6.1.5).

6.1.1 Le choix de la courbure

Une fois le contour discret d'un objet extrait, plusieurs représentations dérivées de la chaîne de coordonnées du contour peuvent être calculées. Citons la courbure locale, la distance centroïde, la distance angulaire, ou les coordonnées complexes [Gonzales and Woods, 2002]. Ces représentations fournissent des fonctions périodiques (pour les contours fermés), invariantes par translation et rotation. Dans notre application nous nous intéressons à la géométrie 1D des traits, même si nous utilisons leurs contours. Les formes que nous étudions sont donc très étirées ou d'un certain point de vue monodimensionnelles. Les distances centroïde, angulaire et complexe font appel au barycentre de la forme comme point de référence, c'est pourquoi elles ne sont pas adaptées à notre étude, car peu représentatives de ce que nous voulons atteindre.

La courbure locale d'un trait est une caractéristique très proche du geste de l'artiste. Un point de forte courbure dans un trait par exemple coïncide avec un fort ralentissement du geste [Plamondon and Privitera, 1999], et il représente visuellement quelque chose de remarquable à l'oeil. De nombreuses études psychophysiques ont démontré que ces points jouaient un rôle important dans la reconnaissance de contours. Les études originales d' [Attneave, 1954], de [Gollin, 1960] et de [Biedermann, 1987] ont montré que les points de forte courbure pris de manière isolés (avec un voisinage) suffisaient pour faire reconnaître à des sujets des objets avec un bon taux de succès. A l'inverse, éliminer ces points abaisse fortement ce taux. De nombreuses travaux ont permis de confirmer et compléter l'intuition d'Attneave [Wilson, 1985, Winter et al., 2002, Feldman and Singh, 2005, Ghosh and Petkov, 2006]. Les points de fortes courbures sont aussi utilisés en compression de formes, appelée aussi approximation polygonale [Wu, 2003]. Kristjánsson et Tse ont aussi montré plus globalement que les discontinuités de courbure étaient essentielles en tant qu'indices pour la reconnaissance rapide de forme [Kristjánsson and Tse, 2001]. Depuis la courbure est donc logiquement devenu un mode de représentation très fréquent en reconnaissance automatique de forme [Mokhtarian and Mackworth, 1992, Liu and Srinath, 1990, Agam and Dinstein, 1997, Del Bimbo and Pala, 1999, Kopf et al., 2005].

6.1.2 Calcul de la courbure locale

Il existe plusieurs manières de calculer la courbure le long d'un contour. Mokhtarian et Mackworth utilisent une formule impliquant la dérivée première et seconde de la direction après un filtrage gaussien pour enlever le bruit [Mokhtarian and Mackworth, 1986]. Cette méthode utilise l'estimation des dérivées premières et secondes en chaque point d'un contour ce qui impose l'utilisation d'un fort lissage Gaussien pour pouvoir être robuste. Ce lissage induit une forte détérioration géométrique et topologique des formes. Liu et Srinath utilise le gradient local perpendiculairement au contour obtenu avec un filtre de Sobel [Liu and Srinath, 1990]. Agam et Dinstein, utilise la différence de directions en chaque point, en utilisant le point précédent et suivant [Agam and Dinstein, 1997]. C'est l'approche que nous utiliserons à la Section 6.2.3. Toutes ces méthodes sont assez sensibles au bruit. Augmenter la fenêtre de calcul en prenant plus de points, ou en sous échantillonnant le contour, permet d'obtenir des fonctions de courbure plus lisses, mais les points de fortes courbure peuvent être sous-estimés. [Urdiales et al., 2002] proposent une mesure adaptative. [Han and Poston, 2001] proposent une estimation de la courbure en un point basée sur l'accumulation de la distance à une corde glissante.

6.1.3 Modes d'utilisation de la courbure

Pour décrire les contours extraits, nous pouvons utiliser classiquement des descripteurs multidimensionnels, comme les descripteurs de Fourier [Gonzales and Woods, 2002], ou des descriptions multi échelles [Mokhtarian and Mackworth, 1992, Del Bimbo and Pala, 1999, Wang et al., 1999, Kopf et al., 2005]. Les modèles autorégressifs [Dubois and Glanz, 1986] permettent aussi ce type de description, mais ont été jugés moins efficaces que les descripteurs de Fourier par exemple, dans le cadre de la classification de formes 2D [Kaupipinen et al., 1995]. De plus ces modèles sont mal adaptés à la présence de points de fortes courbures, car ceux ne sont fidèlement modélisés qu'au prix d'un grand nombre de coefficients.

Notre problématique est légèrement différente. Nous ne souhaitons pas faire de la reconnaissance d'objets, mais de l'analyse d'impact pictural. Comme nous l'avons remarqué à la Section 2.1.5, des sujets différents peuvent avoir le même impact pictural. Il est vrai aussi que le sujet a une influence forte sur l'impact visuel. Le dessin *chaotique* d'une branche de sapin aura difficilement le même impact que l'aspect rond et doux du dessin d'une feuille de chêne. Nous voyons par ailleurs qu'au niveau sémantique, nous pouvons tout aussi bien mettre ces deux dessins en commun dans une même catégorie "feuille d'arbre", ou au contraire les distinguer car ce sont des arbres différents. Le recherche sémantique s'applique mal au domaine des beaux-arts car les artistes n'hésitent pas à représenter des objets protéiformes. Par ailleurs les contours que nous détectons peuvent être très nombreux dans un dessin. En recherche et reconnaissance de forme, l'objet est au contraire souvent décrit par un seul contour.

Notre problématique n'est pas de retrouver un contour donné indépendamment d'un certain nombre de transformations par exemple. D'un certain point de vue, notre étude est plus proche d'une analyse de texture [Tuceryan and Jain, 1993, Heeger and Bergen, 1995]. Nous cherchons à caractériser l'effet visuel de traits par l'utilisation des marginales de mesures faites le long des contours.

C'est pourquoi à la Section 6.2 nous utiliserons un ensemble de descripteurs de niveaux bas (marginales sur la courbure) et moyens (caractéristiques de la répartition des points d'inflexions, présence et caractéristique de croisements de traits, points de fortes courbures etc).

6.1.4 Points d'intérêts, coins et jonctions

Dans les œuvres artistiques, il a été observé par [Willats, 1997, Leyton, 2006] que les points de fortes courbures sont une caractéristique essentielle de l'impact visuel d'une œuvre. Ces points sont appelés aussi points d'intérêts dans l'image (que nous noterons PDI dans tout ce chapitre), ou aussi jonctions et simplement coins. Parfois les coins sont vus comme un type particulier de jonctions (jonction dite en L). [Schmid et al., 2000] dresse un état de l'art des méthodes de détection des PDI proposées jusqu'en 2000. Ces méthodes s'organisent en trois familles selon ce qu'elles utilisent : les valeurs du signal image (par exemple, les descripteurs SIFT [Lowe, 2004]), un modèle paramétrique des jonctions (par exemple, [Sinzinger, 2008]) ou les contours dans l'image. Nous détaillerons dans la section suivante un peu plus cette dernière famille car nous utilisons aussi les contours des objets.

Détection des coins et jonctions

La recherche en détection de PDI le long de courbes a produit une littérature très riche. Les méthodes qui ont été proposées peuvent être regroupées en plusieurs catégories. Nous trouvons des méthodes utilisant une mesure locale autre que la courbure [Medioni and Yasumoto, 1986, Horaud et al., 1990, Pikaz and Dinstein, 1994, Ji and Haralick, 1998, Marji and Siy, 2003, Poyato et al., 2004], d'autres utilisent l'espace multi-échelle linéaire [Mokhtarian and Mackworth, 1992, Rattarangsi and Chin, 1992, Mokhtarian, 1995, Mokhtarian and Suomela, 1998, He and Yung, 2004, Zhang et al., 2007] ou des décompositions en ondelettes [Fayolle et al., 2000] et enfin d'autres utilisent directement la courbure.

Dans cette dernière catégorie, [Ghosh and Petkov, 2006] utilisent un seuil sur la mesure $|\kappa(i)|L(i)$ où $\kappa(i)$ est la courbure au i^{me} PDI et $L(i)$ la distance en nombre de points le long du contour entre le i^{me} PDI et le plus proche point j où la courbure vaut à nouveau la valeur $\kappa(j) \approx \kappa(i)$. Cette approche privilégie les points assez isolés de courbure forte. Wu en 2003 propose une méthode adaptative de détection des points dits *dominants* utilisant

la courbure locale mesurée sur un support de longueur variable [Wu, 2003]. L'objectif est de calculer la meilleure approximation polygonale. [Gu and Tjahjadi, 2000] proposent une méthode de mise en correspondance de formes à deux échelles, décrites par les coins les plus significatifs. Ceux ci sont extraits avec un seuil sur la courbure et un seuil sur la distance minimale entre deux PDI. [Neumann and Teisseron, 2002] utilisent les oscillations du contour au voisinage des extrema de courbure pour permettre de sélectionner les PDI.

Caractérisation d'un point d'intérêt

Pour caractériser un point d'intérêt, nous pouvons par exemple utiliser la valeur de la courbure seule. Mais d'autres facteurs en son voisinage peuvent aider à sa caractérisation. Dans le domaine de la recherche en perception visuelle, différentes caractéristiques ont été étudiées telles que l'angle de rotation formé par les perpendiculaires estimées aux points d'inflexions au voisinage du PDI [Hoffman and Singh, 1997], ou le même angle estimé aux points médians entre le PDI et les PDI voisins [Pasupathy and Connor, 1999], ou encore directement aux PDI voisins [Pasupathy and Connor, 1999]. Zusne propose une mesure de compacité du PDI utilisant la surface et le périmètre couverts jusqu'aux PDI voisins [Zusne, 1970]. Winter et al. effectuent une étude comparative de ces méthodes avec aussi d'autres mesures de compacité [Winter et al., 2002]. Les auteurs concluent que la meilleure corrélation entre mesure quantitative et impression perceptive est obtenue avec les mesures d'angles et en particulier celle estimée aux PDI voisins.

Selon le théorème de Jordan, si un contour est fermé, l'espace topologie complémentaire au contour est délimité en deux composantes connexes qui sont l'*intérieur* et l'*extérieur* du contour. L'oeil voit habituellement un contour fermé comme celui d'un objet posé sur un arrière plan. Cet effet perceptif créé aussi une discrimination fond-forme qui rappelle les études de la Gestalt [Köhler, 1967]. Alors, les PDI selon qu'ils sont concaves ou convexes n'ont pas le même impact perceptif, pour toutes autres caractéristiques prises égales (valeur de la courbure, et angle par exemple). Cette distinction a été observée et démontrée par Feldman et Singh [Feldman and Singh, 2005] et Fantoni et al. [Fantoni et al., 2005]. C'est pourquoi dans le cas des contours fermés il est intéressant de distinguer les PDI convexes ou concaves.

6.1.5 Caractérisation du style d'écriture

Le dessin et l'écriture sont analogue sur plusieurs points. Les lettrines dans les enluminures sont souvent considérées comme des dessins. La majorité des *kanji* chinois ont aussi une origine figurative. Une personne ne sachant pas décoder de tels idéogrammes (chinois, arabes, ...) peut y voir un geste purement esthétique abstrait. Les écritures hiéroglyphes égyptiennes ou aborigènes sont encore plus proches de ce que peut être un dessin. Les civilisations ont souvent développé leur alphabet à partir de dessins ayant chacun une

signification [Clouzot, 1997]. Les polices d'écritures ou styles typographiques sont autant de styles différents appliqués à des alphabets communs. Les différences d'écriture d'un même mot par plusieurs personnes sont du même ordre que les différences de styles entre plusieurs artistes dessinant un même objet. L'étude graphologique de l'écriture d'une personne pousse même le raisonnement au point de faire un lien entre le portait psychologique du scripteur et son style d'écriture à travers son impact visuel.

De nombreuses méthodes ont été proposées pour caractériser le style d'une écriture indépendamment du texte écrit. Les objectifs et applications sont multiples, et comprennent l'aide à la reconnaissance d'écriture, l'authentification de documents judiciaire et l'authentification de signatures. Pour un état de l'art des méthodes développées dans le cadre de ces applications, le chapitre 2 de l'ouvrage de référence de [Mullet, 2006] peut être consulté.

Il semble que la caractérisation de l'aspect de l'écriture par opposition à son sens littéral est une problématique proche de la notre. Non seulement parce qu'elle aborde une matière où le fond et la forme sont liés, mais aussi parce que cette matière est issue d'un procédé biomécanique proche de celui des dessins. L'écriture supporte néanmoins un certain nombre de contraintes géométriques et sémantiques. L'alphabet des artistes est illimité par opposition à celui de l'écriture. Les principaux obstacles à la reconnaissance automatique d'écriture résident dans la variabilité allographique, c'est à dire une certaine liberté qu'a le scripteur sur le choix du modèle des lettres (cursive, droite, etc.). En dessin, cette variabilité est illimitée. L'écriture est de plus géométriquement contrainte. C'est pourquoi des approches statistiques ou fréquentielles permettent d'obtenir de bons résultats en authentification de signatures par exemple.

Par ailleurs le sens syntaxique de l'écrit modifie profondément le regard du lecteur, et ce regard n'est pas le même que celui du spectateur face à un dessin. Le processus psychomoteur du scripteur est en réalité différent de celui de l'artiste. Si le geste physique a de nombreux points communs, le processus cognitif est lui très éloigné [Van Sommers, 1984]. L'écriture est une activité quasiment spontanée qui ne fait appel à aucun jugement continu du scripteur quant à son rendu [Schomaker, 1991]. L'écriture peut être modélisée par un mouvement de moteur asservi, répondant à des paramètres biologiques (fatigue, mobilité naturelle de l'ensemble bras poignet main, etc.) et historiques (enseignement perçu, culture, etc.) [Schomaker et al., 1989]. Le spectateur peut également recevoir un impact visuel différent d'une écriture selon sa culture, comme l'ont montré Tse et Cavanagh à propos des caractères chinois [Tse and Cavanagh, 2000]. L'artiste ne peut pas se contenter de reproduire de manière spontanée un alphabet appris. Il doit effectuer une analyse visuelle et mentale continue. Son vocabulaire et ses sujets de représentation étant illimités, la dissociation entre sujet et style devient réellement impossible. Vouloir dissocier ces deux contenus est un problème mal posé car nous ne pouvons pas remonter ce processus cognitif et créatif jusqu'à l'artiste.

6.2 Analyse du contenu géométrique

Nous proposons dans cette section un ensemble de méthodes pour analyser le contenu pictural des dessins à partir de leurs ensembles d'adhérences. Nous lisons tout d'abord très légèrement les lignes de niveaux des ensembles d'adhérence (Section 6.2.1) avant d'estimer le rayon moyen de l'outil utilisé pour les traits de dessin (Section 6.2.2). Cette information sera très utile par la suite pour ajuster les méthodes de détection des points d'intérêts. Nous calculons de plus la courbure le long des lignes de niveaux (Section 6.2.3). Ensuite, nous détectons dans l'image les points d'inflexions (Section 6.2.4), les points d'intérêts candidats (Section 6.2.5), les extrémités de traits (Section 6.2.8), les croisements de traits (Section 6.2.6) et les coins (Section 6.2.7).

6.2.1 Lissage des contours

La courbure est une information très sensible au bruit. C'est pourquoi un lissage des lignes de niveaux formant les ensembles d'adhérence est nécessaire. Nous souhaitons utiliser un lissage qui préserve la structure topologique des lignes de niveaux. Un des lissages les plus connus est le lissage Gaussien. Ce lissage ne préserve pas la topologie des courbes. En particulier, il peut fusionner ou scinder les lignes de niveaux de la carte topographique. Nous avons choisi d'utiliser le lissage affine géométrique [Moisan, 1998] qui est un lissage par courbure. Appliqué à l'image en niveaux de gris, ce lissage se traduit au niveau de la carte topographique des lignes de niveaux par l'équation aux dérivées partielles (EDP) suivante. Une ligne de niveau $s \mapsto \mathcal{L}(s, 0)$ évolue en fonction de t selon l'EDP :

$$\frac{\partial \mathcal{L}}{\partial t}(s, t) = \kappa^{1/3}(s, t) \mathbf{N}(s, t), \quad (6.1)$$

où κ représente la courbure locale, s l'abscisse curviligne, et $\mathbf{N}$ le vecteur normal à $\mathcal{L}(\cdot, t)$ en $\mathcal{L}(s, t)$. Ce lissage respecte notamment la structure topologique des lignes au niveau théorique et pratique [Moisan, 1998]. Ceci nous assure par exemple que les lignes de niveaux ne se croisent pas après lissage. Bien que l'invariance affine caractéristique de ce lissage ne soit pas indispensable dans notre cas, nous avons utilisé l'implémentation de ce lissage disponible dans la librairie Megawave [Froment et al., 2007].

Nous devons ensuite choisir à quelle échelle est lissé le contour. Le lissage affine étant réalisé par une EDP, un paramètre d'arrêt est nécessaire. Dans le cadre de l'implémentation de Moisan, ce paramètre d'échelle T est normalisé de telle sorte qu'à l'échelle T finale, un cercle de rayon T a complètement disparu par lissage. Ceci permet de calibrer ce paramètre d'arrêt par rapport à l'échelle minimale des détails que nous souhaitons conserver. Dans le cadre de cette application, nous utiliserons toujours le paramètre $T = 0.5$. Ceci signifie qu'après le lissage, les détails de dimension 1 pixel ont disparu. L'implémentation

6.2. Analyse du contenu géométrique

de Moisan permet aussi de fixer un pas d'échantillonnage relativement constant (abscisse curviligne) comme nous le verrons dans la Section expérimentale. Le pas d'échantillonnage des lignes de niveaux sera fixé également à $\Delta s = 0.5$.

Remarquons que nous n'utilisons qu'une seule échelle de lissage, très faible, égale à un pixel. En ce sens, notre méthode se distingue d'une approche multi-niveaux. Nous n'utilisons pas le lissage comme un outil d'analyse à différentes échelles (comme [Mokhtarian, 1995]) mais seulement comme une méthode d'élimination du bruit pixelique.

6.2.2 Estimation du rayon de l'outil

Le rayon de l'outil est une donnée importante de notre analyse car elle permettra par la suite de régler automatiquement les paramètres d'extraction des éléments géométriques tels que les extrémités de traits, les jonctions et les coins. Elle constitue également un descripteur géométrique. A partir des ensembles d'adhérences, nous pouvons extraire assez facilement le rayon de l'outil moyennant l'hypothèse suivante. Nous supposons qu'un agrégat de traits d'artiste est constitué de traits ayant été réalisés avec un outil laissant une trace d'épaisseur constante. Alors, pour un agrégat Φ décrit par un ensemble d'adhérence composé de n lignes $\mathcal{L}_i$, le rayon de l'outil R_Φ est estimé par :

$$R_\Phi \approx \frac{\text{surface}}{\text{périmètre}} = \frac{-\sum_{i=1}^n \text{sign}(\mathcal{L}_i) * S_i}{\sum_{i=1}^n P_i}, \quad (6.2)$$

où est S_i et P_i sont la surface et le périmètre respectivement de la ligne $\mathcal{L}_i$. Le signe de $\mathcal{L}_i$ correspond au sens d'orientation du gradient lorsque nous allons de l'extérieur vers l'intérieur de $\mathcal{L}_i$ (Section 5.2.1). L'hypothèse ci dessus est restrictive, mais elle permet une estimation simple de l'épaisseur de l'outil. Dans un cas plus général, il faudrait envisager une méthode plus locale de suivi de contour. Cette méthode pourrait consister à suivre un contour de l'ensemble d'adhérence et à projeter le points dans les directions perpendiculaires au contour pour trouver les point les plus proches. Une analyse des modes de l'histogramme des épaisseurs [Desolneux et al., 2003] sur l'ensemble d'adhérence permettrait de conduire à l'estimation de plusieurs épaisseurs différentes. Nous pourrions aussi envisager une méthode de recherche locale de constance de largeur telle qu'étudiée par les Gestaltistes [Köhler, 1967, Kanizsa, 1996].

6.2.3 Calcul de la courbure

Utilisation de la variation de l'orientation

La courbure $\kappa(s_i)$ est calculée en chaque point d'abscisse s_i comme étant la variation $\Delta\theta$ de l'orientation de la tangente approximée par le segment reliant deux points

consécutifs mesurée dans le sens trigonométrique divisée par Δs (Figure 6.2).

$$\kappa(s_i) = \theta'(s_i) \approx \frac{\Delta\theta}{\Delta s} = \frac{\theta_{s_i s_{i+1}} - \theta_{s_{i-1} s_i}}{(s_i s_{i+1} + s_{i-1} s_i)/2}. \quad (6.3)$$

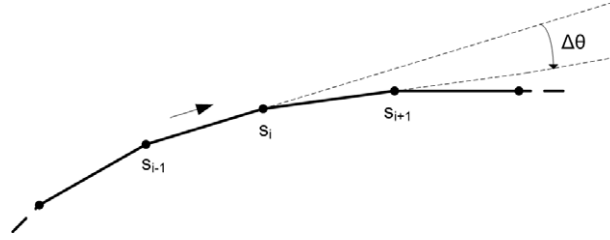


FIG. 6.2: Calcul de la courbure. La courbure se calcule en chaque point comme étant la variation de l'orientation en parcourant la courbe par rapport au pas d'échantillonnage Δs .

En effet, si nous notons $\mathbf{T}(s) = \mathcal{L}'(s)$ le vecteur tangent à la ligne de niveau $\mathcal{L}$ alors, le vecteur normal $\mathbf{N}(s)$ (formant avec $\mathbf{T}(s)$ une base orthonormale directe du plan) permet de définir le rayon de courbure $1/\kappa(s)$: $\mathbf{N}(s) = 1/\kappa(s)\mathbf{T}'(s)$. Donc si

$$\mathbf{T}(s) = \begin{pmatrix} \cos(\theta(s)) \\ \sin(\theta(s)) \end{pmatrix},$$

alors,

$$\mathbf{T}'(s) = \theta'(s) \begin{pmatrix} -\sin(\theta(s)) \\ \cos(\theta(s)) \end{pmatrix} = \theta'(s)\mathbf{N}(s).$$

Nous voyons alors que dans le cadre de l'utilisation d'une abscisse curviligne, nous avons $\kappa(s) = \theta'(s)$. Cette estimation de la courbure rendue possible par l'utilisation d'une abscisse curviligne (pas d'échantillonnage constant), et est moins sensible au bruit et moins coûteuse qu'une méthode utilisant les dérivées premières et seconde du signal par exemple [Mokhtarian, 1995].

Convention pour le signe de la courbure

La courbure en chaque point est une mesure qui peut être positive ou négative, selon le sens de parcours choisi le long d'une courbe fermée (sens horaire ou anti-horaire) et selon la géométrie locale (convexe ou concave). Nous choisissons les deux conventions suivantes. Les lignes de niveaux négatives sont parcourues dans le sens horaire. Ceci a pour conséquence que la courbure sera négative dans les régions convexes et positive dans les régions concaves. Inversement, les lignes de niveaux de type positif sont parcourues

dans le sens anti-horaire pour inverser le signe de la courbure par rapport aux lignes de niveaux de type négatif.

Ces conventions nous assurent tout d'abord que les éléments picturaux que nous cherchons à détecter (croisements de traits, extrémités, etc.) sont représentés par des portions de ligne de courbure de même signe quelque soit leur lieu d'apparition dans l'agrégat auxquels ils appartiennent (Figure 6.3). Par exemple, une extrémité de trait représentée par une ligne de niveau de type négatif ou positif aura toujours une courbure de signe négatif (par exemple e sur la Figure 6.3 à droite). Inversement, les croisements de traits seront toujours le siège d'une courbure de signe positif.

La courbure des traits (et non de leurs contours) est en effet une information importante du contenu pictural géométrique [Willats, 1997, Leyton, 2006, Grabli et al., 2004]. Nous souhaitons calculer la courbure des traits à partir des lignes de niveaux passant par leurs contours. Ces conventions nous permettent de sélectionner indirectement une partie des valeurs de courbure qui soit assez représentative de la courbure le long des traits, bien que nous utilisions leurs contours. En effet les valeurs $\kappa(s) \leq 0$ permettent de suivre un seul côté des traits, sans prendre en compte les points liés aux croisements. Ces valeurs intègrent néanmoins les points liés aux extrémités de traits (Figure 6.3). Nous verrons à la Section 6.3.1 comment diminuer leur impact.

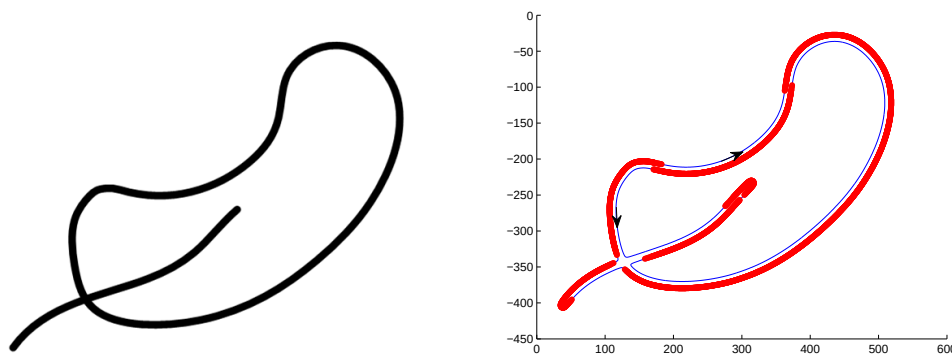


FIG. 6.3: Convention choisie pour le signe de la courbure. À gauche un dessin synthétique, à droite les lignes de niveaux extraites. Les lignes de niveaux négatives sont parcourues dans le sens horaire. Leur courbure est alors négative dans les régions convexes et positive dans les régions concaves. Inversement, les lignes de niveaux de type positif sont parcourues dans le sens anti-horaire pour inverser le signe de la courbure par rapport aux lignes de type négatif. Les portions de lignes de niveaux en bleu correspondent à des courbures positives $\kappa(s) \geq 0$ avec le sens de parcours de la ligne indiqué par des flèches. Les positions de lignes en rouge correspondent à des courbures négatives $\kappa(s) \leq 0$.

