Contributions pour la Recherche d'Images par Composantes Visuelles

Julien Fauqueur

Thèse

présentée par **Julien FAUQUEUR**

pour obtenir le grade de docteur de l'Université de Versailles
Saint-Quentin en Yvelines
spécialité : INFORMATIQUE

Contributions pour la Recherche d'Images par Composantes Visuelles

---

Soutenue le 21 novembre 2003 devant la Commission d'examen composée de :

Del Bimbo	Alberto	Rapporteur	Professeur Université de Florence, Italie
Berrut	Catherine	Rapporteur	Professeur Université Joseph Fourier, Grenoble
Joly	Philippe	Examinateur	Maitre de Conférence Université de Toulouse
Boujemaa	Nozha	Directeur de thèse	Directrice de Recherche INRIA Rocquencourt
Oria	Vincent	Examinateur	Maitre de Conférence New Jersey Institute of Technology, USA
Bouzeghoub	Mokrane	Examinateur	Professeur Université de Versailles Saint Quentin

Thèse préparée à l'INRIA-Rocquencourt, projet IMEDIA
Julien.Fauqueur@inria.fr - http://www-rocq.inria.fr/~fauqueur/

Note sur la navigation dans ce document électronique

Afin de profiter du confort offert par le format PDF pour la navigation dans la forme électronique de ce document, nous suggérons à l'utilisateur l'usage des liens internes et de l'historique si le logiciel Acrobat Reader est utilisé.

Les références aux citations bibliographiques, aux sections et chapitres, aux figures et tableaux sont *interactives* : un clic de souris sur une référence conduit directement à la partie concernée du document. La commande CTRL + ← (sous linux et ALT+← sous windows) permet de revenir au lieu de la citation. Plus généralement, pour reculer ou avancer dans l'historique de la navigation du document, les commandes suivantes peuvent être invoquées :

Sous Linux :
– CTRL + ← : reculer dans l'historique
– CTRL + → : avancer dans l'historique
Sous Windows :
– ALT + ← : reculer dans l'historique
– ALT + → : avancer dans l'historique
N'hésitez pas à utiliser ces raccourcis pour un meilleur confort de lecture.

Remerciements

Je remercie tout particulièrement