6.2.4 Extraction des points d'inflexions

[Lowe, 1985] utilise les points d'inflexions le long des contours d'une forme comme éléments caractéristiques pour sa reconnaissance. Dans le domaine artistique, [Leyton, 2006] utilise de même les points d'inflexions pour découper un contour en segments caractéristiques. Nous choisissons dans cette section d'extraire les points d'inflexions à l'aide d'une méthode relativement simple plutôt que de chercher par exemple à détecter les parties plates d'une courbe avec une méthode *a contrario* plus complexe [Sur, 2004]. Bien que le lissage affine ait éliminé un grand nombre d'oscillations pixeliques le long des lignes de niveaux, les variations à très petites échelles peuvent créer des points d'inflexions peu perceptibles. Pour éviter de les détecter, nous appliquons tout d'abord un seuil $\kappa_t = 1/(k_c R_\Phi)$ sur la courbure absolue $|\kappa|$ tel que : $|\kappa(i)| < \kappa_t \Rightarrow \kappa(i) = 0$. Nous verrons comment fixer le paramètre k_c à la Section 6.4.3.

Les points d'inflexions sont estimés comme étant les milieux des segments des domaines où la courbure est inférieure à κ_t en valeur absolue lorsque la courbure aux deux extrémités du segment change de signe (Figure 6.4). Sur la Figure 6.4 est visible à droite un dessin avec les points d'inflexions superposés à une partie de dessin.

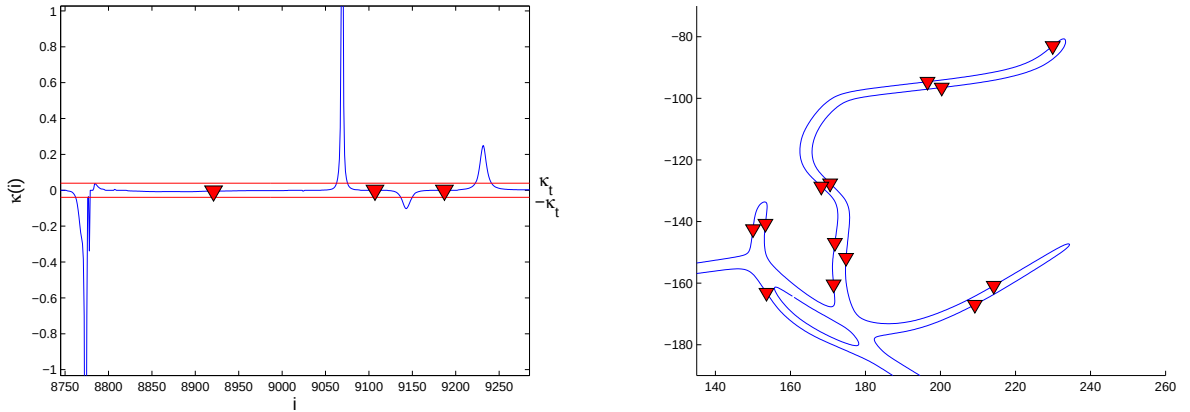


FIG. 6.4: Estimation des points d'inflexion. A gauche, les points d'inflexions (triangles rouges) sont estimés comme étant les milieux des segments des domaines où la courbure est inférieure à κ_t en valeur absolue lorsque la courbure aux deux extrémités du segment change de signe. Les deux droites horizontales rouge à gauche symbolisent le seuil κ_t . A droite, un exemple de dessins avec les points d'inflexions estimés.

6.2.5 Extraction des extrema de courbure

Cette section présente la méthode proposée pour extraire un ensemble d'extrema de courbures. Ces extrema sont par la suite candidats à la classification selon la taxonomie

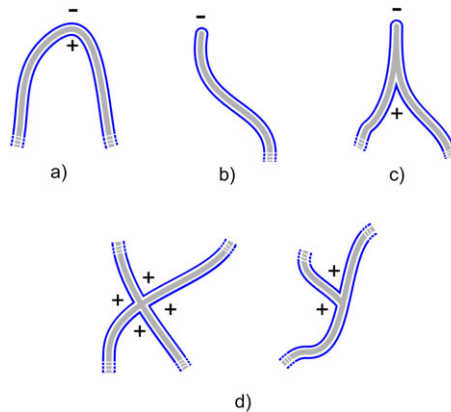


FIG. 6.5: Une taxonomie des points d'intérêts dans les œuvres au trait reposant sur les points de fortes courbures. a) un point de forte courbure simple. b) une extrémité de trait. c) un coin d) Deux types de jonctions : en X et en T. Les signes + et - symbolisent les signes de la courbure le long des lignes de niveaux dans les régions d'intérêt formées par les lignes de niveaux en noir. Les traits sont représentés en gris.

des points de fortes courbures visible sur la Figure 6.5. A partir des portions du signal de courbure extraites à la section précédente respectant $|\kappa(i)| > \kappa_t$ (où κ_t a été défini à la section précédente), nous estimons de la même manière les passages par zéro de la dérivée du signal. Cette sélection assez permissive conduit à un ensemble d'extrema dont beaucoup d'éléments correspondent au type a) de la Figure 6.5. Le but ici est de minimiser la perte possible d'un élément des autres catégories. Les méthodes que nous présentons dans les sections suivantes permettent d'isoler chacun des types sur cet ensemble de candidats. Enfin remarquons que selon le signe de la courbure, l'extremum peut être positif ou négatif. Grâce à la convention de parcours choisie à la Section 6.2.3, une extrémité de trait est toujours située en un extremum négatif, et un croisement de trait est le siège de plusieurs extrema positifs. Un coin (catégorie c) de la Figure 6.5) est le lieu d'un extremum positif et d'un extremum négatif. La Figure 6.6 présente des exemples d'extrema superposés à un dessins.

6.2.6 Croisements de traits

Les croisements ou jonctions de traits sont des caractéristiques importantes d'un dessin. Ils symbolisent souvent la présence d'une occlusion dans la scène 3D et sont corrélés à un certain degré de perspective dans le système de projection de [Willats, 1997].

Un extremum positif indique un croisement de traits potentiel. Pour décider si cet extremum appartient à un croisement nous construisons tout d'abord un disque de rayon $k_j R_\Phi$ centré sur l'extremum. Rappelons que R_Φ est l'estimation du rayon de l'outil utilisé

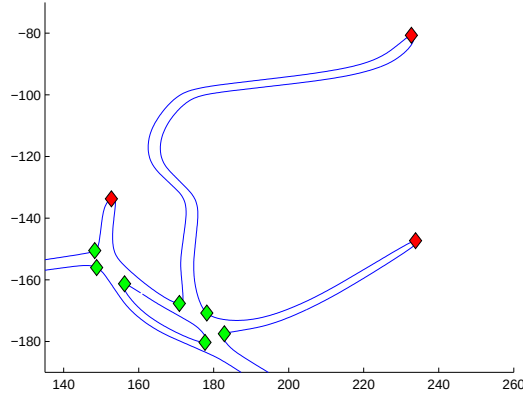


FIG. 6.6: Extrema de courbure seuillés détectés. Les extrema positifs (resp. négatifs) sont représentés en vert (resp. rouge).

par l'artiste. Si ce disque contient trois segments ou plus d'une ligne de niveau appartenant à l'agrégat auquel est associé l'extremum, alors cet extremum sera classé comme faisant partie d'un croisement de traits. Un croisement unique de traits étant souvent le siège de plusieurs extrema de ce type, nous fusionnons les extrema à une distance Euclidienne inférieure à $k_j R_\Phi$ après avoir analysé l'ensemble de l'agrégat. Le choix du paramètre k_j sera détaillée à la Section 6.4.4.

Lorsque le disque contient exactement deux sections de lignes de niveaux, l'extremum est potentiellement un coin (catégorie c de la Figure 6.5). Les coins sont caractérisés en détails à la Section 6.2.7 qui suit. Remarquons que nous pourrions classer plus précisément les jonctions (en X, en T, ...) à partir du nombre de sections de lignes de niveaux. Pour plus de simplicité et de robustesse nous avons décidé de constituer uniquement une catégorie liée aux croisements de traits, sachant qu'ils sont souvent synonymes d'une occlusion dans la scène 3D.

6.2.7 Coins

Nous définissons un coin comme étant un point de forte courbure positif le long d'un trait qui ne soit pas causé par un croisement. Ce type d'élément pictural a un impact fort comme le remarque [Leyton, 2006] pour qui les discontinuités de la courbure expriment une tension émotionnelle conséquente à une déformation picturale de la forme dessinée. L'importance de ces structures sur l'impression picturale est illustrée sur la Figure 6.8. Pour pouvoir calculer un descripteur global de l'intensité des coins dans une image, nous recherchons une caractéristique qui soit additive. C'est pourquoi nous utiliserons la surface relative du coin par rapport à l'image.

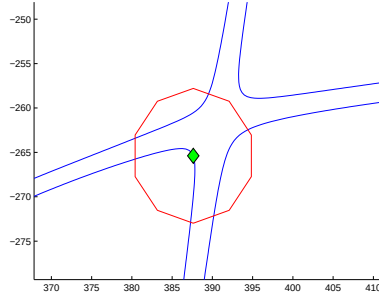


FIG. 6.7: Analyse de la région proche d'un extremum positif. Un disque est centré sur l'extremum. Si ce disque contient trois segments ou plus d'une ligne de niveau appartenant à l'agrégat auquel est associé l'extremum, alors cet extremum sera classé comme faisant partie d'un croisement de traits. Un croisement unique de traits étant souvent le siège de plusieurs extrema de ce type, nous fusionnons les extrema qui sont distants de moins de $k_j R_\Phi$ après avoir analysé l'ensemble de l'agrégat.

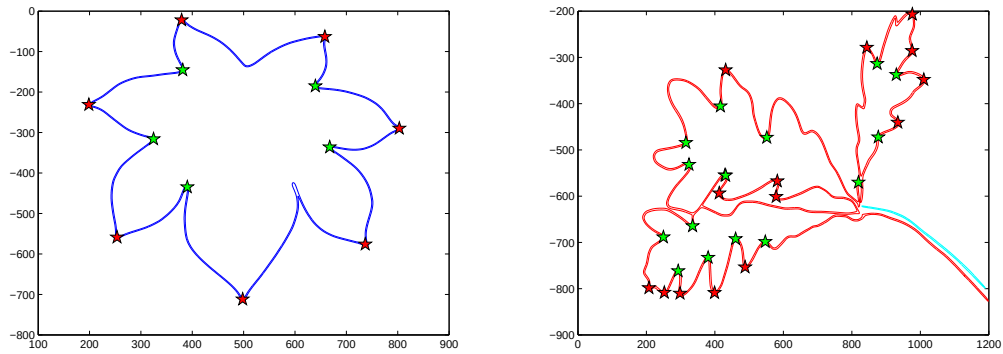


FIG. 6.8: Présence de coins dans une œuvre. Les étoiles rouges (resp. vertes) sont des coins positifs (resp. négatifs). Chaque ensemble d'adhérence est représenté par une couleur différente.

Pour mesurer la surface relative d'un coin en un extremum p_m , nous appliquons l'algorithme suivant. Nous construisons itérativement le polygone $p_{m-i} \dots p_m \dots p_{m+i}$ pour $i > 0$. Nous itérons $i \leftarrow i + 1$ tant que ce polygone ne contient aucun autre point que l'ensemble des $\{p_{m-i}, \dots, p_{m+i}\}$. Dès que cette propriété n'est plus vérifiée, la force du coin est estimée comme étant la surface relative (par rapport à la surface total de l'œuvre) couverte par le polygone maximal.

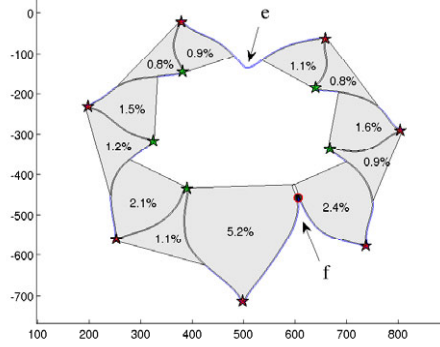


FIG. 6.9: Force des coins. La surface relative des régions grises par rapport à la surface totale de l'œuvre correspond à l'estimation choisie de la force d'un coin. En e , le maximum n'a pas une courbure supérieure à $2\kappa_t$ et ne fait donc pas parti des extrema candidats. En f , les deux traits se rejoignent très progressivement. L'extremum est alors classé comme étant un croisement de trois traits.

Cette estimation de la surface a l'avantage d'être simple. Nous imposons un critère d'arrêt simultané de chaque côté de l'expansion du polygone pour que l'itération s'arrête à la même distance géodésique. Nous aurions pu dissocier le processus de construction itérative en considérant le polygone $p_{m-i_1} \dots p_m \dots p_{m+i_2}$ avec arrêt de l'incrémentation de i_1 ou i_2 dès qu'un non respect de la condition d'arrêt apparaît. Cette stratégie nous semble néanmoins moins robuste car elle conduit à une surestimation de l'impact visuel dans le cas des coins asymétriques (c.f. Figure 6.10).

Indépendamment de sa surface relative, un coin peut être concave ou convexe (Figure 6.8). Selon le cas, l'impact pictural est différent comme l'illustre les dessins de la Figure 6.8. Cette particularité a été étudiée expérimentalement et modélisée par [Fantoni et al., 2005, Feldman and Singh, 2005]. Décider si un coin est convexe ou concave est un problème qui en pratique est souvent mal posé. En effet, si le trait décrivant un objet est ouvert, il peut être difficile de définir l'intérieur et l'extérieur de cet objet. Il arrive aussi que le contour ne décrive simplement pas celui d'un objet visible. C'est pourquoi nous proposons ci-après un type (positif ou négatif) qui décrit l'orientation du coin par rapport au barycentre G de l'ensemble d'adhérence auquel il appartient. Nous disons qu'un coin est positif s'il est orienté dans la direction opposée à G et inversement. Cette orientation est estimée à partir de celle du vecteur $\overrightarrow{p_{m-i_\Phi} p_m} + \overrightarrow{p_{m+i_\Phi} p_m}$ où i_Φ est telle que $i_\Phi \Delta s \approx R_\Phi$. Nous verrons que cette mesure correspond souvent avec la convexité ou la concavité réelle

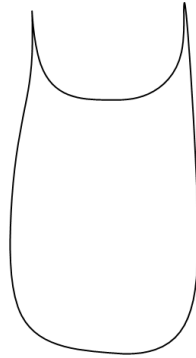


FIG. 6.10: Coin asymétrique. Nous aurions pu dissocier le processus de construction itérative en considérant le polygone $p_{m-i_1} \dots p_m \dots p_{m+i_2}$ avec arrêt de l'incrémement de i_1 ou i_2 dès qu'un non respect de la condition d'arrêt apparaît. Cette stratégie conduit à une surestimation de l'impact visuel dans le cas des coins asymétriques Ici en utilisant cette approche, les deux coins auraient chacun une surface relative proche de la surface totale de la forme.

du coin lorsque celle ci est définie. La situation typique où cette correspondance n'est pas valide est celle où la forme se replie vers elle-même comme l'illustre la Figure 6.11.

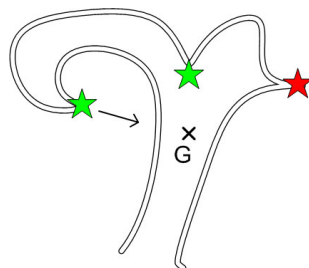


FIG. 6.11: Coin positif ou négatifs. Nous disons qu'un coin est positif s'il est orienté dans la direction opposée au barycentre G et inversement. La situation typique où ce type est mal corrélé à la convexité ou concavité réel du coin est celui où la forme décrit un recourbement complet vers elle-même comme à gauche sur cette figure.

6.2.8 Extrémités de traits de dessins

Les extrémités de traits sont un élément important du contenu pictural dans les œuvres au trait. Ils sont le lieu d'une *rupture* visuelle forte. Ils permettent aussi d'estimer la longueur moyenne des traits en mesurant la longueur total des traits divisée par le nombre total d'extrémités. Pour une même longueur totale de traits, une œuvre constituée uniquement de traits courts et isolés a un impact différent d'une œuvre faite d'un seul trait fermé par exemple.

Pour extraire les extrémités de traits, nous analysons l'ensemble des extrema négatifs. Pour chaque extremum négatif p_m appartenant à un ensemble d'adhérence Φ , nous considérons les deux points le long de la ligne de niveau p_i, p_j où $i < m < j$ qui sont à une distance géodésique $k_e R_\Phi$ de p_m (au pas d'échantillonnage prêt). Nous adoptons la règle de décision suivante. Si la distance entre p_i et p_j est inférieure à $3R_\Phi$, et si il n'existe aucun point p_q dans le triangle $p_i p_m p_j$ tel que $q < i$ et $q > j$, alors p_m est une extrémité de trait. Le choix du paramètre k_e sera détaillé à la Section 6.4.4.

Cette règle de décision est illustrée sur la Figure 6.12. Le seuil $3R_\Phi$ sur la distance entre p_i et p_j adopte une marge d'erreur de 50% sur R_Φ puisque pour un trait parfaitement segmenté cette distance devrait être égale à $2R_\Phi$.

Remarquons que nous n'utilisons pas la même méthode que celle utilisée pour détecter les croisements. L'utilisation d'un disque centré autour d'une extrémité pour analyser le nombre de sections de contour couverts introduit en effet un risque non nul de fausse détection à cause de portion de contours voisines passant à proximité de l'extrémité. En utilisant le triangle $p_i p_m p_j$, nous pouvons facilement limiter la surface d'étude à celle située *sous* l'extremum (Figure 6.12a).

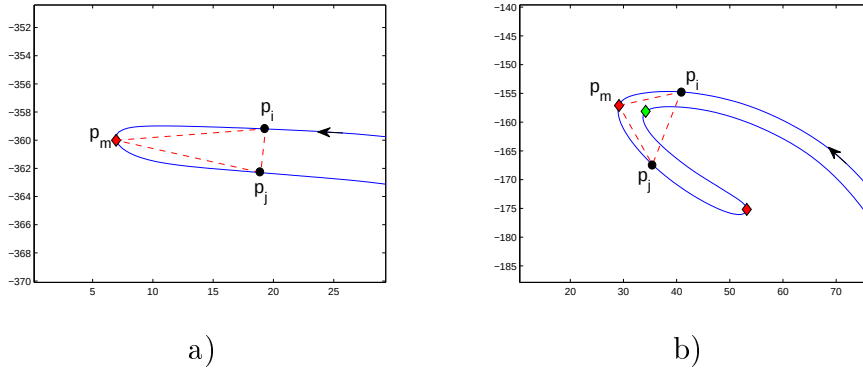


FIG. 6.12: Détection des extrémités de traits. A gauche un extremum négatif classé comme étant une extrémité de trait. A droite l'extremum présenté n'est pas une extrémité de trait car le triangle contient des points p_q ne faisant pas partie de la portion de contour $p_i \leq p_q \leq p_j$.

6.3 Descripteurs et mesure de similarité

Dans cette section nous présentons les descripteurs utilisés pour l'indexation des dessins, ainsi que la distance choisie pour comparer deux dessins.

6.3.1 Descripteurs

Pour illustrer et évaluer les méthodes proposées, la partie expérimentale utilisera un schéma de recherche d'images par l'exemple. À partir de l'analyse du contenu pictural présentée dans les sections précédentes, 11 descripteurs scalaires sont calculés. Ces descripteurs sont résumés dans le Tableau 6.1.

Les trois premiers descripteurs sont calculés sur la distribution des valeurs négatives de courbures supérieures à $-\kappa_t$. Ces valeurs permettent de ne considérer localement qu'un côté de la ligne de niveau représentative du contour d'un trait, en vertu de l'orientation définie à la Section 6.2.3 et sur la Figure 6.3. Nous n'utilisons aussi pour ces mesures que les valeurs $\kappa(i) \leq \kappa_t$ afin de ne pas corrélérer ces trois premiers descripteurs avec les suivants qui sont partiellement basés sur l'utilisation du critère κ_t . Ceci permet aussi de limiter l'influence des extrémités de traits sur ces mesures. Ces trois descripteurs donnent une indication de la courbure des traits indépendamment du reste du contenu géométrique (extrémités, croisements, etc.). Les valeurs de courbures utilisées sont normalisées par la diagonale du support du dessin afin que ces descripteurs soient invariants par changement de résolution de l'œuvre.

Les descripteurs 4 à 8 sont les densités linéaires respectivement des points d'inflexions, extrémités de traits, croisements de traits, coins positifs et négatifs. Chaque densité linéaire correspond au nombre total de ces éléments dans l'image divisé par la longueur totale des lignes de niveaux des ensembles d'adhérence extrait dans l'image. Ces densités linéaires sont aussi normalisées par la diagonale de l'image afin qu'elles soient invariantes par changement de résolution.

Les descripteurs 9 et 10 sont les sommes des surfaces relatives respectivement des coins positifs et négatifs dans l'image. Bien que ces deux descripteurs soient en partie redondants par rapport aux descripteurs 4 et 8, ils introduisent néanmoins une information supplémentaire. Nous avons observé expérimentalement que cette redondance n'était pas néfaste aux résultats de classification. Il a été observé également qu'une analyse ACP dégradait les résultats. Enfin le dernier descripteur est la valeur moyenne des rayons de l'outil R_Φ (introduit à la Section 6.2.2) pondérés par la longueur de chaque ensemble d'adhérence Φ , et normalisée par la diagonale de l'image.

Soulignons que les densités linéaires d'extrémités et de croisements ont été également proposées et étudiées par [Julesz, 1986] dans le cadre de la perception *préattentive* et la discrimination de textures à base de traits, appelées *textons*. Des expériences sur quelques images de textons de Julesz sont présentées à la Section 6.4.10.

TAB. 6.1: Ensemble de mesures proposées pour décrire le contenu pictural

Descripteurs
1. moyenne de la distribution des courbures $\kappa(i) \in [-\kappa_t, 0]$
2. écart type de la distribution des courbures $\kappa(i) \in [-\kappa_t, 0]$
3. kurtosis de la distribution des courbures $\kappa(i) \in [-\kappa_t, 0]$
4. densité linéaire des points d'inflexions
5. densité linéaire des extrémités de traits
6. densité linéaire des croisements de traits
7. densité linéaire des coins positifs
8. densité linéaire des coins négatifs
9. somme des surfaces relative des coins positifs
10. somme des surfaces relative des coins négatifs
11. rayon moyen pondéré des rayons d'agrégats dans l'image

6.3.2 Mesure de similarité

Après avoir été calculé, chaque descripteur du Tableau 6.1 est centré puis normalisé par la moyenne et l'écart-type de la distribution de ce descripteur sur la base. Si nous notons $V^i(k), k \in [1, 11]$ les 11 descripteurs normalisés pour une œuvre i , alors la mesure de similarité $d(V^1, V^2)$ entre deux œuvres est la distance Euclidienne pondérée suivante :

$$d(V^1, V^2) = \sqrt{\sum_{k=1}^{11} w_i (V_i^1(k) - V_i^2(k))^2}, \quad (6.4)$$

où $W_{11D} = (w_1, \dots, w_{11})$ est un vecteur de pondération à 11 dimensions.

6.4 Expériences

Des expériences sont tout d'abord réalisées sur des images de synthèse (Section 6.4.1 à Section 6.4.4) pour étudier la précision de la mesure de courbure et de mesure du rayon de l'outil, ainsi que pour ajuster les paramètres k_c , k_e et k_j vus en Section 6.2.4, 6.2.8 et 6.2.6. La Section 6.4.5 présente la base d'œuvres classées que nous avons utilisé pour comparer quantitativement notre méthode avec une approche différente présentée à la Section 6.4.7. Dans toutes les expériences, les images sont codées sur 8 bits en niveaux de gris, et le pas de quantification pour extraire la carte topographique est 1.

6.4.1 Estimation de la courbure

Le lissage affine que nous utilisons utilise l'équation aux dérivées partielles présentée à la Section 6.2.1. L'implémentation que nous utilisons est celle de [Moisan, 1998] disponible dans la librairie Megawave [Froment et al., 2007]. Elle conserve toutes les propriétés théoriques de ce lissage (monotonie, invariance affine et morphologique). Le temps d'évolution final T de l'équation est normalisé de telle manière qu'un cercle de rayon T disparaît complètement à la fin du lissage. Dans toutes nos expériences, nous utilisons $T = 0.5$ pixel. Sachant que la résolution réelle des dessins que nous utiliserons dans la base présentée à la Section 6.4.5 est de l'ordre de $4\text{pixels}/\text{mm}^2$, nous voyons que ce lissage fait disparaître les détails de l'ordre de 0.25mm^2 .

L'implémentation de Moisan permet également d'obtenir un échantillonnage des courbes relativement constant le long des lignes de niveaux. En effet, l'implémentation utilisée nous garantit en pratique que le pas d'échantillonnage appartient à l'intervalle $[\Delta s, 2\sqrt{2}\Delta s]$. Dans toutes les expériences nous fixons ce pas à $\Delta s = 0.5\text{pixels}$.

Pour illustrer la précision de la mesure de courbure, la méthodologie a été testée sur une image de spirale logarithmique (Figure 6.13). Cette spirale a été tracée sous forme paramétrique dans Matlab puis pixelisée. La courbure sur ce type de spirale est inversement proportionnelle à l'abscisse curviligne. Les mesures produites sont comparées avec les valeurs exactes sur la Figure 6.14. Nous voyons que les valeurs suivent assez bien la courbe théorique.

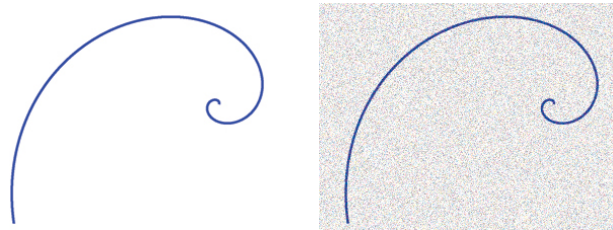


FIG. 6.13: Spirales logarithmiques. A gauche une image d'une spirale non bruitée, et à droite la même image avec un bruit additif Gaussien ($\sigma = 5$). La courbure théorique le long de telles spirales est inversement proportionnelle à l'abscisse curviligne.

6.4.2 Estimation du rayon de l'outil

L'estimation du rayon de l'outil utilisé pour chaque agrégat est une information importante car elle permet d'ajuster les paramètres des méthodes subséquentes d'analyse du contenu pictural. Pour évaluer la robustesse de cette mesure, nous avons utilisé deux dessins synthétiques avec un outil dont le rayon est contrôlable. Les dessins ont été réalisés

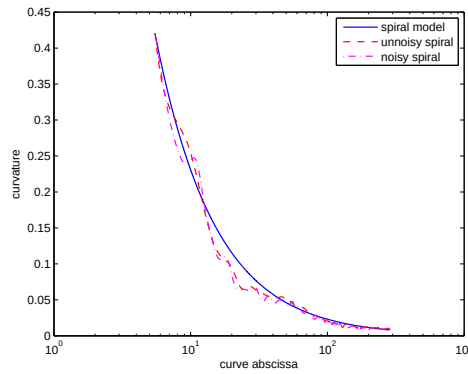


FIG. 6.14: Courbure le long d'une spirale logarithmique. La courbure a été calculée sur les images de la Figure 6.13 et comparée aux valeurs théoriques (bleu).

avec le logiciel Gimp. Le premier est visible à gauche de la Figure 6.15 et présente des agrégats de traits de différents rayons : 3, 7 et 15 pixels. L'estimation de ces rayons est faite avec l'Equation 6.2. L'erreur moyenne sur cet exemple est de 12.5%. Sur l'exemple à droite de la Figure 6.15 qui présente un très grand agrégat, la longueur des traits tend à diminuer l'erreur d'estimation. L'erreur moyenne sur cet exemple est de 1.8%.

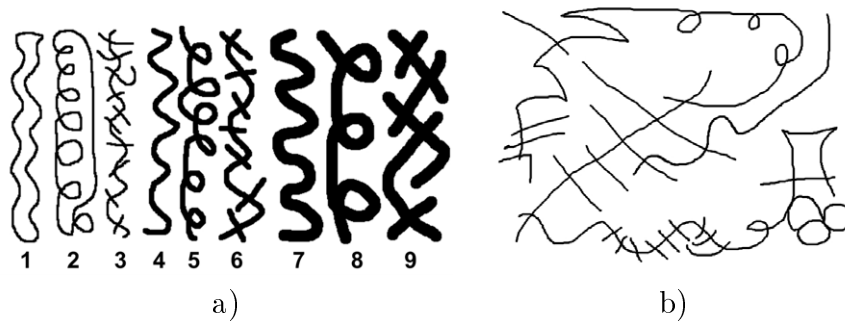


FIG. 6.15: Dessins synthétiques utilisés pour l'évaluation de la méthode de mesure du rayon de l'outil R_Φ . A gauche, 9 agrégats, numérotés de 1 à 9 de gauche à droite dans le Tableau 6.2.

TAB. 6.2: Estimation du rayon de l'outil sur la Figure 6.15a.

	1	2	3	4	5	6	7	8	9	Moyenne
Erreur	5%	6%	2%	14%	27%	26%	5%	15%	13%	12.5%

6.4.3 Paramètre k_c

Les paramètres k_c , k_e et k_j ont été fixés empiriquement par de très nombreuses expériences sur les dessins, et également sur des figures synthétiques. Nous présentons dans cette section et la suivante une illustration sur des figures synthétiques des compromis que ces paramètres amènent à réaliser.

Le paramètre k_c a été introduit à la Section 6.2.4 pour fixer le seuil $\kappa_t = 1/(k_c R_\Phi)$ sur la courbure permettant d'isoler un ensemble de candidats à la taxonomie des extrema. Ce seuil sépare en deux la distribution des courbures utilisées pour calculer les trois premiers descripteurs. Il doit être fixé de manière à ce que ces descripteurs soient indépendants du nombre d'extrémités de traits dans l'image.

Nous fixons empiriquement k_c en utilisant deux groupes de 5 images synthétiques. Une image de chaque groupe pour $R_\Phi = 5$ est présentée sur la Figure 6.16. A gauche de cette figure, un dessin a été tracé avec un seuil trait, en utilisant un outil de dessin vectoriel. Le même dessin a été ensuite découpé en plusieurs traits. Chacun de ces dessins a été reproduits avec cinq rayons d'outils R_Φ différents : 3, 5, 9, 13 et 17 pixels. Ces 10 images ont été ensuite pixelisées.

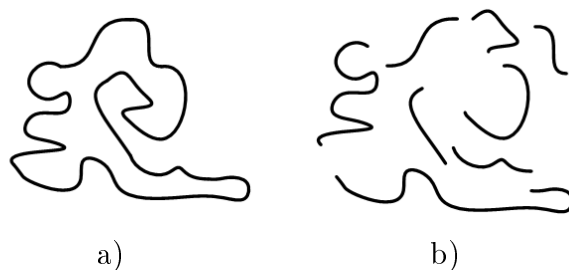


FIG. 6.16: Images synthétiques utilisées pour fixer le seuil sur la courbure κ_t avec $R_\Phi = 5$. A gauche, un dessin constitué d'un seul trait fermé. A droite, le même dessin découpé en plusieurs morceaux. Ces dessins ont été réalisés avec un outil de dessin vectoriel et ainsi ont pu être répétés à l'identique avec plusieurs épaisseurs de traits. Toutes ces images ont ensuite été pixelisées à la même résolution.

Idéalement, le seuil κ_t est fixé de manière à ce que les trois premiers descripteurs de ces dessins soient égaux. Ceci n'est pas vrai lorsque k_c est trop faible puisque les extrémités et leurs voisinages (de fortes courbures) sont intégrées dans le calcul des descripteurs. Inversement, si nous augmentons trop k_c , nous risquons de rejeter un grand nombre de points qui sont le long d'un trait. Ces deux effets sont illustrés sur la Figure 6.17. Nous avons empiriquement trouvé un compromis sur ces dessins avec $k_c = 5$. Dans le reste des expériences nous utiliserons donc $\kappa_t = 1/(5R_\Phi)$.

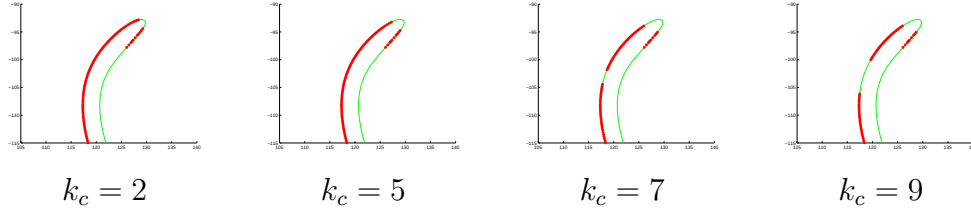


FIG. 6.17: Influence de k_c . Détail de l'extrémité visible au centre de la Figure 6.16b. Si k_c est trop faible, les extrémités et leurs voisinages (de fortes courbures) sont intégrés dans le calcul des descripteurs. Inversement, si nous augmentons trop k_c , nous risquons de rejeter un grand nombre de points qui sont le long d'un trait. Rappelons ici que nous nous intéressons uniquement aux valeurs $\kappa \leq 0$.

6.4.4 Paramètres k_e et k_j

Les extrémités sont détectées en analysant les portions de contours de longueur $k_e R_\Phi$ autour d'un extremum négatif (Section 6.2.8). Augmenter k_e accroît le risque de manquer les extrémités qui sont à une distance inférieure à $k_e R_\Phi$ d'un croisement de traits. Inversement, abaisser k_e accroît le risque de détecter certaines jonctions *étirées* de deux traits comme étant une extrémité. Une illustration de ces deux effets opposés en faisant varier k_e est présentée sur la Figure 6.19. Nous verrons ce risque sur des dessins réels à la Section 6.4.6. L'image synthétique présentée sur la Figure 6.18a nous a permis de trouver un compromis entre ces deux risques en fixant $k_e = 5$.

Les croisements de traits sont détectés en analysant la région couverte par un cercle polygonal de rayon $k_j R_\Phi$. Augmenter k_j accroît le risque de considérer des régions trop grandes, suffisamment complexe pour provoquer une fausse détection de croisement. Abaisser k_j accroît le risque de manquer des croisements. Une illustration de ces deux effets opposés en faisant varier k_j est présentée sur la Figure 6.20. La valeur théorique de k_j devrait s'adapter à chaque croisement, et à l'angle relatif local formé par les traits en présence. Pour plus de simplicité, nous avons utilisé le dessin de la Figure 6.18b pour fixer empiriquement ce seuil à $k_j = 5$.

Dans le reste des expériences nous utiliserons donc les valeurs $k_e = k_j = 5$.

6.4.5 Base de donnée classée selon l'impact pictural

Dans les expériences suivantes, nous utiliserons une base de 105 œuvres à base de traits. Nous avons classé manuellement, et avec notre propre subjectivité, cette base en 14 classes d'images en fonction de l'impact pictural. La base au complet est visible en Annexe A. Un aperçu de la base est visible sur la Figure 6.21. La moitié de cette base (classes 1 à 7) provient d'un catalogue d'une exposition réalisée au Centre Pompidou en

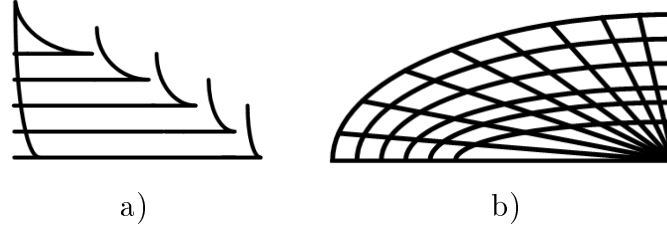


FIG. 6.18: Images synthétiques utilisées pour fixer empiriquement k_e (à gauche) et k_j (à droite).

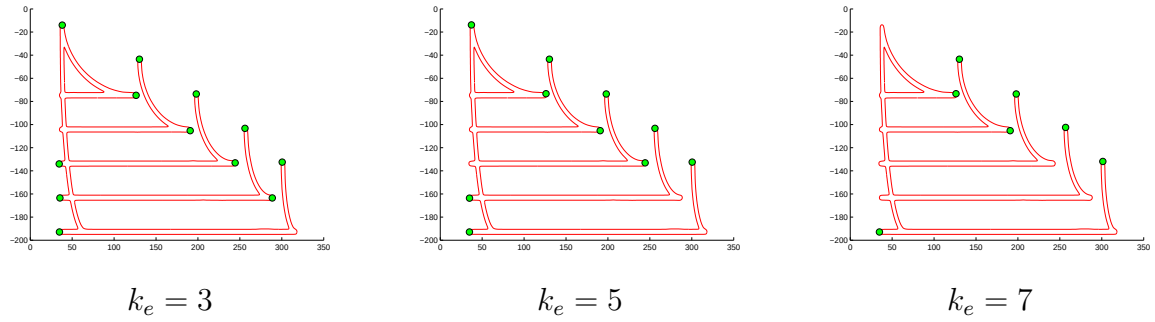


FIG. 6.19: **Influence de k_e .** En vert, les extrémités détectées. Augmenter k_e accroît le risque de manquer les extrémités qui sont à une distance inférieure à $k_e R_\Phi$ d'un croisement de traits. Inversement, abaisser k_e accroît le risque de détecter certaines jonctions *étirées* de deux traits comme étant une extrémité.

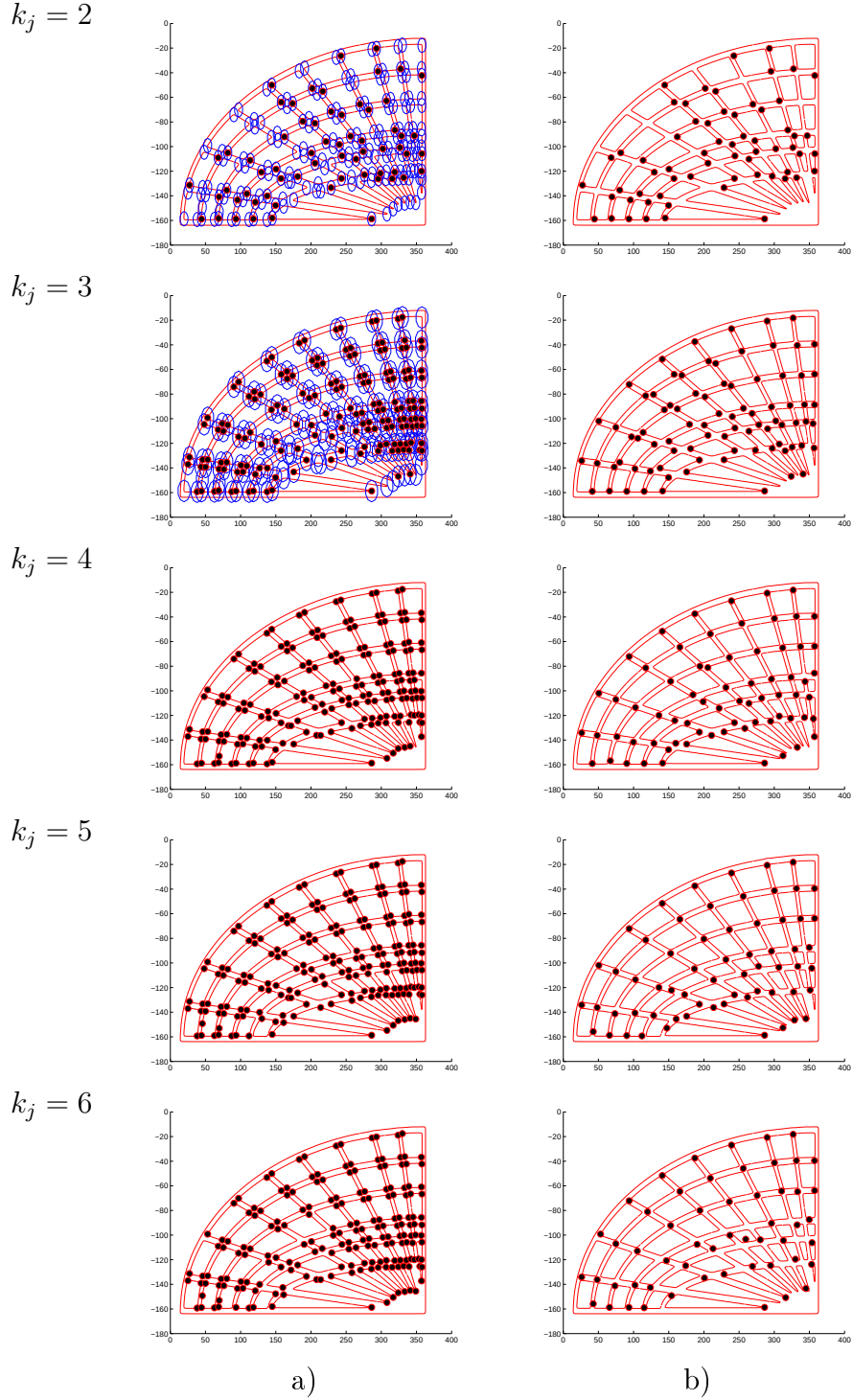


FIG. 6.20: Influence de k_j . a) En noir, les éléments de croisement détectés, avant fusion des éléments distants de moins que $k_j R_\Phi$. Les deux premières lignes montrent aussi en bleu les cercles de rayons $k_j R_\Phi$. Pour plus de lisibilité, ces cercles ne sont pas présentés pour $k_j \geq 3$. b) En noir, les croisements détectés finaux.

2002 regroupant des dessins faits par Henri Matisse et Ellsworth Kelly. Ces dessins portent essentiellement sur des plantes et feuilles de végétaux [Remi Labrusse, 2002]. Ces images ont été numérisées par un scanner HP Scanjet 8200 avec une résolution de 4800 ppi. Trois classes (8 à 11) proviennent de fac-similé numérique de carnets de croquis de Picasso, édité par la RMN en 2005 [RMN, 2006]. Enfin, trois classes (12 à 14) proviennent de dessins faits par ordinateur, extrait d’une base commerciale d’illustration appelée ArtExplosion, contenant 750000 cliparts [Explosion, 2002].

L’ordre de grandeur des temps de calculs pour extraire les descripteurs sur cette base est d’environ 15 secondes par dessin en utilisant un PC Pentium IV 4.3GHz. La majeure partie de ce temps est consacrée à l’extraction des lignes de niveaux significatives maximales, dont la méthodologie a été présentée au Chapitre 5. Pour effectuer une requête sur la base, le temps de calcul est quasiment instantané sachant que nous utilisons une distance Euclidienne sur un vecteur de 11 descripteurs.

6.4.6 Détection des extrémités, croisements et coins

Un exemple de dessin dont les éléments picturaux tels que définis à la Section 6.2 ont été extraits est visible sur la Figure 6.22. Sur cette figure, les points verts représentent les extrémités de traits, les points noirs sont les croisements de traits, et les étoiles rouges (resp. vertes) sont les coins positifs (resp. négatifs).

Le Tableau 6.3 résume les résultats de classification pour les extrémités et croisements de traits, calculés sur un ensemble de 200 occurrences de chaque élément manuellement extraits sur une partie de la base classée. Le taux de rappel et le taux de précision dans ce tableau sont ceux classiques utilisés par exemple par [Del Bimbo, 1999]. Le rappel est le ratio entre le nombre d’éléments vrais positifs et le nombre total d’éléments classés comme positifs. Le taux de précision est le ratio entre le nombre d’éléments vrais positifs et le nombre total réel d’éléments positifs.

TAB. 6.3: Taux de classification des extrémités et croisements sur une base de 200 occurrences.

	Extrémités	Croisements
Taux de Rappel	95.6%	92.5%
Taux de Précision	78.4%	95.1%

Ainsi, si une extrémité de trait est présente en un point, il y a 95.6% de chance de la détecter. Parmi tous les éléments qui ont été classés comme étant une extrémité de traits, 75.4% sont vraiment des extrémités de traits. Ce ratio assez faible est dû à deux type d’erreurs : les erreurs de détection des contours et les extrémités *ambigües*. Les erreurs de détection des contours ont été présentées à la Section 5.4.3 dans le précédent chapitre.

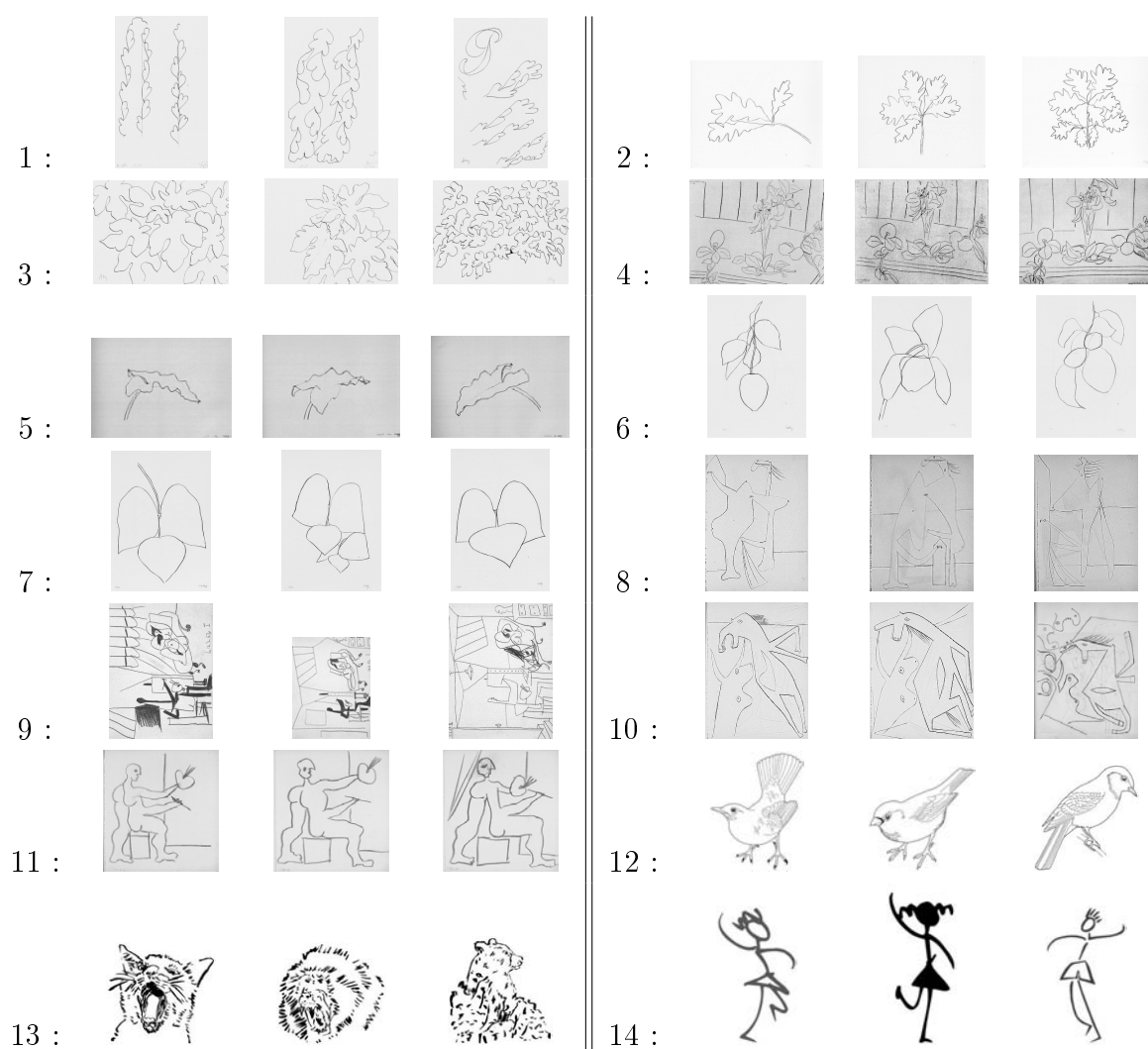


FIG. 6.21: Aperçu de la base d'œuvres classées. Nous utiliserons cette base aux Section 6.4.6 et 6.4.7. La base au complet est visible à l'Annexe A.

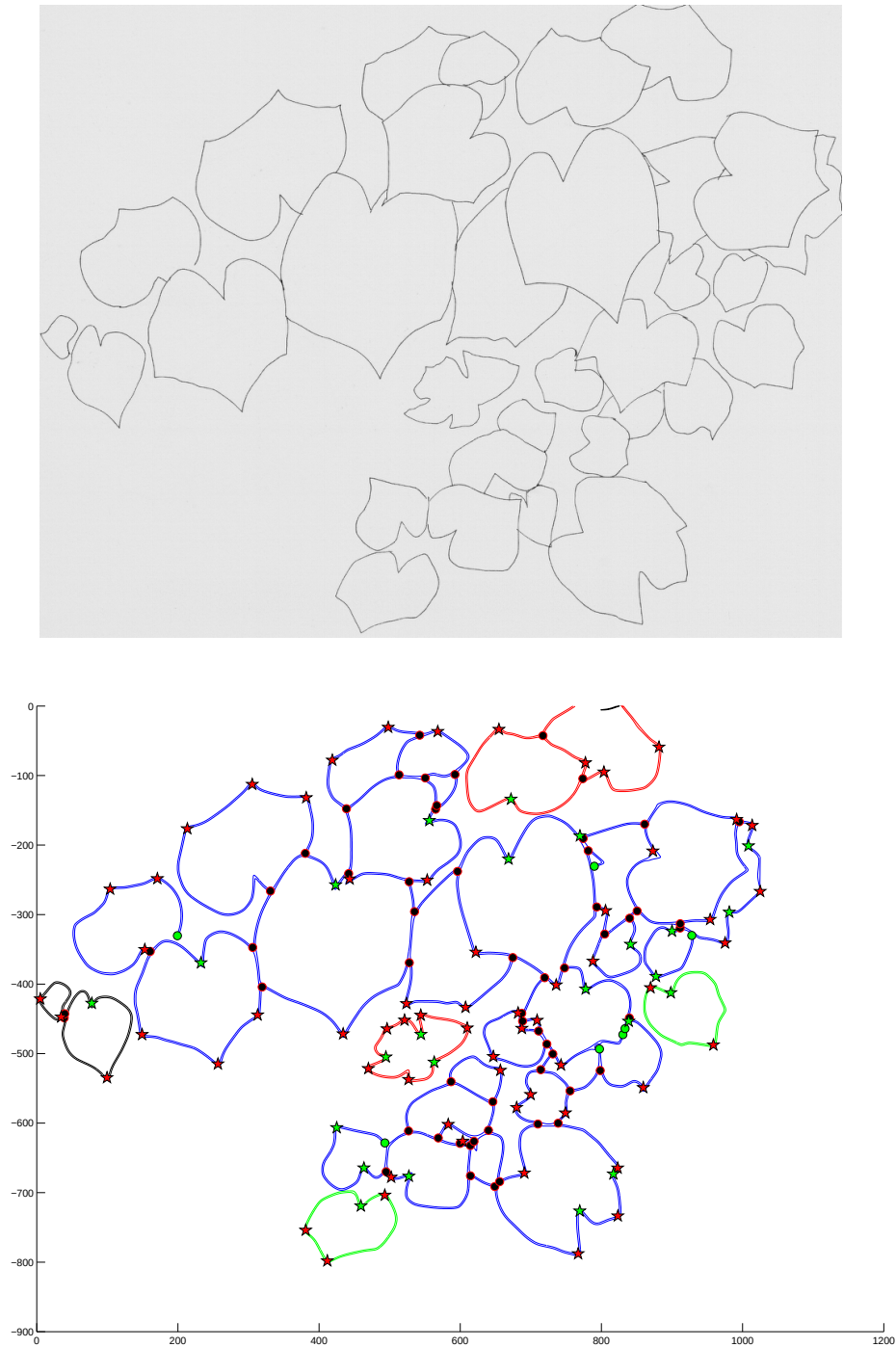


FIG. 6.22: Un exemple de dessin analysé. Les points verts sont les extrémités de traits, les points noirs sont les croisements de traits, et les étoiles rouges (resp. vertes) sont les coins positifs (resp. négatifs). Chaque agrégat est représenté par une couleur de trait différente.

Une telle erreur est visible sur le deuxième exemple de la Figure 6.22 le long de la tige de la plante. Ce que nous appelons une extrémité ambiguë se rencontre dans une situation où deux traits se rejoignent et s'étirent sur une longueur supérieure au seuil $k_e R_\Phi$ (Section 6.2.8). Dans ces situations (que nous avons vu sur la Figure 6.19), est-ce une extrémité de trait ? Sommes nous certain que l'artiste a relevé son outil ? Il est parfois difficile de conclure de manière péremptoire même avec un examen attentif du dessin. Toutes ces situations d'ambiguïtés, si elles ont été détectées comme étant le siège d'une extrémité ont été comptabilisées comme étant des faux positifs dans notre protocole expérimental. Sans prendre en compte ces ambiguïtés, le taux de rappel est de 85.1%.

Les taux de rappel et de précision des croisements de traits sont au dessus de 90%. Les 7.5% de croisements qui sont manqués par notre méthode sont essentiellement dus à une fusion abusive des extrema classés comme appartenant à un croisement de trait dans une région où plusieurs croisements sont très proches les uns des autres (Section 6.2.6 et Figure 6.20). La contrepartie positive de cette perte est une assez bonne modélisation unitaire des croisements puisqu'un croisement réel est en moyenne représenté par 1.03 croisements détectés.

Notre définition des coins étant empirique, il serait difficile et biaisé de classer un ensemble de coins pour évaluer la capacité à les détecter. Nous pouvons par contre évaluer l'adéquation entre notre définition du type des coins (positif ou négatif selon leur orientation par rapport au barycentre de l'objet) et la nature réelle des coins détectés lorsqu'ils décrivent un objet (convexe ou concave). Dans le cas précis où le coin appartient au contour extérieur d'un objet identifiable sans ambiguïté, il y a en moyenne 94.3% de chance que le signe du coin soit correct. Dans l'ensemble des tests effectués pour cette expérience, 10% des coins détectés n'appartenait pas à des contours d'objets identifiables. Cette valeur faible est due au contenu de la base utilisée qui est très figurative.

6.4.7 Résultats d'indexation comparés

Méthodes de comparaisons

Dans cette section, nous comparons notre méthode sur la base classifiée avec une approche utilisant l'espace multi-échelle de courbure (CSS) proposée par [Mokhtarian and Mackworth, 1992] et faisant parti de la norme MPEG-7. Cet approche classique a inspiré de nombreuses recherche sur la détection des points d'intérêts (Section 6.1.4). Nous utilisons dans cette section une méthode récente reposant sur le CSS pour détecter les points d'intérêts (extrémités, coins, jonctions...) dans une image proposée par [He and Yung, 2004].

Cette méthode extrait tout d'abord une carte binaire des contours à partir de l'image en niveaux de gris en utilisant un détecteur de type Canny. Les pixels de contours voi-

sins sont reliés pour former un ensemble de courbes dans l'image. Un espace CSS à deux échelles est utilisé pour fixer un seuil adaptatif κ_a sur la courbure permettant de sélectionner un ensemble de points d'intérêts parmi les extrema de courbure. Cette méthode sera notée par la suite *CSS-Canny*. La méthode obtenue en remplaçant la carte des contours de Canny par les lignes de niveaux extraites avec la méthode proposée au chapitre précédent sera appelée *CSS-LL*. En utilisant chacune de ces méthodes, nous calculons pour chaque dessin un vecteur de 4 descripteurs. Les trois premières dimensions sont composées des mêmes descripteurs que dans le Tableau 6.1 mais calculés sur les distributions de courbure extraites à partir du CSS à l'échelle pixelique. Le seuil κ_t sur la distribution de courbure, utilisé pour calculer ces descripteurs est le seuil κ_a fourni par la méthode proposée par [He and Yung, 2004]. Le quatrième descripteur est la densité linéaire totale des points d'intérêts trouvés avec chacune de ces méthodes. Ces descripteurs sont normalisés de la même manière qu'à la Section 6.3.

Les points d'intérêts détectés par cette méthode regroupent en une seule catégorie les extrémités de traits, croisements et coins. Une première comparaison équitable et directe de ces méthodes avec l'approche que nous proposons est d'utiliser aussi un vecteur de descripteurs à quatre dimensions, composés des descripteurs 1 à 3 (Tableau 6.1) et d'un quatrième descripteur qui est la moyenne des descripteurs 5 à 8. Ce quatrième descripteur correspond à la densité linéaire des points caractéristiques. Nous appelons *LL-moy* cette méthode utilisant les lignes de niveaux, et les méthodes proposées dans ce chapitre.

Une comparaison suivante possible est de ne pas faire la moyenne des descripteurs 5 à 8 de manière à illustrer l'apport représenté par une distinction plus fine des éléments picturaux dans le cadre de notre problématique. Cette troisième méthode que nous appellerons *LL-diff*, est équivalente à celle utilisant la distance décrite à la Section 6.3 avec un vecteur de pondération $W_{11D} = [1, 1, 1, 0, 1, 1, 1, 1, 0, 0, 0]$. Pour cette raison, nous utiliserons pour les méthodes LL-moy, CSS-Canny et CSS-LL un vecteur de pondération $W_{4D} = [1, 1, 1, 4]$.

La dernière étape de comparaison est d'utiliser tous les descripteurs que nous proposons dans le Tableau 6.1. Nous appellerons *LL-all* cette quatrième méthode de comparaison. Elle est obtenue en utilisant un vecteur de pondération $W_{11D} = [1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1]$.

Indices de comparaisons

La distance Euclidienne pondérée nous permet d'ordonner la base de donnée en une liste allant de l'image la plus proche (la requête elle-même lorsque nous utilisons une image de la base) à la plus éloignée selon cette mesure. En suivant l'ordre des images résultats fournie par cette distance, nous noterons N_x le nombre d'images résultats nécessaires pour retrouver x images correctement classées, c'est à dire dont la classe est la même que celle de la requête. La Figure 6.23 montre les courbes du taux de précision x/N_x par rapport au rappel défini par x . Cette définition du rappel est différente de celle utilisée à la Section

6.4.6. Elle est choisie ici car elle permet de réaliser la moyenne de plusieurs courbes même lorsque les classes de la base sont de tailles différentes.

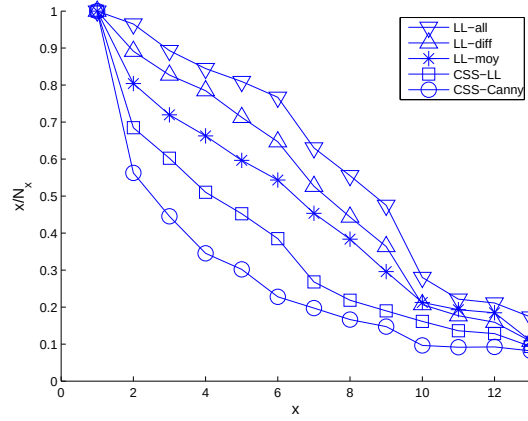


FIG. 6.23: Courbe précision-rappel moyenne. En suivant l'ordre des images résultats fournie par la distance Euclidienne pondérée, nous notons N_x le nombre d'images résultats nécessaires pour retrouver x images correctement classées, c'est à dire dont la classe est la même que celle de la requête.

Pour évaluer différemment les performances des méthodes nous utilisons l'indice appelé *décompte de gain cumulé* (ou *discounted cumulative gain*, DCG) [Shilane et al., 2004]. Cette mesure cumule les contributions g_r associés à chaque image de rang r dans la liste des résultats ordonnée selon la distance utilisée. La contribution g_r de la r^{eme} image vaut $1/\log_2(r)$ si elle appartient à la même classe que la requête et 0 sinon. Ce cumul de gain est normalisé par le cumul maximal possible (si toutes les images de la classe arrivent dans les premières) Pour une requête appartenant à une classe de taille C dans une base de taille N , le DCG est alors donné par :

$$DCG = \frac{1 + \sum_{r=2}^N g_r}{1 + \sum_{j=2}^C 1/\log_2(j)}.$$

Enfin, nous utilisons également le score statistique appelé score *Bull-Eye* (ou Bull-Eye Percentage, BEP) [Shilane et al., 2004], qui correspond à la moyenne sur l'ensemble des classes des moyennes des taux de précisions mesurés sur les $2C$ premières images, où C est la taille de la classe à laquelle appartient la requête. Les scores DCG et BEP sont résumés dans le Tableaux 6.4 et le Tableau 6.5.

Les résultats de la comparaison entre CSS-Canny et CSS-LL soulignent l'apport bénéfique dans notre problématique des lignes de niveaux comme support d'étude plutôt qu'une carte des contours extraites à partir du détecteur de Canny-Deriche. La comparaison entre CSS-LL et LL-moy démontre aussi que l'analyse des points d'intérêts picturaux

TAB. 6.4: Scores DCG par classe pour toutes les méthodes utilisées.

	LL-all	LL-diff	LL-moy	CSS-LL	CSS-Canny
Moyenne	93.7%	88.2%	82.5%	74.4%	65.3%
Classe 1	99.8%	96.2%	90.1%	100%	78.7%
Classe 2	98.9%	92.7%	79.7%	87.5%	76.1%
Classe 3	96%	92.2%	94.5%	64.7%	56%
Classe 4	98.2%	99.1%	94.5%	93.3%	57.4%
Classe 5	93.3%	83.6%	73.4%	87.5%	64%
Classe 6	88.3%	88.7%	85.4%	89.8%	75.6%
Classe 7	92.6%	86.5%	83.1%	61.1%	54.1%
Classe 8	93.8%	87.8%	76%	66.2%	57.6%
Classe 9	83.9%	80.2%	66.4%	63.3%	56.7%
Classe 10	76.2%	67.11%	63.5%	67.3%	56.1%
Classe 11	75.3%	66.1%	57.8%	66.3%	38%
Classe 12	96.8%	91.5%	78.67%	59.4%	69.1%
Classe 13	100%	88.6%	91%	61.4%	76.7%
Classe 14	98.4%	89.7%	91.1%	66.1%	64.9%

TAB. 6.5: Scores Bull-Eye par classe pour toutes les méthodes utilisées.

	LL-all	LL-diff	LL-moy	CSS-LL	CSS-Canny
Moyenne	88.2%	83.6%	72.8%	62.8%	50.1%
Classe 1	100%	94.4%	91.7%	100%	61.1%
Classe 2	97.9%	85.7%	59.2%	85.7%	79.6%
Classe 3	93.7%	92.2%	93.7%	53.1%	34.4%
Classe 4	100%	100%	100%	88.9%	44.4%
Classe 5	98%	77%	63%	68%	44%
Classe 6	91.7%	82.7%	82.7%	87.5%	50.1%
Classe 7	87.5%	87.5%	81.2%	37.5%	37.5%
Classe 8	91.3%	88.9%	71.6%	37.4%	29.6%
Classe 9	72%	76%	48%	52%	52%
Classe 10	50%	56.2%	50%	56.2%	43.7%
Classe 11	66.6%	55.5%	33.33%	66.6%	33.3%
Classe 12	97.5%	92.6%	67.9%	43.2%	69.1%
Classe 13	100%	97.5%	91.3%	50.6%	86.4%
Classe 14	95.3%	84.6%	85.8%	53.2%	34.9%

6.4. Expériences

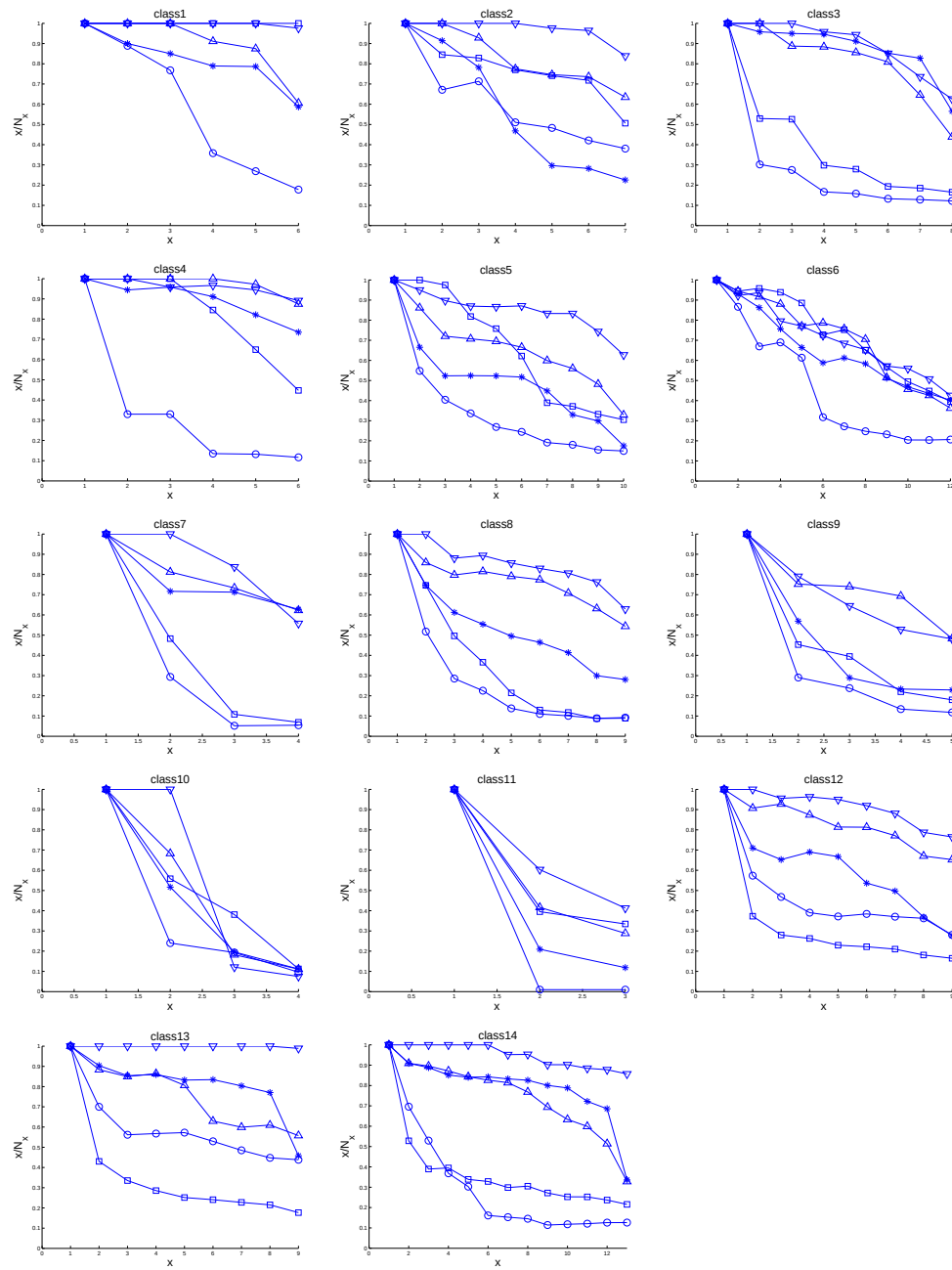


FIG. 6.24: Courbes précisions rappel pour chaque classe. Même légende que sur la Figure 6.23.

par notre approche, indépendamment de leur type, est d'une plus grande précision par rapport à celle utilisant les CSS. Les résultats de la méthode LL-diff soulignent le gain apporté en caractérisant chaque point d'intérêt. Enfin, les descripteurs supplémentaires proposées dans la méthode LL-all pour mieux caractériser ces éléments apportent aussi un gain de qualité des résultats.

6.4.8 Exemples de requête sur la base classée

Un grand nombre d'exemples de requêtes sur la base classée de 105 dessins sont visibles sur les Figure 6.25 et 6.26. Ces résultats permettent d'apprécier la performance de la méthode proposée (correspondant à la méthode *LL - all*). Quatre exemples d'erreurs de *classification* sont visibles sur les Figures 6.27 et 6.28. Sur la Figure 6.27 à gauche, les deux premiers résultats (au centre en haut et à droite en haut) n'appartiennent pas à la même classe que la requête (en haut à gauche). Ces erreurs ont néanmoins un impact pictural relativement proche de la requête. De même, sur les Figure 6.27 à droite et 6.28 à gauche, nous pouvons intuitivement deviner que les descripteurs obtenus par la densité de croisements, de coins et d'extrémités doivent en effet être assez proches entre ces erreurs et la requête. L'association de ces erreurs avec la requête est néanmoins moins satisfaisante. Sur la Figure 6.28 à droite l'écart sémantique est tellement grand entre les représentations d'oiseaux et les fausses images similaires qu'il est difficile de ne pas vouloir dissocier totalement ces images.

6.4.9 Exemples de requête sur une base non classée

Dans cette section, nous avons enrichi la base précédente avec 154 œuvres supplémentaires non classées. Ces exemples permettent d'illustrer non pas la robustesse de la méthode vis à vis des classes (organisée selon un critère subjectif), mais de visualiser des résultats d'indexation généraux sur une base deux fois plus riche. Ces résultats permettent également d'apprécier la capacité de ce critère à associer des œuvres auxquelles nous n'aurions peut-être pas pensé (sérendipité). Les Figures 6.29 et 6.30 montrent plusieurs résultats sur cette base.

6.4.10 Discrimination de textons

[Julesz, 1986] proposa une théorie de la perception des textures réalisées à partir de traits rectilignes. Nous présentons dans cette section quatre résultats illustrant le comportement sur ces textures de la méthode proposée dans ce chapitre. Sur la Figure 6.31 en haut, les textons ont les mêmes distributions de courbure et des densités linéaires de croisements, extrémités et coins différentes. Notre vision préattentive comme la méthode

6.4. Expériences

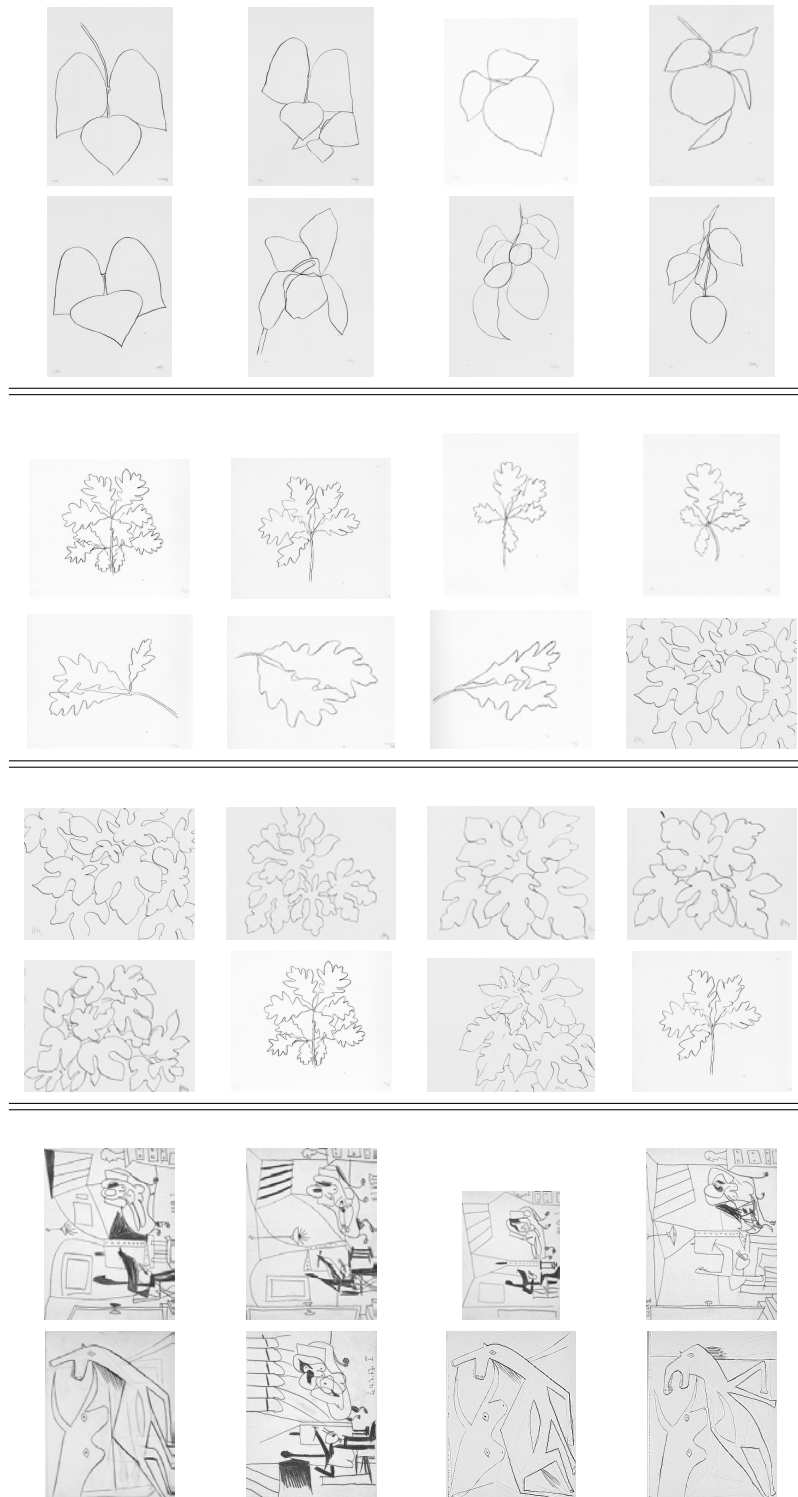


FIG. 6.25: Exemple de requête sur la base classée de 105 œuvres

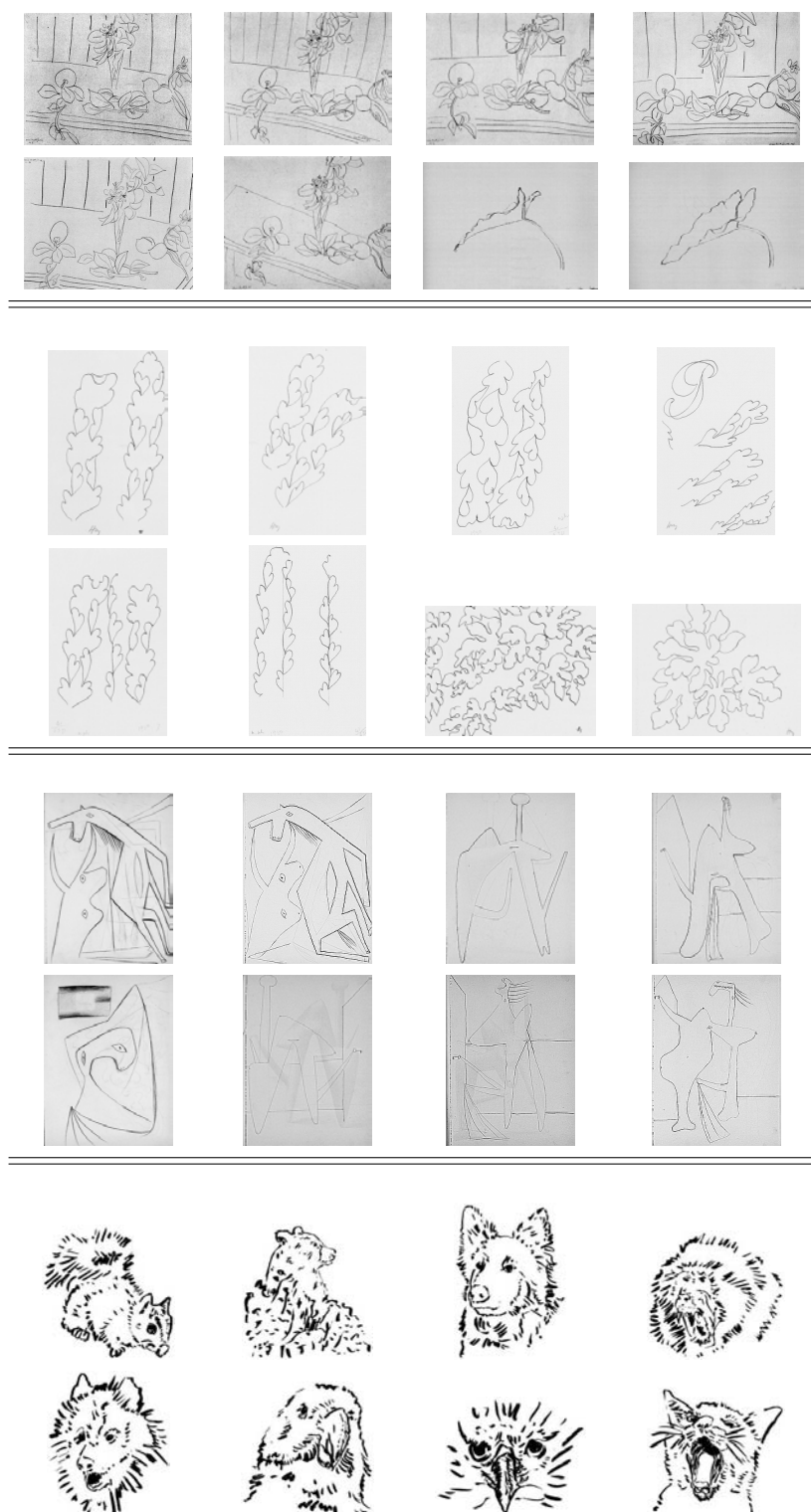


FIG. 6.26: Exemple de requête sur la base classée de 105 œuvres



FIG. 6.27: Exemple d'erreur. A gauche, les deux premiers résultats (au centre en haut et à droite en haut) n'appartiennent pas à la classe 7 comme la requête (en haut à gauche). Ces erreurs ont néanmoins un impact pictural relativement proche de la requête. A droite, les deux derniers résultats n'appartiennent pas à la classe 3 comme la requête, bien que toutes les images de la classe ne soient pas encore trouvées. Nous devinons intuitivement que les descripteurs comme la densité de croisements, de coins et d'extrémités entre ces erreurs et la requête doivent probablement être assez proches.

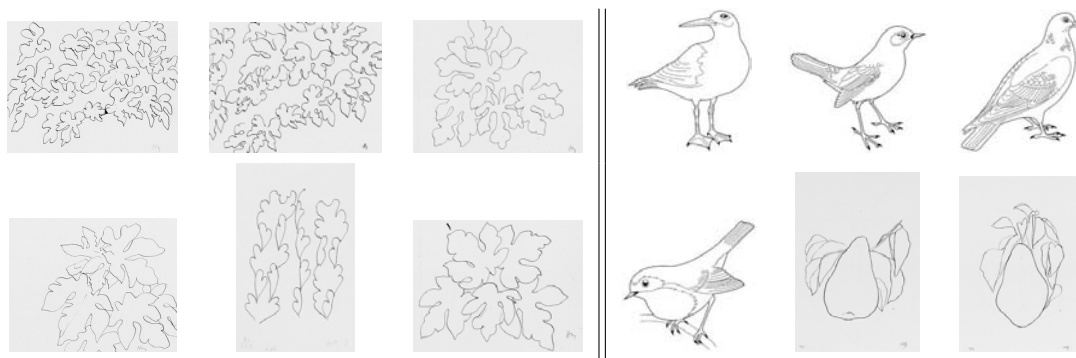


FIG. 6.28: Exemple d'erreur. A gauche, l'avant dernier résultat n'appartient pas à la classe 3 comme la requête. Là aussi nous devinons intuitivement qu'une grande partie des descripteurs sont probablement assez proches. A droite, les deux derniers résultats ne font pas partie de la classe 12 comme la requête. Même si nous pouvons comprendre que les descripteurs puissent être assez proches (les densités de croisement et coins semblent similaires), l'écart sémantique est tellement fort qu'il est difficile de ne pas vouloir dissocier totalement ces images.



FIG. 6.29: Exemple de requête sur la base non classée de 259 œuvres

6.4. Expériences

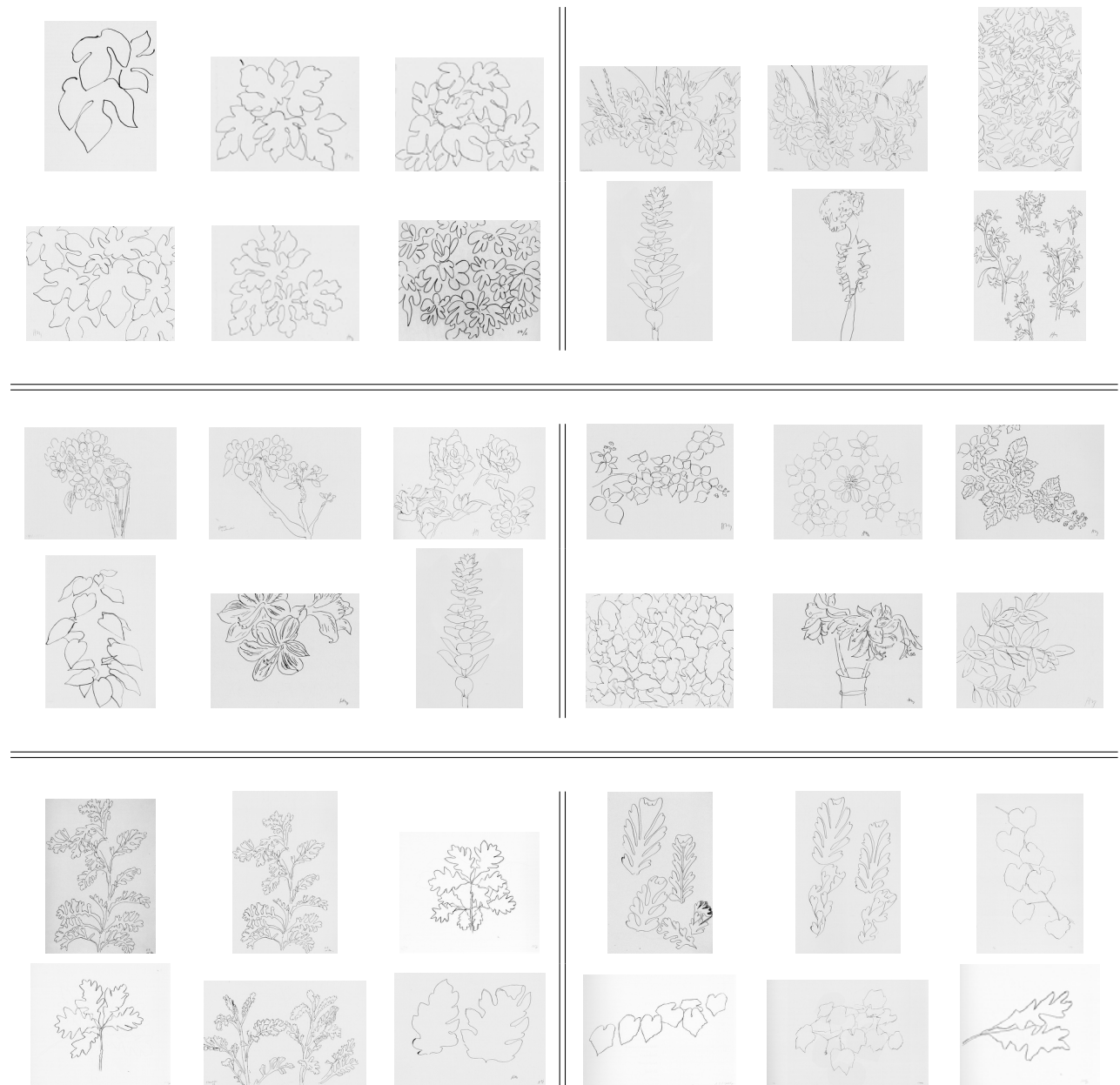


FIG. 6.30: Exemple de requete sur la base non classée de 259 œuvres

proposée distinguent les textons facilement. En bas, les différents textons ont les mêmes distributions de courbure et les mêmes densités linéaires de coins et extrémités. Notre vision préattentive comme la méthode proposée n'arrivent pas à les distinguer. Sur la Figure 6.32 en haut, pentagones et hexagones sont parallèles entre eux respectivement. Notre vision préattentive les distinguent selon Julesz au contraire d'en bas où ils sont orientés aléatoirement. Notre méthode dans les deux cas les distinguerait légèrement à cause d'un nombre de coins différents. Cette limite pourrait être résolue en intégrant des informations liées à l'orientation locale absolue le long des contours. Dans nos expériences et notre problématique de recherche d'images (qui est différente de la discrimination de textures), nous avons préféré conserver une certaine invariance aux rotations.

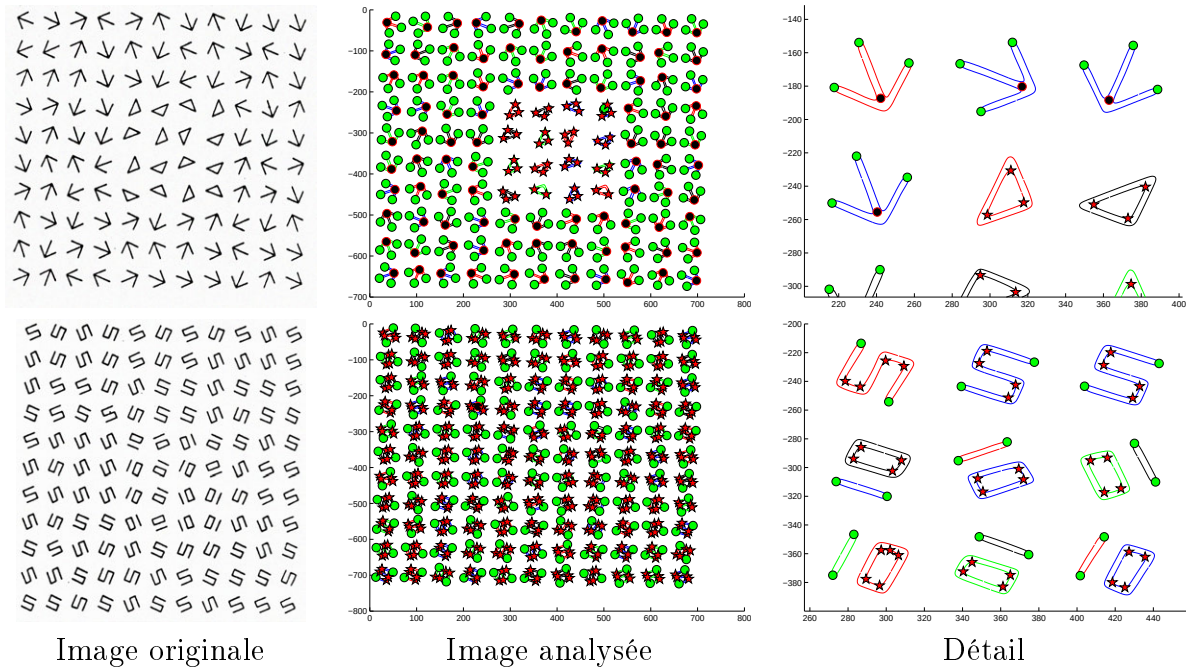


FIG. 6.31: Discrimination de textons. En haut, les textons ont les mêmes distributions de courbures et des densités linéaires de croisements, extrémités et coins différents. Notre vision préattentive comme la méthode proposée distinguent les textons facilement. En bas, les différents textons ont les mêmes distributions de courbures et les mêmes densités linéaires de coins et extrémités. Notre vision préattentive comme la méthode proposée n'arrivent pas à les distinguer.

6.5 Conclusion

Nous avons proposé dans ce chapitre un ensemble de méthodes permettant, dans une oeuvre au trait, d'extraire la courbure des traits et plusieurs éléments picturaux importants tels que les extrémités de traits, les croisements et les coins. A l'aide de ces méthodes, nous avons aussi proposé un ensemble de descripteurs statistiques globaux

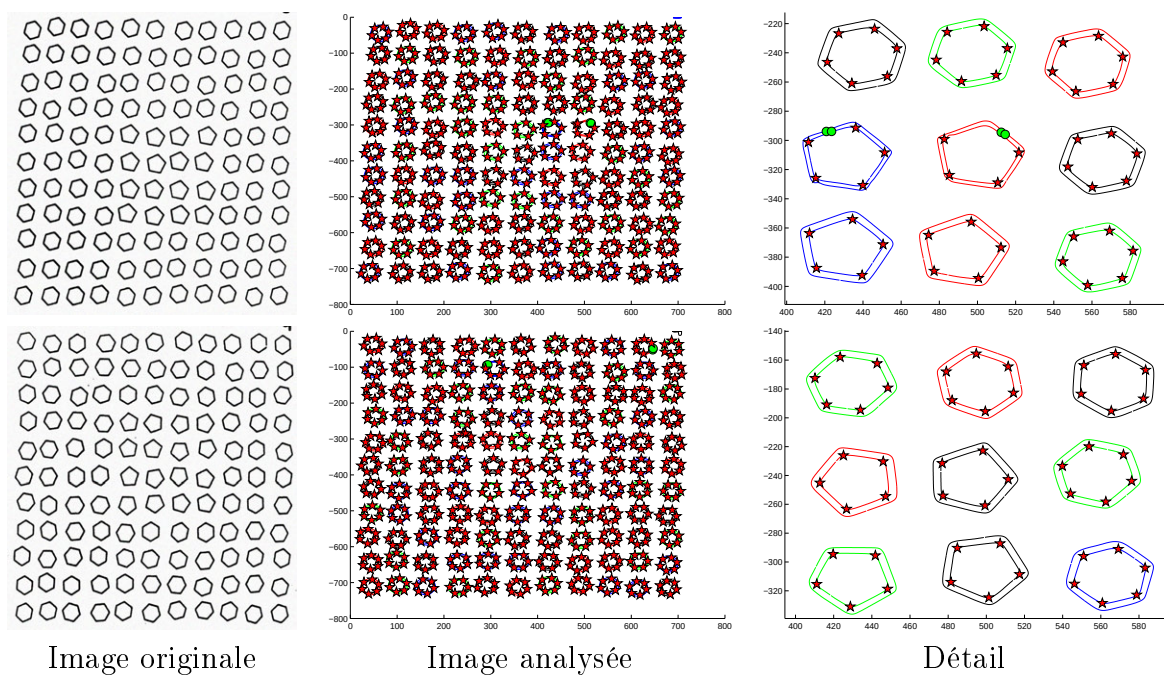


FIG. 6.32: Discrimination de textons. En haut, pentagones et hexagones sont parallèles entre eux respectivement. Notre vision préattentive les distinguent selon Julesz au contraire d'en bas où ils sont orientés aléatoirement. Notre méthode dans les deux cas les distingueraient à cause d'un nombre de coins différents.

permettant une comparaison directe entre deux œuvres. Cette description a été comparée à une approche différente, ainsi qu'à des versions simplifiées de notre méthode. Ceci a permis d'illustrer l'apport et la qualité des différents descripteurs sur une base classée.

L'approche originale proposée permet d'aborder une problématique rarement abordée dans la littérature. Elle rend compte pourtant d'un type de primitive particulier et abondamment exploité par l'artiste.

Le critère que nous cherchons à étudier est plus subjectif qu'un critère de reconnaissance sémantique du contenu. Nous pensons pourtant que ce type de contenu est plus facilement atteignable par des méthodes automatiques. La subjectivité de l'impact pictural ne s'exprime pas en terme de mots ou de concepts hauts-niveaux. L'impact pictural est au contraire plus *bas niveau*, car il se traduit essentiellement en termes géométriques.

En conclusion de cette partie sur la géométrie des œuvres au trait, soulignons que nous n'avons pas proposé de critère de similarité comme dans le Chapitre 4. Un critère d'arrêt similaire utilisant une méthode *a contrario* pourrait néanmoins être proposé en supposant l'indépendance des descripteurs et en apprenant leurs marginales sur la base. Cette indépendance est néanmoins assez discutable ici, et par ailleurs, la taille des bases que nous avons constituées ne permet pas d'apprécier l'intérêt du critère.

Chapitre 7

Conclusion et perspectives

Conclusion

Nous avons étudié dans cette thèse deux aspects du contenu pictural d'une oeuvre. Le premier concerne la composition en termes d'organisation spatiale des couleurs. Nous avons proposé une méthode de comparaison de cette composition sous la forme d'un problème de transport entre deux images. Cette approche se différencie des méthodes classiques de recherche d'images par régions homogènes par le fait que nous n'utilisons pas de méthodes de segmentation. Nous proposons d'utiliser une forme de sur-segmentation homogène et non supervisée en combinaison avec une distance de transport. Cette approche donne des résultats très satisfaisants sur un grand nombre de bases dont certaines dépassant 10000 images. Nous avons montré également que cette solution apporte des résultats qualitatifs plus robustes que les méthodes à base de segmentation. Nous avons également proposé un seuil automatique sur la distance de transport permettant de répondre au problème, rarement abordé de manière adaptative, du nombre d'images à retourner à l'utilisateur. Nous avons montré que ce critère donnait généralement un bon ordre de grandeur du nombre d'images pertinentes contenues dans une base.

Le deuxième aspect que nous avons abordé concerne l'étude de la géométrie des lignes dans une oeuvre et de l'impact pictural qu'elles produisent. L'impact pictural est un aspect intuitif de l'oeuvre que nous avons défini comme étant l'impact visuel transmis chez le spectateur par le contenu pictural d'une oeuvre, par opposition à l'impact que peut produire ce qui a été sémantiquement représenté par l'artiste. Cette définition couvre en fait un grand nombre d'éléments picturaux qui peuvent être extraits et analysés puisqu'ils sont souvent à des niveaux bas ou intermédiaires dans l'échelle des contenus d'une image. Nous avons proposé un ensemble de méthodes pour extraire les extrémités des traits, les croisements et les coins, les points d'inflexions ainsi que le rayon de l'outil utilisé. Ces éléments géométriques font partie du vocabulaire sur les primitives 1D que l'artiste peut exploiter dans un dessin au trait. Ces éléments géométriques sont analysés à partir de

contours des traits que nous détectons en utilisant une méthode automatique de sélection des lignes de niveaux proposée également dans cette thèse.

Perspectives de recherche

Le principal inconvénient de notre approche proposée pour l'étude de l'organisation spatiale des couleurs reste son coût calculatoire. Une requête sur une base de cent mille images demanderait plusieurs minutes sur un ordinateur classique. Même si l'interrogation d'une base est une opération qui peut être parallélisée il est intéressant d'optimiser le coût de cette méthode. Plusieurs approximations de la distance EMD ont été proposées dans la littérature, mais aucune ne permet d'espérer un gain significatif dans le cadre de distributions aussi grandes que celles que nous utilisons ($n = 10 \times 15$). Par manque de temps, nous n'avons pas cherché dans cette thèse à tester une approche imposant une contrainte spatiale de déplacement maximal des pixels. Une telle approche pourrait conduire à une matrice des coûts creuse, ce qui permettrait l'utilisation de méthodes plus performante pour la résolution du problème d'affectation. Nous pourrions également étudier des méthodes multi-niveaux (*coarse to fine*) de calculs de distance. Ces méthodes permettent de limiter le calcul de la distance avec la distance EMD sur une partie restreinte de la base, l'autre partie étant constituée d'images qui sont très éloignées et qui peuvent être détectées avec une distance moins coûteuse.

Dans le prolongement de notre étude des bipixels, il pourrait être de plus intéressant d'étudier la possibilité d'un problème de transport de petites régions ou *patches* de pixels comme cela a été notamment proposé par [Ling and Okada, 2007]. Par ailleurs, nous avons vu que pour certaines bases dont le contenu est particulier (comme les photographies en noir et blanc), une modification de la distance d_e utilisée permettait une amélioration des résultats. Ils serait intéressant d'étudier plus profondément les paramètres et les possibilités offertes par l'adaptation de cette distance à d'autres contenus particuliers comme les dessins, les photographies, les illustrations homogènes (comme les cliparts), etc.

La méthodologie proposée pour l'étude de l'impact pictural lié aux traits a été restreinte dans cette thèse aux dessins au trait. Il serait intéressant de pouvoir extraire les primitives 1D dans tout type d'oeuvres. En suivant la terminologie de Willats, une perspective pourrait être ensuite de proposer une méthode d'analyse de l'impact pictural liée aux formes 2D présentes dans une oeuvre. Combinés à une approche analysant les primitives de textures, une méthodologie complète de l'étude du contenu géométrique des oeuvres pourrait être proposée.

Par ailleurs, nous avons souligné aussi l'intérêt d'une approche analysant les primitives 1D pour la communauté de recherche en synthèse d'images non réalistes. Une perspective intéressante serait d'intégrer la méthodologie proposée dans la partie II pour permettre le transfert d'impact pictural ou de style à des images de synthèse dont la scène 3D est connue.

Un domaine d'application qui pourrait bénéficier directement des contributions apportées dans cette thèse pour l'étude des lignes est le domaine de la reconnaissance de caractères. La méthode de sélection que nous proposons peut être appliquée aux documents imprimés sur fond homogène, et elle constitue un prétraitement robuste pour soit binariser l'image, ou obtenir les contours des lettres. Concernant les documents cursifs, il serait intéressant de tester l'approche proposée pour l'authentification de documents ou de signatures.

Contexte applicatif

Au niveau du contexte applicatif, plusieurs perspectives nous semblent dignes d'un intérêt particulier au domaine du patrimoine culturel. Le parcours d'un visiteur sur le site internet d'un musée peut s'articuler selon différents scénarios. L'un de ces scénarios est celui où le visiteur se laisse guider par son goût pour l'aspect et l'esthétique d'un tableau. Chaque fiche d'œuvre pourrait proposer des *tremplins* au visiteur vers d'autres d'œuvres en lui signalant (par exemple sur le côté de l'écran ou en bas de page) quelques reproductions qui pourraient probablement lui plaire ou l'intéresser. Ceci consisterait par exemple à afficher sur chaque page d'œuvre les 3 ou 4 œuvres les plus similaires en utilisant cette œuvre comme requête (selon l'un des critères proposées dans cette thèse). Ce concept de *teasing* est notamment souvent exploité sur les sites de vente de livres sur internet comme Amazon.

Cette même idée serait d'ailleurs tout aussi intéressante pour les sites de galeries d'arts ou de ventes de reproductions (posters). Si un visiteur a une certaine sensibilité pour une OSC particulière, ou pour l'impact pictural d'un dessin (et plus directement son style), il est probable que de telles suggestions de tremplins vers d'autres images soient susceptibles de l'intéresser fortement.

Nous avons implanté récemment un tel système à l'Atelier Circulaire qui a collaboré avec nous durant cette thèse. Ce système fonctionne sur une base à l'Atelier regroupant les 1073 œuvres de la base ATCI. Cette base est actuellement utilisée par une seule personne en interne pour organiser les archives et la collection de l'Atelier. C'est pourquoi nous n'avons pas présenté de résultats dans cette thèse quant à l'utilisation de ce système. Il est actuellement à l'étude d'étendre l'utilisation de ce système au site internet de l'Atelier via un système de gestion du contenu [Hurtut, T. and Cheriet, F. and Chronopoulos, M., 2007].

Nous essayons également de monter un projet de recherche de 3 ans en collaboration avec l'artiste Carlos Calado, l'Atelier Circulaire, l'Ecole Polytechnique de Montréal, le Conseil des Arts du Canada et le CRSNG¹. Dans ce projet un système complet de production d'estampes numériques serait développé, de la conception jusqu'à l'impression 3D

¹Conseil de recherches en sciences naturelles et en génie du Canada

de la matrice. Ce système serait collaboratif en étant implanté au sein de l'atelier² et il intégrerait entre autres une base d'estampes numériques enrichie continuellement par les travaux des artistes. Il est prévu dans ce projet que le système de gestion de cette base utilise la méthodologie développée pour les traits 1D et intègre également une étude des primitives 2D.

Il serait également intéressant de mesurer l'intérêt que pourraient avoir des professionnels des musées, et en particulier les conservateurs, pour des outils de recherche incluant les critères proposés dans cette thèse. Nous n'avons pas eu l'opportunité de pouvoir mener une telle étude dans cette thèse. Les musées et institutions avec lesquels nous avons été en contact ont malheureusement été généralement réticents à ce type d'expériences. Les raisons invoquées ont toujours été celles de l'atteinte possible aux droits d'auteurs si nous demandions un échantillon de leurs bases pour effectuer nos tests préliminaires. De ce point de vue, nous tenons à remercier particulièrement l'IRHT pour nous avoir fourni leur base. Nous remercions également très chaleureusement l'Atelier Circulaire et ses artistes membres pour nous avoir confié leurs oeuvres à des fins de numérisation, toujours écouté avec une oreille patiente et curieuse, et pour avoir été prêts à nous accompagner dans les multiples projets que nous leur avons soumis.

²L'Atelier Circulaire compte environ une centaine d'artistes membres.

Annexe A

Bases d'images utilisées dans les expériences

A.1 Bases d'images utilisées pour l'étude de l'organisation spatiale des couleurs

Les expériences étudiant l'organisation spatiale des couleurs dans une oeuvre utilise 6 bases d'images différentes, résumées dans la Tableau A.1. Nous détaillons ici leurs caractéristiques, ainsi que leur origine. Le contenu de chacune d'elle est particulier, et leur taille varie de 1000 images à 25000 images. Nous avons aussi inclu une base de photographies non artistiques (CLIC) pour observer l'intérêt et le comportement des méthodes proposées sur une base de photos classiques.

La base IRHT

La première et principale base d'images que nous avons utilisé est celle qui nous a été fournie par l'Institut de Recherche et d'Histoire des Textes (IRHT)¹. Cette base contient 1499 enluminures numérisées avec une grande résolution et dont les couleurs ont été calibrées. Cette base est utilisée pour illustrer la majorité des expériences préliminaires à celles des recherches d'images.

¹IRHT, CNRS, 40 avenue d'Iena, 76116 Paris. URL : <http://www.irht.cnrs.fr/>

La base ATCI

La base ATCI est constituée de 1073 estampes originales. Une estampe est une oeuvre sur papier généralement imprimée à l'aide d'une presse, et issue d'une matrice (plaque de cuivre, bois, zinc, pierre lithographique ...) gravée en creux, en relief ou à plat par l'artiste. Ces oeuvres sont généralement imprimées à un tirage limité à quelques dizaines d'exemplaires. La base ATCI nous a été fournie par l'Atelier Circulaire² à Montréal. Environ 100 artistes travaillent dans cet atelier collectif. L'atelier dispose d'une collection propre enrichie continuellement par les artistes qui pour chaque édition, donne une épreuve dite *d'atelier*. C'est cette base qui a été numérisée par l'Atelier Circulaire à 70% et par moi-même à 30%. J'ai par ailleurs conçu le système d'acquisition et d'archivage mis à la disposition de l'Atelier Circulaire [Hurtut, T. and Cheriet, F. and Chronopoulos, M., 2007].

La base MIXCA

La base MIXCA est constituée de 25032 peintures, dessins et objets issus des collections de 3 grands musées Canadiens : le Musée des Beaux-Arts de Montréal, le Musée des Beaux-Arts du Canada, et le Musée Canadien des Civilisations à Hull. Ces images ont été collectées sur le site du Musée Virtuel du Canada³.

La base LCPD

Cette base est constituée de 21094 portraits photographiés au format vertical en noir et blanc. Elle nous a été fournie par Nies Huijsmans de l'université Leiden⁴.

La base CLIC

La base CLIC-CEA⁵ est une base d'environ 40 classes composées de photographies généralistes (Botanique, Animaux, Architecture ...). Elle contient au total 15200 photographies.

²Atelier collectif d'artistes à Montréal. 5445, av. de Gaspé espace 503 Montréal (Québec) H2T 3B2.
URL : <http://www.atelier-circulaire.qc.ca/>

³www.museevirtuel.ca

⁴Leiden 19th-Century Portrait Database <http://nies.liacs.nl> : 1860/

⁵<http://www.irit.fr/RFIEC/collection/images/clic.htm>

Institut	Contenu	Taille	Label
Institut de Recherche en Histoire des Textes	Enluminures	1500	IRHT
Atelier Circulaire	Estampes	1000	ATCI
Leiden 19 th -Century Portrait Database	Photographies noir et blanc	20000	LCPD
Musée virtuel du Canada	Peintures, dessins, objets photographiés	25000	MIXCA
CLIC-CEA	Photographies non artistiques	15000	CLIC

TAB. A.1: Bases d'images utilisées dans les expériences de la Partie I de la thèse

A.2 Bases d'images utilisées pour l'étude du contenu géométrique

Les expériences pour l'étude du contenu géométrique dans une oeuvre à base de traits ont été menées à partir de 3 bases d'images différentes dont nous avons mélangé certaines parties pour constituer une base classée (Tableau A.2). Nous détaillons ici leurs caractéristiques, ainsi que leur origine. Le contenu de chacune d'elle est particulier, et leur taille varie de 60 images à 2200 images. Dans toutes les expériences les images sont converties en niveaux de gris codé sur 8 bits.

La base Kelly-Matisse

Cette base d'images est composée de 225 dessins et lithographies au crayon de deux artistes, Matisse (110 oeuvres) et Ellsworth Kelly (115 oeuvres). La majorité de cette base a été constitué à partir du catalogue de l'exposition [Remi Labrusse, 2002] qui s'est tenu au Centre Pompidou en 2002 (environ 150 images). Le conservateur responsable de cette exposition souligne dans le catalogue la sensibilité artistique qui ressort des oeuvres présentées à travers les variations et imperfections des lignes. Ces oeuvres ont toutes pour sujet des plantes ou feuilles d'arbres. Nous avons aussi utilisé deux autres catalogues de ces deux artistes portant sur le même thème pour enrichir cette base [Grammont, 2003, Kelly, 2005].

La numérisation a été faite avec un scanner HP Scanjet 8200 à une résolution de 4800 ppi (8bits). La taille réelle des oeuvres est connue grâce à ces catalogues, et toutes les oeuvres ont été ensuite ramenée à la même résolution spatiale de $4\text{pixels}/\text{mm}^2$.

La base Picasso

Cette base contient 60 croquis de Picasso issu de plusieurs fac-similés numériques réalisés par la Réunion des Musée Nationaux et édités en 2006 dans un DVD-Rom [RMN,

A.3. Remerciements

2006]. Les images sont de tailles 800×600 . Tous les carnets originaux ont la même taille de pages.

La base ArtExplosion

La base ArtExplosion contient environ 200000 dessins faits par ordinateur. Nous en avons sélectionné 2200 qui sont réalisés à base de lignes.

Origine	Contenu	Taille	Label
Numérisation de catalogues d'expositions	Dessins et lithographies de plantes	225	Kelly-Matisse
Fac-similé numérique de 3 carnets de croquis de Picasso	Dessins au crayon	60	Picasso
Sélection de la base ArtExplosion	Cliparts	2200	Explosion

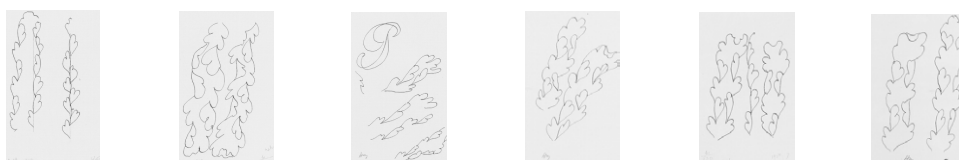
TAB. A.2: Bases d'images utilisées dans les expériences de la Partie II de la thèse

La base classée

A partir de trois bases précédentes, nous avons constitué une base de 105 oeuvres classée subjectivement par moi-même en 14 catégories. Chaque classe est composée de dessins ayant un impact pictural similaires. Ces 14 classes sont visibles sur les Figures [A.1](#) à [A.3](#).

A.3 Remerciements

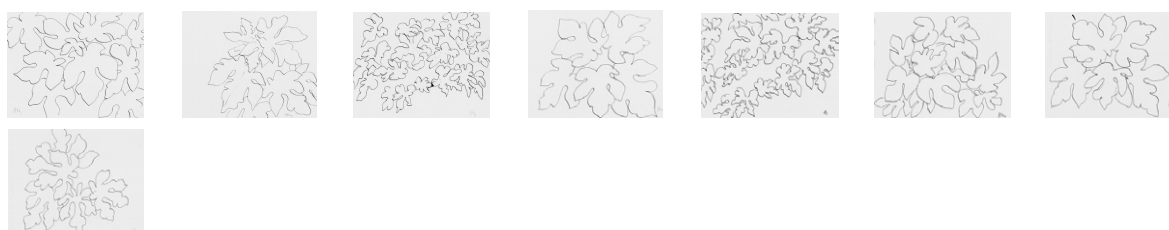
Nous remercions vivement Gilles Kagan de l'IRHT, Maria Chronopoulos à l'Atelier Circulaire (ainsi que tous les artistes qui y travaillent et qui nous ont prêté leurs oeuvres), le Professeur Nies Huijsmans de l'université Leiden de nous avoir donné l'autorisation d'utiliser leurs bases de données.



Classe 1



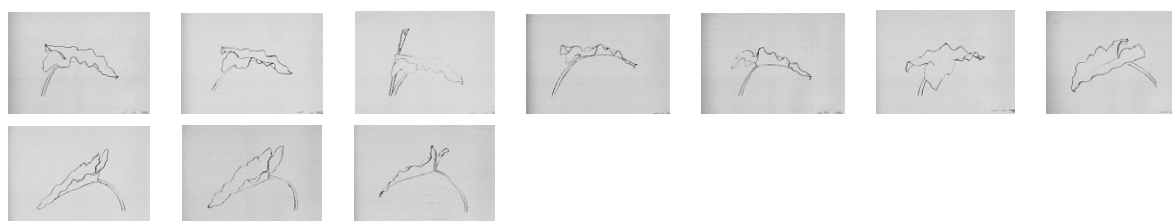
Classe 2



Classe 3



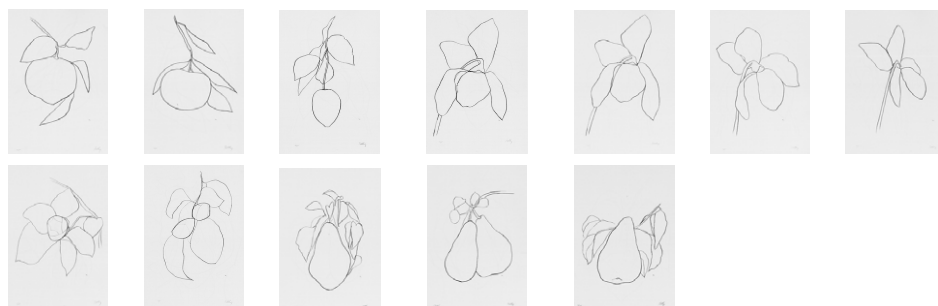
Classe 4



Classe 5

FIG. A.1: Base d'oeuvres classées. Classes 1 à 5.

A.3. Remerciements



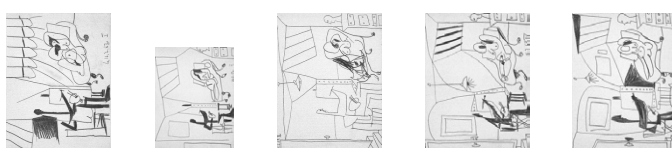
Classe 6



Classe 7



Classe 8



Classe 9

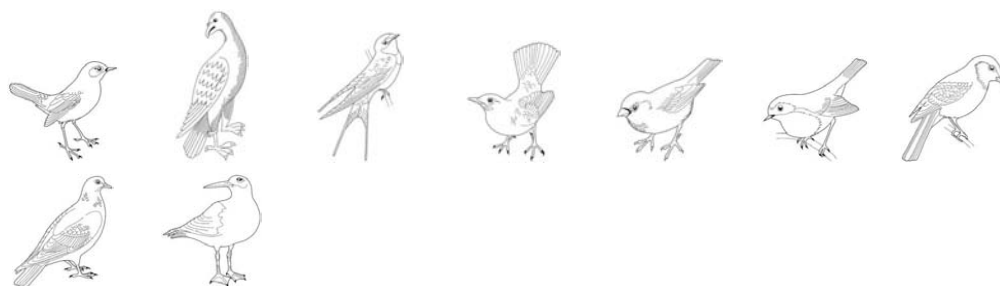


Classe 10

FIG. A.2: Base d'oeuvres classées. Classes 6 à 10.



Classe 11



Classe 12



Classe 13



Classe 14

FIG. A.3: Base d'oeuvres classées. Classes 11 à 14.

Annexe B

Problème de transport optimal à deux classes de poids total égaux

B.1 Notations et définition du problème

Dans cette section, nous nous intéressons au transport optimal de deux classes vers deux classes dans le cas où la masse totale des deux classes de départ est égale à la masse totale des deux classes d'arrivée. Nous utilisons la méthode classique de résolution d'un simplexe adaptée au problème des transports, que l'on peut trouver par exemple dans [Simonnard, 1962]. La distance EMD (voir Equation (3.1)), dans le cadre du problème 2×2 correspondant au transport d'une classe de départ $\mathbf{C}_d = \{(\mathcal{C}_1^d, p_1^d), (\mathcal{C}_2^d, 1 - p_1^d)\}$ vers une classe d'arrivée $\mathbf{C}_a = \{(\mathcal{C}_1^a, p_1^a), (\mathcal{C}_2^a, 1 - p_1^a)\}$ ayant le même poids total $p_1^d + p_2^d = p_1^a + p_2^a = 1$ est défini par :

$$d(\mathbf{C}_d, \mathbf{C}_a) = \min_{x_{kp}} \sum_{k=1,2} \sum_{p=1,2} d_e(\mathcal{C}_k^d, \mathcal{C}_p^a) x_{kp}, \quad (\text{B.1})$$

où les *flux* x_{kp} suivent les contraintes suivantes :

$$\forall k = 1, 2, \forall p = 1, 2, x_{kp} \geq 0$$

,

$$\begin{aligned} \forall k = 1, 2, \quad \sum_{p=1,2} x_{kp} &= p_k^d, \\ \forall p = 1, 2, \quad \sum_{k=1,2} x_{kp} &= p_p^a, \end{aligned}$$

et où la distance $d_e(\cdot, \cdot)$ est une distance au sol, associée à la distance EMD $d(\cdot, \cdot)$ ci dessus.

B.2. Chemins de départs

Pour suivre la méthode classique de résolution, nous devons d'abord construire la matrice 2×2 des flux x_{kp} et des coûts $c_{kp} = d_e(\mathcal{C}_k^d, \mathcal{C}_p^a), \forall k, p = 1, 2$ tel qu'indiqué sur la Figure B.1.

p_1^d	x_{11}	x_{12}	avec les coûts	c_{11}	c_{12}
p_2^d	x_{21}	x_{22}		c_{21}	c_{22}
	p_1^a	p_2^a			

FIG. B.1: Mode de représentation du problème des transports entre deux bicouleurs A gauche la matrice des flux x_{kp} avec en marge les différents poids p . A droite, la matrice des coûts $c_{kp} = d_e(\mathcal{C}_k^1, \mathcal{C}_p^2), \forall k, p = 1, 2$ associés.

B.2 Chemins de départs

La technique générale consiste à trouver un chemin de départ dans le tableau des coûts puis à l'optimiser en le modifiant jusqu'à ce qu'il respecte une condition d'optimisation que nous allons expliciter. Pour trouver un flux de départ, nous utilisons ici la méthode du coin supérieur, c'est à dire ici en choisissant $x_{11} = \min\{p_1^d, p_1^a\}$. Nous traiterons le cas de dégénérescence $p_1^d = p_1^a$ en fin de section. Nous supposons donc d'abord que $p_1^d > p_1^a$ ou $p_1^d < p_1^a$. Nous voyons alors que :

Cas A : si $p_1^d < p_1^a$, les flux deviennent	0	p_1^d	0
	p_2^d	.	.
		$p_1^a - p_1^d$	p_2^a
Cas B : si $p_1^d > p_1^a$, les flux deviennent	$p_1^d - p_1^a$	p_1^a	.
	p_2^d	0	.
		0	p_2^a

FIG. B.2: Première étape de la recherche d'un chemin de départ.

L'étape suivante consiste à éliminer la ligne ou colonne dont le poids a été annulé, puis à considérer la case immédiatement à droite ou au dessus dans les cases restantes. À nouveau pour cette case nous devons fixer le flux au minimum des deux poids associés en ligne et colonne. Dans le *cas A* de la Figure B.2, il s'agit de $x_{21} = \min\{p_1^a - p_1^d, p_2^d\}$, or $p_1^a - p_1^d - p_2^d = p_1^a - 1 < 0$ donc $\min\{p_1^a - p_1^d, p_2^d\} = p_1^a - p_1^d$. De même, dans le deuxième cas de la Figure B.2, nous avons $x_{12} = \min\{p_1^d - p_1^a, p_2^a\} = p_1^d - p_1^a$. Les tableaux deviennent donc :

Les poids restant à déplacer se simplifient et sont bien égaux aux poids restant à atteindre (e.g. $p_2^d - (p_1^a - p_1^d) = p_2^a$ pour le *cas A*). Les tableaux finaux des deux chemins de départ possibles sont donc :

Cas A : si $p_1^d < p_1^a$, les flux sont	0	<table><tr><td>p_1^d</td><td>0</td></tr><tr><td>$p_1^a - p_1^d$</td><td>.</td></tr></table>	p_1^d	0	$p_1^a - p_1^d$	.
	p_1^d	0				
$p_1^a - p_1^d$	.					
$p_2^d - (p_1^a - p_1^d)$	<table><tr><td>0</td><td>p_2^a</td></tr></table>	0	p_2^a			
0	p_2^a					
Cas B : si $p_1^d > p_1^a$, les flux sont	0	<table><tr><td>p_1^a</td><td>$p_1^d - p_1^a$</td></tr><tr><td>0</td><td>.</td></tr></table>	p_1^a	$p_1^d - p_1^a$	0	.
	p_1^a	$p_1^d - p_1^a$				
0	.					
p_2^d	<table><tr><td>0</td><td>$p_2^a - (p_1^d - p_1^a)$</td></tr></table>	0	$p_2^a - (p_1^d - p_1^a)$			
0	$p_2^a - (p_1^d - p_1^a)$					

FIG. B.3: Deuxième étape de la recherche d'un chemin de départ.

Cas A : si $p_1^d < p_1^a$, les flux sont	0	<table> <tr> <td>p_1^d</td> <td>0</td> </tr> <tr> <td>$p_1^a - p_1^d$</td> <td>p_2^a</td> </tr> </table>	p_1^d	0	$p_1^a - p_1^d$	p_2^a
	p_1^d	0				
$p_1^a - p_1^d$	p_2^a					
0	<table> <tr> <td>0</td> <td>0</td> </tr> </table>	0	0			
0	0					

Cas B : si $p_1^d > p_1^a$, les flux sont	0	<table> <tr> <td>p_1^a</td> <td>$p_1^d - p_1^a$</td> </tr> <tr> <td>0</td> <td>p_2^d</td> </tr> </table>	p_1^a	$p_1^d - p_1^a$	0	p_2^d
	p_1^a	$p_1^d - p_1^a$				
0	p_2^d					
p_2^d	<table> <tr> <td>0</td> <td>0</td> </tr> </table>	0	0			
0	0					

FIG. B.4: Chemins de départs possibles finaux

B.3 Condition d'optimalité

Ces deux chemins sont-ils optimaux? La condition d'optimalité utilise le problème dual (voir [Simonnard, 1962]). Elle consiste à résoudre tout d'abord le système constitué des équations $u_k + v_p = c_{kp}$ aux inconnues duales u_k, v_p pour tous les couples (k, p) qui sont associés à un flux x_{kp} non nul. Le système est résolu en fixant arbitrairement une inconnue, e.g. $u_1 = 1$. La condition d'optimalité est remplie si $D_{kp} = u_k + v_p - c_{kp} \leq 0$ pour tous les couples (k, p) qui ont un flux x_{kp} nul. Dans notre situation, il n'y a qu'un seul flux nul dans chaque cas de figure, et les conditions sont $D_{12} \leq 0$, et $D_{21} \leq 0$. Si nous résolvons les systèmes nous voyons qu'ils sont symétriques, et nous avons $D_{12} = D_{21} = c_{11} + c_{22} - c_{21} - c_{12}$ dans les deux cas de la Figure B.4. Nous noterons ce déterminant D .

Si $D \leq 0$, le chemin trouvé est donc optimal, et alors la distance EMD se calcule à partir des flux trouvés dans le tableau B.4. Sinon, il faut améliorer les deux chemins trouvés en diminuant le flux de rang pair dont le coût est minimal sur le cycle fermé du chemin précédent commençant par le flux nul. Dans les deux cas A et B, les cases de rang paire sont les cases (1, 1) et (2, 2), et celle dont le flux est minimal est nécessairement la case (2, 2). Dans le cas A, cela revient donc à diminuer x_{11} de p_2^a , et dans le cas B, cela revient à diminuer x_{11} de p_2^d . Les chemins sont alors indiqués sur la Figure B.5.

Nous voyons que ces deux chemins optimisés sont égaux. De plus, le déterminant se résume ici à $D_{22} = c_{12} + c_{21} - c_{22} - c_{11} = -D$ qui est négatif en cas de non optimalité des

B.3. Condition d'optimalité

$$\begin{array}{ll}
 \text{Cas A si } p_1^d < p_1^a \text{ et } D > 0, \text{ les flux sont} & \begin{array}{cc} 0 & \begin{array}{|c|c|} \hline p_1^d - p_2^a & p_2^a \\ \hline 1 - p_1^d & 0 \\ \hline 0 & 0 \end{array} \\ 0 & \end{array} \\
 \text{Cas B si } p_1^d > p_1^a \text{ et } D > 0, \text{ les flux sont} & \begin{array}{cc} 0 & \begin{array}{|c|c|} \hline p_1^a - p_2^d & 1 - p_1^a \\ \hline p_2^d & 0 \\ \hline 0 & 0 \end{array} \\ p_2^d & \end{array}
 \end{array}$$

FIG. B.5: Chemins dans le cas où $D > 0$

chemins précédents. La distance EMD se calcule à nouveau à partir des flux optimaux.

Nous obtenons donc la règle de calcul suivante :

Algorithme 4 Calcul de la distance EMD entre deux classes de poids total égaux

```

si  $D = c_{11} + c_{22} - c_{21} - c_{12} > 0$  alors
     $d(\mathbf{C}_d, \mathbf{C}_a) = p_2^d c_{21} + (p_1^a - p_2^d) c_{11} + p_2^a c_{12}$ 
sinon
    si  $p_1^d < p_1^a$  alors
         $d(\mathbf{C}_d, \mathbf{C}_a) = p_1^d c_{11} + p_1^a c_{22} + (p_1^a - p_1^d) c_{21}$ 
    sinon
         $d(\mathbf{C}_d, \mathbf{C}_a) = p_1^a c_{11} + p_1^d c_{22} + (p_2^a - p_2^d) c_{21}$ 
    fin si
fin si

```

Cas dégénéré : dans notre situation il y a dégenescence du problème si et seulement si $p_1^d = p_1^a$. Il s'en déduit aussi que $p_2^d = p_2^a$. On réappliquant la même méthode (recherche de chemin de départ, puis condition d'optimalité) on obtient que $d(\mathbf{C}_d, \mathbf{C}_a) = p_1^a c_{11} + p_1^d c_{22}$ si $D \leq 0$ ce qui est un cas particulier mais couvert par l'Algorithme 4. La même formule est obtenue si $D > 0$

Bibliographie

- [Agam and Dinstein, 1997] Agam, G. and Dinstein, I. (1997). Geometric separation of partially overlapping nonrigid objects applied to automatic chromosome classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 11(19) :1212–1222.
- [Aggarwal et al., 2002] Aggarwal, G., T.V., A., and Ghosal, S. (2002). An image retrieval system with automatic query modification. *IEEE Trans. on Multidimedia*, 4(2) :201–214.
- [Ales et al., 1991] Ales, J.-P., Froment, J., and Morel, J.-M. (1991). Reconstruction visuelle et généricité. *Intellectica*, 1 :11–35.
- [Almansi et al., 2003] Almansi, A., Desolneux, A., and Vamech, S. (2003). Vanishing points detection without any a priori information. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25(4) :502–507.
- [Alvarez et al., 1999] Alvarez, L., Gousseau, Y., and Morel, J.-M. (1999). The size of objects in natural and artificial images. *Advances in Imaging and Electron Physics, Academic Press*, 111 :167–242.
- [Arnheim, 1954] Arnheim, R. (1954). *Art and visual perception*. California Press.
- [Arnheim, 1969] Arnheim, R. (1969). *Visual thinking*. California Press.
- [Attneave, 1954] Attneave, F. (1954). Some informational aspects of visual perception. *Psychol. Rev.*, 61(3) :183–93.
- [Bach et al., 1996] Bach, J., Fuller, C., Gupta, A., Hampapur, A., Horowitz, B., Humphrey, R., Jain, R., and Shu, C. (1996). Virage image search engine : an open framework for image management. *Proceedings of SPIE*, 2670 :76.
- [Bae and Durand, 2004] Bae, S. and Durand, F. (2004). Statistical analysis and transfer of pictorial styles. *MIT Workshop, Oxygene*.
- [Barnet, 1985] Barnet, S. (1985). *A short guide to writing about art*. Little, Brown Boston.
- [Bentley, 1975] Bentley, J. (1975). Multidimensional binary search trees used for associative searching. *Communications of the ACM*, 18(9) :509–517.
- [Berezhnoy et al., 2005] Berezhnoy, I., Postma, E., and van den Herik, H. (2005). Authentic : Computerized Brushstroke Analysis. *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on*, pages 1586–1588.

- [Berezhnoy et al., 2004] Berezhnoy, I., Postma, E., and van den Herik, J. (2004). Digital analysis of van gogh's complementary colours. *BNAIC*, pages 163–170.
- [Berezhnoy et al., 2007] Berezhnoy, I., Postma, E., and van den Herik, J. (2007). Computer analysis of Van Gogh's complementary colours. *Pattern Recognition Letters*, 28(6) :703–709.
- [Berretti et al., 1999] Berretti, S., Del Bimbo, A., and Vicario, E. (1999). Weighting spatial arrangement of colors in content based imageretrieval. *Multimedia Computing and Systems, 1999. IEEE International Conference on*, 1.
- [Berretti et al., 2000] Berretti, S., Del Bimbo, A., and Vicario, E. (2000). The Computational Aspect of Retrieval by Spatial Arrangement. *International Conference on Pattern Recognition*, page 1.
- [Berretti et al., 2002] Berretti, S., Del Bimbo, A., and Vicario, E. (2002). Spatial arrangement of color in retrieval by visual similarity. *Pattern Recognition*, 35 :1661–1674.
- [Berretti et al., 2003] Berretti, S., Del Bimbo, A., and Vicario, E. (2003). Weighted walk-throughs between extended entities for retrieval by spatial arrangement. *Multimedia, IEEE Transactions on*, 5(1) :52–70.
- [Bersani, 1968] Bersani, J. (1968). *Encyclopædia universalis*. Encyclopædia universalis.
- [Biedermann, 1987] Biedermann (1987). Recognition-by-components : a theory of human image understanding. *Psychol. Rev.*, 94 :115–147.
- [Bird et al., 1999] Bird, C. L., Elliott, P. J., and Hayward, P. M. (1999). Content-Based Retrieval for European Image Libraries.
- [Blake and Zisserman, 1998] Blake, A. and Zisserman, A. (1998). *Active Contours*. Springer Verlag.
- [Bordenave et al., 2006] Bordenave, C., Gousseau, Y., and Roueff, F. (2006). The dead leaves model : a general tessellation modeling occlusion. *Adv. in Appl. Probab*, 38(1) :31–46.
- [Bottou et al., 1998] Bottou, L., Haffner, P., Howard, P., Simard, P., Bengio, Y., and LeCun, Y. (1998). High quality document image compression with DjVu. *Journal of Electronic Imaging*, 7(3) :410–425.
- [Boujemaa et al., 2001] Boujemaa, N., Fauqueur, J., Ferencat, M., Fleuret, F., Gouet, V., Saux, B., and Sahbi, H. (2001). Ikona : Interactive generic and specific image retrieval. *Proceedings of the International workshop on Multimedia Content-Based Indexing and Retrieval (MMCBIRŠ2001)*, pages 25–28.
- [Burford et al., 2003] Burford, B., Briggs, P., and Eakins, J. (2003). A Taxonomy of the Image : On the Classification of Content for Image Retrieval. *Visual Communication*, 2(2) :123.
- [Burkard and Čela, 1998] Burkard, R. and Čela, E. (1998). *Linear assignment problems and extensions*. Karl-Franzens-Univ. Graz & Techn. Univ. Graz.
- [Burkard et al., 2008] Burkard, R., Dell'Amico, M., and Martello, S. (2008). *Assignment Problems*. SIAM Monographs on Discrete Mathematics and Applications. To appear.

- [Canny, 1986] Canny, J. (1986). A computational approach to edge detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 8 :679–698.
- [Cao et al., 2005] Cao, F., Musé, P., and Sur, F. (2005). Extracting Meaningful Curves from Images. *Journal of Mathematical Imaging and Vision*, 22(2) :159–181.
- [Carr and Leonard, 1992] Carr, D. and Leonard, M. (1992). *Looking at Paintings : A Guide to Technical Terms*. J. Paul Getty Museum in association with British Museum Press.
- [Carson et al., 2002] Carson, C., Belongie, S., Greenspan, H., and Malik, J. (2002). Blob-world : image segmentation using em and its application to image querying. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 24(8) :1026–1038.
- [Caselles et al., 1997] Caselles, V., Kimmel, R., and Sapiro, G. (1997). Geodesic active contours. *Int. J. Comput. Vision*, 22 :61–79.
- [Caselles and Monasse, 2002] Caselles, V. and Monasse, P. (2002). Grain Filters. *Journal of Mathematical Imaging and Vision*, 17(3) :249–270.
- [Cavanagh, 1999] Cavanagh, P. (1999). Pictorial art and vision.
- [Cavanagh, 2003] Cavanagh, P. (2003). Les licences de Fra Carnevale. *Pour la Science*, 307 :104–105.
- [Cavanagh, 2005] Cavanagh, P. (2005). The artist as neuroscientist. *Nature*, 434(7031) :301–307.
- [Chang, 1991] Chang, C. (1991). Spatial match retrieval of symbolic pictures. *Journal of Information Science and Engineering*, 7(3) :405–422.
- [Chang et al., 1987] Chang, S., Shi, Q., and Yan, C. (1987). Iconic indexing by 2-D strings. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(3) :413–428.
- [Chen et al., 2002] Chen, C., Del Bimbo, A., Amato, G., Boujemaa, N., Bouthemy, P., Kittler, J., Pitas, I., Smeulders, A., Alexander, K., Kiernan, K., et al. (2002). Report of the DELOS-NSF Working Group on Digital Imagery for Significant Cultural and Historical Materials. *DELOS-NSF Reports, December*.
- [Chen et al., 2005] Chen, C., Wactlar, H., Wang, J., and Kiernan, K. (2005). Digital imagery for significant cultural and historical materials. *International Journal on Digital Libraries*, 5(4) :275–286.
- [Chen et al., 2003] Chen, K., Zhang, X., Ou, Z., and Feng, X. (2003). Recognition of digital curves scanned from paper drawings using genetic algorithms. *Pattern Recognition*, 36(1) :123–130.
- [Chen et al., 2007] Chen, Y., Tsai, M., Cheng, C., Chan, P., and Zhong, Y. (2007). Perimeter Intercepted Length and Color t-Value as Features for Nature-Image Retrieval. *Lecture Notes In Computer Science*, 4570 :185.
- [Cinque et al., 2001] Cinque, L., Ciocca, G., Levialdi, S., Pellicanò, A., and Schettini, R. (2001). Color-based image retrieval using spatial-chromatic histograms. *Image and Vision Computing*, 19(13) :979–986.

- [Ciocca et al., 2002] Ciocca, G., Schettini, R., and Cinque, L. (2002). Image indexing and retrieval using spatial chromatic histograms and signatures. *The First European Conference on Color in Graphics, Imaging and Vision (CGIV)*, pages 255–258.
- [Clouzot, 1997] Clouzot (1997). Les naissances de l’écriture. Vidéo.
- [Conniss et al., 2000] Conniss, L., Ashford, A., and Graham, M. (2000). *Information Seeking Behaviour in Image Retrieval VISOR I Final Report*. Institute for Image Data Research, University of Northumbria at Newcastle.
- [Corridoni et al., 1996] Corridoni, J. M., Del Bimbo, A., De Magistris, S., and Vicario, E. (1996). A visual language for color-based painting retrieval. In *VL ’96 : Proceedings of the 1996 IEEE Symposium on Visual Languages*, page 68, Washington, DC, USA. IEEE Computer Society.
- [Corridoni et al., 1997] Corridoni, J. M., Del Bimbo, A., and Pala, P. (1997). Retrieval of paintings using effects induced by color features. *IEEE Multimedia*, 6(3) :38–53.
- [Corridoni et al., 1998] Corridoni, J. M., Del Bimbo, A., and Pala, P. (1998). Retrieval of paintings using effects induced by color features. *Content-Based Access of Image and Video Database, 1998. Proceedings., 1998 IEEE International Workshop on*, pages 2–11.
- [Criminisi and Stork, 2004] Criminisi, A. and Stork, D. G. (2004). Did the great masters use optical projections while painting? *ICPR*, pages 645–648.
- [Dai and Zhang, 2005] Dai, S.-Y. and Zhang, Y.-J. (2005). Unbalanced region matching based on two-level description for image retrieval. *Pattern Recognition Letter*, 26 :565–580.
- [David, 2004] David, C. (2004). Analyse cognitive de la cohérence interindexeurs lors de l’indexation de documents. Master’s thesis, Université de Montréal.
- [Del Bimbo, 1999] Del Bimbo, A. (1999). *Visual information retrieval*. Morgan Kufman Publishers.
- [Del Bimbo et al., 1998] Del Bimbo, A., Mugnaini, M., Pala, P., and Turco, F. (1998). Visual querying by color perceptive regions. *Pattern Recognition*, 31(9) :1241–1253.
- [Del Bimbo and Pala, 1999] Del Bimbo, A. and Pala, P. (1999). Shape indexing by multi-scale representation. *Image and Vision Computing*, 17 :245–261.
- [Del Bimbo and Vicario, 1998] Del Bimbo, A. and Vicario, E. (1998). Using weighted spatial relationships in retrieval by visual contents. *Proceedings of the IEEE Workshop on Content-Based Access of Image and Video Libraries*, page 35.
- [Deng and Manjunath, 2001] Deng, Y. and Manjunath, B. (2001). Unsupervised segmentation of color-texture regions in images and video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(8) :800–810.
- [Deriche, 1987] Deriche, R. (1987). Using canny’s criteria to derive recursively implemented optimal edge detector. *Int. J. Computer Vision*, 1 :167–187.
- [Desolneux et al., 2001] Desolneux, A., Moisan, L., and Morel, J. (2001). Edge Detection by Helmholtz Principle. *Journal of Mathematical Imaging and Vision*, 14(3) :271–284.

- [Desolneux et al., 2000] Desolneux, A., Moisan, L., and Morel, J.-M. (2000). Meaningful alignments. *International Journal of Computer Vision*, 40 :7–23.
- [Desolneux et al., 2003] Desolneux, A., Moisan, L., and Morel, J.-M. (2003). A grouping principle and four applications. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25(4) :508–513.
- [Desolneux et al., 2008] Desolneux, A., Moisan, L., and Morel, J.-M. (2008). From Gestalt Theory to Image Analysis : A Probabilistic Approach, volume 34 of *Interdisciplinary Applied Mathematics*. Springer-Verlag. in Press.
- [Dubery and Willats, 1972] Dubery, F. and Willats, J. (1972). *Drawing Systems*. Studio Vista.
- [Dubois and Glanz, 1986] Dubois, S. R. and Glanz, F. H. (1986). An autoregressive model approach to two-dimensional shape classification. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 8 :55–66.
- [Durand et al., 2002] Durand, F., Agrawala, M., Gooch, B., Interrante, V., Ostromoukhov, V., and Zorin, D. (2002). Perceptual and artistic principles for effective computer depiction. *Course Notes for SIGGRAPH 2002, San Antonio, Texas*.
- [Dvir et al., 2002] Dvir, G., Greenspan, H., and Rubner, Y. (2002). Context-Based Image Modelling. *Proc. Int Conf. ICPR2002*, pages 162–165.
- [Eakins et al., 2004] Eakins, J., Briggs, P., and Burford, B. (2004). Image retrieval interfaces : A user perspective. *Proc. Third Int. Conf. on Image and Video Retrieval, Springer-Verlag Heidelberg, Dublin, Ireland*, pages 628–637.
- [Earnest, 1962] Earnest, L. (1962). Machine Recognition of Cursive Writing. *Information Processing*, 4 :462–466.
- [Egenhofer and Franzosa, 1991] Egenhofer, M. and Franzosa, R. (1991). Point-set topological spatial relations. *International Journal of Geographical Information Science*, 5(2) :161–174.
- [Explosion, 2002] Explosion, A. (2002). Hallogram. <http://www.hallogram.com/artexplosion/>.
- [Fantoni et al., 2005] Fantoni, C., Bertamini, M., and Gerbino, W. (2005). Contour curvature polarity and surface interpolation. *Vision Research*, 45(1047-1062) :7.
- [Fayolle et al., 2000] Fayolle, J., Riou, L., and Ducotter, C. (2000). Robustness of a multiscale scheme of feature points detection. *Pattern Recognition*, 33(9) :1437–1453.
- [Feldman and Singh, 2005] Feldman, J. and Singh, M. (2005). Information along contours and object boundaries. *Psychological Review*, 112(1) :243–252.
- [Flickner et al., 1995] Flickner, M., Niblack, H., Ashley, W., Dom, J., Gorkani, B., Hafner, M., Lee, J., Petkovic, D., Steele, D., Yanker, D., et al. (1995). Query by image and video content : the QBIC system. *Computer*, 28(9) :23–32.
- [Ford and Fulkerson, 1956] Ford, L. R. and Fulkerson, D. R. (1956). Solving the transportation problem. *Management Science*, 3 :24–32.
- [Fréchet, 1906] Fréchet, M. (1906). *Sur quelques points de calcul fonctionnel*. PhD thesis, s. n.

- [Freeman and Adelson, 1991] Freeman, W. and Adelson, E. (1991). The design and use of steerable filters. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 13(9) :891–906.
- [Freeman et al., 2003] Freeman, W. T., Tenenbaum, J. B., and Pasztor, E. C. (2003). Learning style translation for the lines of a drawing. *ACM Trans. Graph.*, 22(1) :33–46.
- [Frishkopf and Harmon, 1961] Frishkopf, L. and Harmon, L. (1961). Machine Reading of Cursive Script. *Information Theory (4th London Symp.)*, pages 300–316.
- [Froment et al., 2007] Froment, J., Moisan, L., and Morel, J.-M. (2007). Megawave 3.01. <http://www.cmla.ens-cachan.fr/fileadmin/Groupes/Megawave/>.
- [Garain et al., 2006] Garain, U., Paquet, T., and Heutte, L. (2006). On foreground-background separation in low quality document images. *International Journal on Document Analysis and Recognition*, 8(1) :47–63.
- [Georges, 1989] Georges, J. (1989). *L’écriture mémoire des hommes*. Découverte Gallimard.
- [Ghosh and Petkov, 2006] Ghosh, A. and Petkov, N. (2006). Effect of high curvature point deletion on the performance of two contour based shape recognition algorithms. *International Journal of Pattern Recognition and Artificial Intelligence*, 20(6) :913–924.
- [Giraudon, 1987] Giraudon, G. (1987). Chainage efficace des contours. Technical Report 605, INRIA.
- [Gollin, 1960] Gollin, E. S. (1960). Developmental studies of visual recognition of incomplete objects. *Percep. Motor. Skills*, 11 :289–298.
- [Gombrich, 2000] Gombrich, E. H. (2000). *Histoire de l’Art*. Phaidon.
- [Gombrich, 2002] Gombrich, E. H. (2002). *L’art et l’illusion*. Phaidon.
- [Gombrich and Eribon, 1998] Gombrich, E. H. and Eribon, D. (1998). *Ce que l’image nous dit : Entretiens sur l’art et la science*. Diderot.
- [Gong et al., 1996] Gong, Y., Chuan, C., and Xiaoyi, G. (1996). Image indexing and retrieval based on color histograms. *Multimedia Tools and Applications*, 2(2) :133–156.
- [Gonzales and Woods, 2002] Gonzales, R. C. and Woods, R. E. (2002). *Digital image processing, second edition*. Prentice Hall.
- [Gooch and Gooch, 2001] Gooch, B. and Gooch, A. (2001). *Non-Photorealistic Rendering*. AK Peters, Ltd. Natick, MA, USA.
- [Goto and Aso, 1999] Goto, H. and Aso, H. (1999). Extracting curved text lines using local linearity of the text line. *International Journal on Document Analysis and Recognition*, 2(2) :111–119.
- [Gousseau, 2003] Gousseau, Y. (2003). Comparaison de la composition de deux images, et application à la recherche automatique. In *proceedings of GRETSI 2003*, Paris, France.
- [Gousseau and Roueff, 2006] Gousseau, Y. and Roueff, F. (2006). Modeling occlusion and scaling in natural images. To appear in SIAM Multiscale Modeling and Simulations.

- [Grabli, 2005] Grabli, S. (2005). Le style dans le rendu non-photoréaliste de dessins au trait — à partir de scènes 3d : une approche programmable. Master's thesis, Université Joseph Fourier de Grenoble (UJF).
- [Grabli et al., 2004] Grabli, S., Turquin, E., Durand, F., and Sillion, F. X. (2004). Programmable style for npr line drawing. *Eurographic Symposium on Rendering*.
- [Grammont, 2003] Grammont, C. (2003). *Matisse et l'arbre*. Hazan, Le Cateau-Cambrésis : Musée Matisse.
- [Greenspan et al., 2000] Greenspan, H., Dvir, G., and Rubner, Y. (2000). Region Correspondence for Image Matching via EMD Flow. *IEEE Workshop on Content-based Access of Image and Video Libraries (CBAIVL-2000)*, pages 27–31.
- [Grimson and Huttenlocher, 1991] Grimson, W. and Huttenlocher, D. (1991). On the verification of hypothesized matches in model-based recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 13(12) :1201–1213.
- [Gu and Tjahjadi, 2000] Gu, Y. H. and Tjahjadi, T. (2000). Coarse-to-fine planar object identification using invariant curve features and B-spline modeling. *Pattern Recognition*, 33(9) :1411–1422.
- [Gunsel et al., 2005] Gunsel, B., Sariel, S., and Icoğlu, O. (2005). Content-Based Access to Art Paintings. *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, 2.
- [Hafner et al., 1995] Hafner, J., Shawhney, H. F., Equitz, W., Flickner, M., and Niblack, W. (1995). Efficient color histogram indexing for quadratic form distance function. *IEEE Transaction on Pattern Recognition and Machine Intelligence (PAMI)*, 17(7) :729–736.
- [Hamel and Strothotte, 1999] Hamel, J. and Strothotte, T. (1999). Capturing and re-using rendition styles for non-photorealistic rendering. In Brunet, P. and Scopigno, R., editors, *Computer Graphics Forum (Eurographics '99)*, volume 18(3), pages 173–182. The Eurographics Association and Blackwell Publishers.
- [Han and Fan, 1994] Han, C. and Fan, K. (1994). Skeleton generation of engineering drawings via contour matching. *Pattern recognition*, 27(2) :261–275.
- [Han and Poston, 2001] Han, J. and Poston, T. (2001). Chord-to-point distance accumulation and planar curvature : a new approach to discrete curvature. *Pattern Recognition Letters*, 22(10) :1133–1144.
- [Haralick, 1984] Haralick, R. (1984). Digital step edges from zero crossing of second directional derivatives. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 6 :58–68.
- [He and Yung, 2004] He, X. and Yung, N. (2004). Curvature scale space corner detector with adaptive threshold and dynamic region of support. *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, 2.
- [Heeger and Bergen, 1995] Heeger, D. and Bergen, J. (1995). Pyramid-based texture analysis/synthesis. *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pages 229–238.

- [Hertzmann et al., 2002] Hertzmann, A., Oliver, N., Curless, B., and Seitz, S. M. (2002). Curve analogies. In *EGRW '02 : Proceedings of the 13th Eurographics workshop on Rendering*, pages 233–246, Aire-la-Ville, Switzerland, Switzerland. Eurographics Association.
- [Hoffman and Singh, 1997] Hoffman, D. and Singh, M. (1997). Saliency of visual parts. *Cognition*, 63(1) :29–78.
- [Holt et al., 1997] Holt, B., Weiss, K., Niblack, W., Flickner, M., and Petkovic, D. (1997). The QBIC Project in the Department of Art and Art History at UC Davis. *Proceedings of the ASIS Annual Meeting*, 34 :189–195.
- [Horaud et al., 1990] Horaud, R., Veillon, F., and Skordas, T. (1990). Finding geometric and relational structures in an image. *European Conference on Computer Vision*, pages 374–384.
- [Huang et al., 1997] Huang, J., Kumar, S., Mitra, M., Zhu, W., and Zabih, R. (1997). Image indexing using color correlograms. *CVPR*, page 762.
- [Huang et al., 1999] Huang, J., Kumar, S. R., Mitra, M., Zhu, W. J., and Zabih, R. (1999). Spatial color indexing and applications. *International Journal of Computer Vision*, 35(3) :245–268.
- [Hubel and Wiesel, 1962] Hubel, D. and Wiesel, T. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, 160(1) :106–154.
- [Hueckel, 1971] Hueckel, M. (1971). An operator which locates edges in digitized pictures. *J. Ass. Comp. Mach.*, 18 :113–125.
- [Hurtut and Cheriet, 2007] Hurtut, T. and Cheriet, F. (2007). Automatic closed edge detection using level lines selection. In *International Conference on Image Analysis and Recognition*, Montreal Canada.
- [Hurtut et al., 2006a] Hurtut, T., Dalazoana, H., Gousseau, Y., and Schmitt, F. (2006a). Spatial color image retrieval without segmentation using thumbnails and the earth mover's distance. In *Color in Graphics, Imaging and Vision, CGIV 06*.
- [Hurtut et al., 2006b] Hurtut, T., Dalazoana, H., Gousseau, Y., Schmitt, F., and Cheriet, F. (2006b). Recherche automatique dans une base de données d'enluminures selon l'organisation spatiale des couleurs. *Le Médiéviste et l'ordinateur*, <http://lemo.irht.cnrs.fr/45/organisation-spatiale-couleurs.htm>.
- [Hurtut et al., 2008] Hurtut, T., Gousseau, Y., and Schmitt, F. (2008). Adaptive image retrieval based on the spatial organization of colors. *Computer Vision and Image Understanding*, *Accepted*.
- [Hurtut, T. and Cheriet, F. and Chronopoulos, M., 2007] Hurtut, T. and Cheriet, F. and Chronopoulos, M. (2007). A novel collaborative website and artworks database management system for artist-run centres. In *ICHIM 2007*.
- [Icoglu et al., 2004] Icoglu, O., Gunsel, B., and Sariel, S. (2004). Classification and indexing of paintings based on art movements. *Proceedings of the EURASIP conference*.
- [Itten, 1961] Itten, P. (1961). *Kunst der Farbe*. Verlag, Ravensburg.

- [Ji and Haralick, 1998] Ji, Q. and Haralick, R. (1998). Breakpoint detection using covariance propagation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8) :845–851.
- [Jiang et al., 2006] Jiang, S., Huang, Q., Ye, Q., and Gao, W. (2006). An effective method to detect and categorize digitized traditional Chinese paintings. *Pattern Recognition Letters*, 27(7) :734–746.
- [Jing et al., 2004] Jing, F., Li, M., Zhang, H., and Zhang, B. (2004). An efficient and effective region-based image retrieval framework. *Image Processing, IEEE Transactions on*, 13(5) :699–709.
- [Jodoin et al., 2002] Jodoin, P.-M., Epstein, E., Granger-Piché, M., and Ostromoukhov, V. (2002). Hatching by example : a statistical approach. In *NPAR*, pages 29–36.
- [Jones-Smith and Mathur, 2006] Jones-Smith, K. and Mathur, H. (2006). Fractal Analysis : revisiting Pollock’s drip paintings. *Nature*, 444(7119) :E9–10.
- [Julesz, 1986] Julesz, B. (1986). Texton gradients : The texton theory revisited. *Biological Cybernetics*, 54(4) :245–251.
- [Kammerer et al., 2003a] Kammerer, P., Langs, G., Sablatnig, R., and Zolda, E. (2003a). Stroke boundary analysis for identification of drawing tools. In Sanfeliu, A. and Ruiz-Shulcloper, J., editors, *Proc. of 8th Iberoamerican Congress on Pattern Recognition*, pages 408–415, Havana, Cuba.
- [Kammerer et al., 2007] Kammerer, P., Lettner, M., Zolda, E., and Sablatnig, R. (2007). Identification of drawing tools by classification of textural and boundary features of strokes. *Pattern Recognition Letters*, 28(6) :710–718.
- [Kammerer et al., 2003b] Kammerer, P., Zolda, E., and Sablatnig, R. (2003b). Computer aided analysis of underdrawings in infrared reflectograms. In Arnold, D., Chalmers, A., and Niccolucci, F., editors, *Proc. of 4th International Symposium on Virtual Reality, Archaeology and Intelligent Cultural Heritage*, pages 19–27, Brighton, United Kingdom.
- [Kampel and Sablatnig, 2007] Kampel, M. and Sablatnig, R. (2007). Rule based system for archaeological pottery classification. *Pattern Recognition Letters*, 28(6) :740–747.
- [Kandinsky, 1975] Kandinsky, W. (1975). *Cours du Bauhaus*. Paris : Denoel Gonthier.
- [Kandinsky et al., 1991] Kandinsky, W., Sers, P., and Leppien, S. (1991). *Point et ligne sur plan : contribution à l’analyse des éléments de la peinture*. Gallimard.
- [Kanizsa, 1996] Kanizsa, G. (1996). *La grammaire du Voir*. Diderot.
- [Kass et al., 1988] Kass, M., Witkin, A., and Terzopoulos, D. (1988). Active contour models. *Int. J. Comp. Vision*, 1 :321–331.
- [Kaupipinen et al., 1995] Kaupipinen, H., Seppanen, T., and Pietikainen, M. (1995). An experimental comparison of autoregressive and fourier-based descriptors in 2d shape classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 17(2) :201–207.
- [Kelly, 2005] Kelly, E. (2005). *Drawn from Nature*. Grand Rapids Art Museum, Yale University Press.

- [Keren, 2002] Keren, D. (2002). Painter Identification Using Local Features and Naive Bayes. *Proc. of the 16 th International Conference on Pattern Recognition*.
- [Keren, 2003] Keren, D. (2003). Recognizing image style and activities in video using local features and naive bayes. *Pattern. Recogn. Letters*, 24 :2913–2922.
- [Kirsh and Levenson, 2000] Kirsh, A. and Levenson, R. (2000). *Seeing through paintings : physical examination in art historical studies*. Yale University Press.
- [Klee, 1964] Klee, P. (1964). *Théorie de l’art moderne*. Gonthier.
- [Klee, 2004] Klee, P. (2004). *Paul Klee, cours du Bauhaus*. Hazan.
- [Kobayasi and Muroya, 2003] Kobayasi, M. and Muroya, T. (2003). A spatial wave-length analysis of coarseness or fineness of color variation in painting arts. *Pattern. Recogn. Letters*, 24 :1737–1749.
- [Köhler, 1967] Köhler, W. (1967). Gestalt psychology. *Psychological Research*, 31(1).
- [Kopf et al., 2005] Kopf, S., Haenselmann, T., and Effelsberg, W. (2005). Enhancing curvature scale space features for robust shape classification. *ICME 2005*.
- [Kristjánsson and Tse, 2001] Kristjánsson, Á. and Tse, P. (2001). Curvature discontinuities are cues for rapid shape analysis. *Perception & Psychophysics*, 63(3) :390–403.
- [Kroner and Lattner, 1998] Kroner, S. and Lattner, A. (1998). Authentication of free hand drawings by pattern recognition methods. *Pattern Recognition, 1998. Proceedings. Fourteenth International Conference on*, 1.
- [Kuhn, 1955] Kuhn, H. W. (1955). The hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2 :83–97.
- [Kushki et al., 2004] Kushki, A., Androustos, P., Plataniotis, K., and Venetsanopoulos, A. N. (2004). Retrieval of images from artistic repositories using a decision fusion framework. *IEEE Trans. on Image Processing*, 13(3) :277–292.
- [Lam et al., 1992] Lam, L., Lee, S., and Suen, C. (1992). Thinning Methodologies-A Comprehensive Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(9) :869–885.
- [Lay and Guan, 2004] Lay, J. A. and Guan, L. (2004). Retrieval for color artistry concepts. *IEEE Trans. on Image Processing*, 13(3) :326–339.
- [Lee et al., 2001] Lee, A., Mumford, D., and Huang, J. (2001). Occlusion models for natural images : A statistical study of a scale invariant dead leaves model. *International Journal of Computer Vision*, 41 :35–59.
- [Lee et al., 1991] Lee, S., Lam, L., and Suen, C. (1991). Performance Evaluation of Skeletonization Algorithms for Document Image Processing. *Proc. 1 st Int. Conf. on Document Analysis and Recognition*, 1 :260–271.
- [Lettner et al., 2004] Lettner, M., Kammerer, P., and Sablatnig, R. (2004). Texture analysis of painted strokes. In Burger, W. and Scharinger, J., editors, *Digital Imaging in Media and Education, Proc. of the 28th Workshop of the Austrian Association for Pattern Recognition (OAGM/AAPR)*, volume 179, pages 269–276. Schriftenreihe der OCG.

- [Lettner and Sablatnig, 2005] Lettner, M. and Sablatnig, R. (2005). Texture based drawing tool classification in infrared reflectograms. In Hanbury, A. and Bischof, H., editors, *Proc. of 10th Computer Vision Winter Workshop*, pages 63–72, Zell an der Pram, Austria.
- [Levina and Bickel, 2001] Levina, E. and Bickel, P. (2001). The earth mover’s distance is the Mallows distance : Some insights from statistics. *Proc. ICCV*, 2 :251–256.
- [Lew, 2000] Lew, M. (2000). Next-generation Web searches for visual content. *Computer*, 33(11) :46–53.
- [Lew et al., 2006] Lew, M., Sebe, N., Djeraba, C., and Jain, R. (2006). Content-based multimedia information retrieval : State of the art and challenges. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP)*, 2(1) :1–19.
- [Lewis et al., 2004] Lewis, P. H., Martinez, K., Abas, F. S., and et al. (2004). An integrated content and metadata based retrieval system for art. *IEEE Trans. on Image Processing*, 13(3) :302–313.
- [Leyton, 2006] Leyton, M. (2006). *The structure of paintings*. SpringerWienNewYork.
- [Li and Wang, 2004] Li, J. and Wang, Z. (2004). Studying digital imagery of ancient paintings by mixtures of stochastic models. *IEEE Trans. on Image Processing*, 13 :340–353.
- [Lindenbaum, 1997] Lindenbaum, M. (1997). An integrated model for evaluating the amount of data required for reliable recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 19(11) :1251–1264.
- [Lindgren, 1965] Lindgren, N. (1965). Machine recognition of human language : Part III-Cursive script recognition. *IEEE Spectrum*, 2(5) :104–116.
- [Ling and Okada, 2007] Ling, H. and Okada, K. (2007). An efficient earth mover’s distance algorithm for robust histogram comparison. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 29(5) :840–853.
- [Lipson et al., 1997] Lipson, P., Grimson, E., and Sinha, P. (1997). Configuration based scene classification and image indexing. *Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR’97)*, page 1007.
- [Lisani et al., 2003] Lisani, J.-L., Moisan, L., Monasse, P., and Morel, J.-M. (2003). On the theory of planar shape. *SIAM Journal on Multiscale Modeling and Simulation*, 1(1) :1–24.
- [Liu and Srinath, 1990] Liu, H. and Srinath, D. (1990). Partial shape classification using contour matching in distance transformation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 11(12) :1072–1079.
- [Liu et al., 2007] Liu, Y., Zhang, D., Lu, G., and Ma, W. (2007). A survey of content-based image retrieval with high-level semantics. *Pattern Recognition*, 40(1) :262–282.
- [Liu et al., 2005] Liu, Y., Zhang, D., Lu, G., and Ma, W.-Y. (2005). Region-based image retrieval with high-level semantic color names. *IEEE Int. Multimedia Modelling Conference*, pages 180–187.

- [Lombardi, 2005] Lombardi (2005). *The classification of style in fine-art painting*. PhD thesis, Pace University.
- [Lombardi et al., 2004] Lombardi, T., Cha, S., and Tappert, C. (2004). A graphical user interface for a fine-art painting image retrieval system. *Proceedings of the 6th ACM SIGMM international workshop on Multimedia information retrieval*, pages 107–112.
- [Lombardi et al., 2005] Lombardi, T., Cha, S., Tappert, C., et al. (2005). A lightweight image retrieval system for paintings. *Proc. SPIE*, 5682 :236–246.
- [Lowe, 1985] Lowe, D. (1985). *Perceptual Organization and Visual Recognition*. Kluwer Academic Publishers Norwell, MA, USA.
- [Lowe, 2004] Lowe, D. (2004). Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2) :91–110.
- [Lucie-Smith, 1999] Lucie-Smith, E. (1999). *Les mouvements artistiques depuis 1945*. Thames & Hudson.
- [Luo et al., 2007] Luo, B., Aujol, J.-F., Gousseau, Y., and Maître, H. (2007). Cartographie des échelle d’une image à partir de la carte topographique. applications aux images satellitaires. In *GRETSI 2007*, Troyes France.
- [Lyu et al., 2004] Lyu, S., Rockmore, D., and Farid, H. (2004). A digital technique for art authentication. *Proceedings of the National Academy of Sciences*, 101(49) :17006–17010.
- [Ma and Manjunath, 1999] Ma, W. and Manjunath, B. (1999). NeTra : A toolbox for navigating large image databases. *Multimedia Systems*, 7(3) :184–198.
- [Ma and Manjunath, 1997] Ma, W. Y. and Manjunath, B. (1997). Edge-flow : a framework fo boundary detection and image segmentation. *CVPR*, pages 744–749.
- [Mandelbrot, 1982] Mandelbrot, B. (1982). *The Fractal Geometry of Nature [M]*. Freeman.
- [Marji and Siy, 2003] Marji, M. and Siy, P. (2003). A new algorithm for dominant points detection and polygonization of digital curves. *Pattern Recognition*, 36(10) :2239–2251.
- [Marr, 1982] Marr, D. (1982). *Vision*. Freeman New York.
- [Marr and Hildreth, 1980] Marr, D. and Hildreth, E. (1980). Thoery of edge detection. *Proc. Rpyal Soc. London*, 207-B :187–217.
- [Martelli, 1972] Martelli, A. (1972). Edge detection using heuristic search methods. *Comp. Graph. and Image Proc.*, 1 :169–182.
- [Matheron, 1968] Matheron, G. (1968). Modèle séquentiel de partition aléatoire. Technical report, CMM.
- [Mattison, 2004] Mattison, D. (2004). Looking for good art image retrieval. *Searcher*, 12(9) :8–19.
- [Medioni and Yasumoto, 1986] Medioni, G. and Yasumoto, Y. (1986). Corner detection and curve representation using cubic B-splines. *Robotics and Automation. Proceedings. 1986 IEEE International Conference on*, 3.
- [Moholy-Nagy, 1955] Moholy-Nagy, L. (1955). *The new vision and Abstract of an artist*. Wittenborn.

- [Moisan, 1998] Moisan, L. (1998). Affine plane curve evolution : a fully consistent scheme. *IEEE Trans. Image Processing*, 7(3) :411–420.
- [Mokhtarian, 1995] Mokhtarian, F. (1995). Silhouette-based isolated object recognition through curvature scale space. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(5) :539–544.
- [Mokhtarian and Mackworth, 1986] Mokhtarian, F. and Mackworth, A. (1986). Scale-based description and recognition of planar curves and two-dimensional shapes. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 8 :2–14.
- [Mokhtarian and Mackworth, 1992] Mokhtarian, F. and Mackworth, A. (1992). A theory of multiscale, curvature-based shape representation for planar curves. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 14 :789–805.
- [Mokhtarian and Suomela, 1998] Mokhtarian, F. and Suomela, R. (1998). Robust image corner detection through curvature scale space. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(12) :1376–1381.
- [Monasse, 2000] Monasse, P. (2000). Representation morphologique d’images numériques et application au recalage d’images. *These de doctorat, Univ. Paris Dauphine*.
- [Monasse and Guichard, 2000] Monasse, P. and Guichard, F. (2000). Fast computation of a contrast-invariant image representation. *Image Processing, IEEE Transactions on*, 9(5) :860–872.
- [Montanari, 1971] Montanari, U. (1971). On the optimal detection of curves in noisy images. *Com. ACM*, 14 :335–345.
- [Mukherjea et al., 1999] Mukherjea, S., Hirata, K., and Hara, Y. (1999). AMORE : A World Wide Web image retrieval engine. *World Wide Web*, 2(3) :115–132.
- [Mullot, 2006] Mullot, R. (2006). *Les documents écrits, de la numérisation à l’indexation par le contenu*. Lavoisier.
- [Mureika et al., 2005] Mureika, J., Cupchik, G., and Dyer, C. (2005). Multifractal Fingerprints in the Visual Arts. *Arxiv preprint physics/0505117*.
- [Musé et al., 2003] Musé, P., Sur, F., Cao, F., and Gousseau, Y. (2003). Unsupervised thresholds for shape matching. In *Image Processing, 2003. Proceedings. 2003 International Conference on*, volume 2, pages 647–650.
- [Musé et al., 2006] Musé, P., Sur, F., Cao, F., Gousseau, Y., and Morel, J.-M. (2006). An *a contrario* decision method for shape element recognition. *International Journal of Computer Vision*, 69(3) :295–315.
- [Nack et al., 2002] Nack, F., Windhouwer, M., Hardman, L., and Pauwels, E. (2002). The role of high-level and low-level features in style-based retrieval and generation of multimedia presentations. *The New Review of Hypermedia and Multimedia*, 7 :39–65.
- [Natsev et al., 2004] Natsev, A., Rastogi, R., and Shim, K. (2004). WALRUS : a similarity retrieval algorithm for image databases. *Knowledge and Data Engineering, IEEE Transactions on*, 16(3) :301–316.
- [Neumann and Teisseron, 2002] Neumann, R. and Teisseron, G. (2002). Extraction of dominant points by estimation of the contour fluctuations. *Pattern Recognition*, 35(7) :1447–1462.

- [Ogier, 2006] Ogier, J. (2006). Madonne : Document Image Analysis Techniques for Cultural Heritage Documents. *International Conference on Digital Cultural Heritage*.
- [Olson, 1998] Olson, C. (1998). Improving the generalized Hough transform through imperfect grouping. *Image and Vision Computing*, 16(9-10) :627–634.
- [Onkarappa and Guru, 2007] Onkarappa, N. and Guru, D. (2007). Modified 9DLT Matrix for Similarity Retrieval of Line-Drawing Images. *Proceedings PReMI*, pages 136–143.
- [Osher and Sethian, 1988] Osher, S. and Sethian, J. (1988). Front propagating with curvature dependant speed : algorithm based on the hamilton-jacobi formulations. *Journal of Computational Physics*, 79 :12–49.
- [Ostromoukhov, 2002] Ostromoukhov, V. (2002). Survol de techniques de rendu-photorealiste (npr). *François X. Sillion (ed.) Synthèse d’images géographiques, Editions Hermes Science*, pages 217–250.
- [Owen, 2007] Owen, P. (2007). Painting. *Encyclopædia Britannica*.
- [P. Barla, 2005] P. Barla, J. Thollot, F. (2005). Geometric clustering for line drawing simplification. *Eurographics Symposium on Rendering*.
- [Panofsky, 1976] Panofsky, E. (1976). *La perspective comme forme symbolique*. Minuit.
- [Pareti and Vincent, 2005] Pareti, R. and Vincent, N. (2005). Global Discrimination of Graphics Styles. *Proceedings of 6th IAPR International Workshop on Graphics Recognition, Hong Kong, August*.
- [Pass and Zabih, 1999] Pass, G. and Zabih, R. (1999). Comparing images using joint histograms. *Journal of Multimedia Systems*.
- [Pass et al., 1996] Pass, G., Zabih, R., and Miller, J. (1996). Comparing images using color coherence vectors. *ACM Multimedia*, pages 65–73.
- [Pasupathy and Connor, 1999] Pasupathy, A. and Connor, C. (1999). Responses to Contour Features in Macaque Area V4. *Journal of Neurophysiology*, 82(5) :2490–2502.
- [Pavlidis, 1993] Pavlidis, T. (1993). Recognition of printed text under realistic conditions. *Pattern Recognition Letters*, 14(4) :317–326.
- [Pentland et al., 2003] Pentland, A., Picard, R., and Sclaroff, S. (2003). Photobook : tools for content-based manipulation of image databases. *Proceedings of SPIE*, 2368 :37.
- [Pettier et al., 1993] Pettier, J., Camillerapp, J., and IRISA, R. (1993). Script representation by a generalized skeleton. *Document Analysis and Recognition, 1993., Proceedings of the Second International Conference on*, pages 850–853.
- [Pettier and Camillerapp, 1993] Pettier, J.-C. and Camillerapp, J. (1993). Segmentation et représentation d’images de traits. Technical Report 2066, INRIA.
- [Pham, 2005] Pham, B. (2005). Image Indexing and Retrieval for a Vietnamese Folk Paintings Gallery. *Digital Image Computing : Techniques and Applications, 2005. DICTA’05. Proceedings*, pages 506–512.
- [Pikaz and Dinstein, 1994] Pikaz, A. and Dinstein, I. (1994). Using simple decomposition for smoothing and feature pointdetection of noisy digital curves. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 16(8) :808–813.

- [Plamondon and Privitera, 1999] Plamondon, R. and Privitera, C. M. (1999). The segmentation of cursive handwriting : an approach based on off-line recovery of the motor-temporal information. *IEEE Trans. Image Processing*, 8(1) :80–91.
- [Plamondon and Srihari, 2000] Plamondon, R. and Srihari, S. (2000). On-line and off-line handwriting recognition : a comprehensive survey. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(1) :63–84.
- [Poyato et al., 2004] Poyato, Á. C., García, N. L. F., Carnicer, R. M., and Madrid-Cuevas, F. J. (2004). A method for dominant points detection and matching 2d object identification. In *International Conference on Image Analysis and Recognition (ICIAR)*, pages 424–431.
- [Prasad et al., 2004] Prasad, B. G., Biswas, K. K., and Gupta, S. K. (2004). Region-based image retrieval using intergrated color, shape, and location index. *Computer Vision and Image Understanding*, 94 :193–233.
- [Rabin et al., 2007] Rabin, J., Delon, J., and Gousseau, Y. (2007). Mise en correspondance de descripteurs éométriques locaux par méthode a contrario. *Groupe de recherche et d’études du traitement du signal et des images (GRETSI)*.
- [Rattarangsi and Chin, 1992] Rattarangsi, A. and Chin, R. (1992). Scale-based detection of corners of planar curves. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 14(4) :430–449.
- [Remi et al., 2002] Remi, C., Frelicot, C., and Courtellemont, P. (2002). Automatic analysis of the structuring of children’s drawings and writing. *Pattern Recognition*, 35(5) :1059–1069.
- [Remi Labrusse, 2002] Remi Labrusse, E. d. C. (2002). *Henri Matisse and Ellsworth Kelly - Dessins de plantes*. Gallimard, Centre Pompidou.
- [Rissanen, 1978] Rissanen, J. (1978). Modeling by shortest data description. *Automatica*, 14(5) :465–471.
- [RMN, 2006] RMN (2006). Les carnets de picasso. DVD-Rom, Réunion des Musées Nationaux.
- [Robinson, 1981] Robinson, J. (1981). Style and significance in Art History and Criticism. *Journal of Aesthetics and Art Criticism*, 40 :5–14.
- [Rubner et al., 2000] Rubner, Y., Tomasi, C., and Guibas, L. J. (2000). The earth mover’s distance as a metric for image retrieval. *International Journal of Computer Vision*, 40(2) :99–121.
- [Rugna and Konik, 2002] Rugna, J. D. and Konik, H. (2002). Color coarse segmentation and regions selection for similar images retrieval. *CGIV*, pages 241–244.
- [Salembier and Garrido, 2000] Salembier, P. and Garrido, L. (2000). Binary partition tree as an efficient representation for imageprocessing, segmentation, and information retrieval. *Image Processing, IEEE Transactions on*, 9(4) :561–576.
- [Salembier et al., 1998] Salembier, P., Oliveras, A., and Garrido, L. (1998). Antiextensive connected operators for image and sequence processing. *Image Processing, IEEE Transactions on*, 7(4) :555–570.

- [Sayre, 1989] Sayre, H. (1989). *Writing about Art*. Prentice Hall Englewood Cliffs, NJ.
- [Schettini et al., 2001] Schettini, R., Ciocca, G., and Zuffi, S. (2001). A survey on methods for colour image indexing and retrieval in image databases. *Color Imaging Science : Exploiting Digital Media*, pages 183–211.
- [Schmid et al., 2000] Schmid, C., Mohr, R., and Bauckhage, C. (2000). Evaluation of Interest Point Detectors. *International Journal of Computer Vision*, 37(2) :151–172.
- [Schomaker, 1991] Schomaker, L. (1991). *Simulation and recognition of handwriting movements : a vertical approach to modeling human motor behavior*. Nijmeegs Instituut voor Cognitie-onderzoek en Informatietechnologie.
- [Schomaker et al., 1989] Schomaker, L., Thomassen, A., and Teulings, H. (1989). A computational model of cursive handwriting. *Computer Recognition and Human Production of Handwriting*, pages 153–177.
- [Seldin et al., 2003] Seldin, Y., Starik, S., and Werman, M. (2003). Unsupervised clustering of images using their joint segmentation. *3rd International Workshop on Statistical and Computational Theories of Vision*.
- [Serra, 1982] Serra, J. (1982). *Image Analysis and Mathematical Morphology*. Academic Press.
- [Sethian, 1996] Sethian, J. (1996). *Level-set methods*. Princeton University Press.
- [Shapiro, 1994] Shapiro, M. (1994). *Style*. Theory and Philosophy of Art.
- [Shilane et al., 2004] Shilane, P., Min, P., Kazhdan, M., and Funkhouser, T. (2004). The Princeton Shape Benchmark. *Shape Modeling Applications, 2004. Proceedings*, pages 167–178.
- [Silverman, 1984] Silverman, R. (1984). *Learning about Art : A Practical Approach*. Romar Arts.
- [Simonnard, 1962] Simonnard, M. (1962). *Programmation linéaire*. Dunod.
- [Sinzing, 2008] Sinzing, E. (2008). A model-based approach to junction detection using radial energy. *Pattern Recognition*, 41(2) :494–505.
- [Smith and Chang, 1997] Smith, J. and Chang, S. (1997). VisualSEEK : a fully automated content-based image query system. *Proceedings of the fourth ACM international conference on Multimedia*, pages 87–98.
- [Smith and Li, 1999] Smith, J. R. and Li, C.-S. (1999). Image classification and querying using composite region templates. *Computer Vision and Image Understanding*, 75(1/2) :165–174.
- [Stanchev et al., 2003] Stanchev, P., Green Jr, D., and Dimitrov, B. (2003). High Level Color Similarity Retrieval. *Int. J. Inf. Theories Appl. v10 i3*, pages 363–369.
- [Stanchev et al., 2006] Stanchev, P., Green Jr, D., and Dimitrov, B. (2006). Some Issues in the Art Image Database Systems. *Journal of Digital Information Management*, 4(4) :227.
- [Stork, 2006] Stork, D. (2006). Computer vision, image analysis, and master art : part 1. *IEEE MultiMedia Magazine*, 13(3) :16–20.

- [Stricker and Dimai, 1997] Stricker, M. and Dimai, A. (1997). Spectral covariance and fuzzy regions for image indexing. *MVA*, 10(2) :66–73.
- [Stricker and Orengo, 1995] Stricker, M. and Orengo, M. (1995). Similarity of color images. *Proc. SPIE Storage and Retrieval for Image and Video Databases*, 2420 :381–392.
- [Sur, 2004] Sur, F. (2004). *A contrario decision for shape recognition*. PhD thesis, Université Paris Dauphine.
- [Swain and Ballard, 1991] Swain, M. J. and Ballard, D. H. (1991). Color indexing. *Int. J. Comput. Vision*, 7(1) :11–32.
- [Tanaka et al., 1999] Tanaka, S., Kurumisawa, J., Plante, A., Iwadate, Y., and Inokuchi, S. (1999). Composition analyzer : computer supported composition analysis of masterpieces. In *C&C '99 : Proceedings of the 3rd conference on Creativity & cognition*, pages 68–75, New York, NY, USA. ACM Press.
- [Tanaka et al., 2000] Tanaka, S., Kurumizawa, J., Inokuchi, S., and Iwadate, Y. (2000). Composition analyzer : support tool for composition analysis on painting masterpieces. *Knowledge-based systems*, 13 :459–470.
- [Taylor et al., 2007] Taylor, R., Guzman, R., Martin, T., Hall, G., Micolich, A., Jonas, D., and Marlow, C. (2007). Authenticating Pollock Paintings Using Fractal Geometry. *Pattern Recognition Letters*, 28(6) :695–702.
- [Taylor et al., 1999] Taylor, R., Micolich, A., and Jonas, D. (1999). Fractal analysis of pollock drip paintings. *Nature*, 399 :422.
- [Taylor et al., 2006] Taylor, R., Micolich, A., and Jonas, D. (2006). Fractal Analysis : Revisiting Pollock’s drip paintings (Reply). *Nature*, 444(7119) :10–11.
- [Tombre, 1997] Tombre, K. (1997). Analysis of Engineering Drawings : State of the Art and Challenges. *Lecture Notes In Computer Science*, pages 257–264.
- [Torre and Poggio, 1986] Torre, V. and Poggio, T. (1986). On edge detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 8 :147–153.
- [Tsai, 2007] Tsai, C. (2007). A review of image retrieval methods for digital cultural heritage resources. *Online Information Review*, 31(2) :185–198.
- [Tse and Cavanagh, 2000] Tse, P. and Cavanagh, P. (2000). Chinese and Americans see opposite apparent motions in a Chinese character. *Cognition*, 74(3) :27–32.
- [Tuceryan and Jain, 1993] Tuceryan, M. and Jain, A. (1993). Texture analysis. *Handbook of Pattern Recognition and Computer Vision*, pages 235–276.
- [Urbach, 2002] Urbach, E. (2002). Shape-Only Granulometries And Gray-Scale Shape Filters. *Mathematical Morphology : Proceedings of the 6th International Symposium-ISMM 2002*.
- [Urdiales et al., 2002] Urdiales, C., Bandera, A., and Sandoval, F. (2002). Non-parametric planar shape representation based on adaptative curvature functions. *Patt. Recogn.*, 35 :43–53.
- [Valle et al., 2006] Valle, E., Cord, M., and Philipp-Foliguet, S. (2006). Content-based retrieval of images for cultural institutions using local descriptors. *Geometric Modeling and Imaging, IEEE International Conference on*.

- [van den Herik and Postma, 2000] van den Herik, H. J. and Postma, E. O. (2000). Discovering the visual signature of painters. *Future Directions for Intelligent Systems and Information Sciences, Physica-Verlag*, pages 129–147.
- [van den Herik and Postma, 2004] van den Herik, H. J. and Postma, E. O. (2004). <http://www.cs.unimaas.nl/i.berezhnoy/>.
- [Van Sommers, 1984] Van Sommers, P. (1984). *Drawing and Cognition : Descriptive and Experimental Studies of Graphic Production Processes*. Cambridge University Press.
- [van Tonder, 2007] van Tonder, G. (2007). Recovery of visual structure in illustrated Japanese gardens. *Pattern Recognition Letters*, 28(6) :728–739.
- [Wang et al., 1998] Wang, J., Wiederhold, G., Firschein, O., and Xin Wei, S. (1998). Content-based image indexing and searching using Daubechies’ wavelets. *International Journal on Digital Libraries*, 1(4) :311–328.
- [Wang et al., 2001] Wang, J. Z., Li, J., and Wiederhold, G. (2001). Simplicity : semantic-sensitive integrated matching for picture libraries. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 23(9).
- [Wang et al., 1999] Wang, Y., Lee, S., and Toraichi, K. (1999). Multiscale curvature-based shape representation using B-splineswavelets. *Image Processing, IEEE Transactions on*, 8(11) :1586–1592.
- [Wei-ning et al., 2004] Wei-ning, W., Ying-lin, Y., and Jian-chao, Z. (2004). Image emotional classification : static vs. dynamic. *Systems, Man and Cybernetics, 2004 IEEE International Conference on*, 7.
- [Widjaja et al., 2003] Widjaja, I., Leow, W. K., and Wu, F.-C. (2003). Identifying painters from color profiles of skin patches in painting images. *Proc. Int. Conf. on Image Processing*.
- [Willats, 1977] Willats, J. (1977). How children learn to draw realistic pictures. *The Quarterly Journal of Experimental Psychology*, 29(3) :367–382.
- [Willats, 1997] Willats, J. (1997). *Art and representation*. Princeton University Press.
- [Willats, 2006] Willats, J. (2006). Ambiguity in drawing. *TRACEY*, <http://www.lboro.ac.uk/departments/ac/tracey/>.
- [Willats and Durand, 2005] Willats, J. and Durand, F. (2005). Defining pictorial style : Lessons from linguistics and computer graphics. *Axiomathes*, 15(2).
- [Williams and Yoon, 2007] Williams, A. and Yoon, P. (2007). Content-based image retrieval using joint correlograms. *Multimedia Tools Applications*, 34 :239–248.
- [Wilson, 1985] Wilson, H. (1985). Discrimination of contour curvature : data and theory. *Journal of the Optical Society of America A*, 2(7) :1191–1199.
- [Winter et al., 2002] Winter, J., Panis, S., and Wagemans, J. (2002). Perceptual saliency of points along the contour of everyday objects : A large-scale study. *Journal of Vision*, 2(7) :487.
- [Wolfflin, 1950] Wolfflin, H. (1950). *Principles of art history*. Dover Publications [New York].

- [Wollheim, 1993] Wollheim, R. (1993). Pictorial Style : Two Views. *The Mind and its Depths*, pages 171–84.
- [Wu, 2003] Wu, W. (2003). An adaptive method for detecting dominant points. *Pattern Recognition*, 36(10) :2231–2237.
- [Wyszecki and Stiles, 2000] Wyszecki, G. and Stiles, W. (2000). Color Science : Concepts and Methods, Quantitative Data and Formulae. *Color Science : Concepts and Methods, Quantitative Data and Formulae, 2nd Edition, by Gunther Wyszecki, WS Stiles,*.
- [Yamamoto et al., 1999] Yamamoto, H., Iwasa, H., Yokoya, N., and Takemura, H. (1999). Content-based similarity retrieval of images based on spatial color distributions. *Image Analysis and Processing, 1999. Proceedings. International Conference on*, pages 951–956.
- [Yan and Jin, 2006] Yan, Y. and Jin, J. (2006). Indexing and Retrieving Oil Paintings Using Style Information. *Lecture Notes In Computer Science*, 3736 :143.
- [Yelizaveta et al., 2004] Yelizaveta, M., Tat-Seng, C., and Aristarkhova, I. (2004). Retrieval of paintings based on concepts defined in art history. *Proc. IWAIT*.
- [Yelizaveta et al., 2005] Yelizaveta, M., Tat-Seng, C., and Irina, A. (2005). Analysis and Retrieval of Paintings Using Artistic Color Concepts. *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on*, pages 1246–1249.
- [Zhang et al., 2004] Zhang, D., Pham, B., and Li, Y. (2004). Modelling traditional Chinese paintings for content-based image classification and retrieval. *Multimedia Modeling Conference, 2004. Proceedings. 10th International*, pages 258–264.
- [Zhang et al., 2007] Zhang, X., Lei, M., Yang, D., Wang, Y., and Ma, L. (2007). Multi-scale curvature product for robust image corner detection in curvature scale space. *Pattern Recognition Letters*, 28(5) :545–554.
- [Zipf, 1949] Zipf, G. (1949). *Human behavior and the principle of least effort*. Addison-Wesley Press Cambridge, Mass.
- [Zusne, 1970] Zusne, L. (1970). *Visual Perception of Form*. Academic Press.

BIBLIOGRAPHIE

Publications

Journaux

T. Hurtut, Y. Gousseau and F. Schmitt.

Adaptive image retrieval based on the spatial organization of colors,
Computer Vision and Image Understanding (CVIU).
(Accepted, in press)

T. Hurtut, Y. Gousseau, F. Cheriet and F. Schmitt.

Pictorial effect analysis of line-drawings using geometrical shape information,
in preparation

Conférences avec actes et comités de lectures

T. Hurtut, F. Cheriet, M. Chronopoulos.

A novel collaborative website and artworks database management system for artist-run centres,
International Conference on Hypermedia and Interactivity in Museums (ICHIM'07), October 2007.

T. Hurtut, F. Cheriet.

Automatic closed edge detection using level lines selection,
International Conference on Image Analysis and Recognition (ICIAR'07), August 2007.

T. Hurtut, H. Dalazoana, Y. Gousseau and F. Schmitt.

Spatial color image retrieval without segmentation using thumbnails and the Earth Mover's Distance,
Color in Graphics, Imaging and Vision (CGIV'06), Leeds IS&T's Conference, June, 2006.

Communications

T. Hurtut, H. Dalazoana, Y. Gousseau, F. Schmitt and F. Cheriet.

Recherche automatique dans une base de données d'enluminures selon l'organisation spatiale des couleurs,

Le médiéviste et l'ordinateur, revue en ligne de l'IRHT, Mai 2005.

T. Hurtut, H. Dalazoana, Y. Gousseau, F. Schmitt and F. Cheriet.

Recherche automatique dans une base de données d'enluminures selon l'organisation spatiale des couleurs,

Signal Image et Arts, organisée par la Section Signal et Image du club EEA, 2 Juin 2006.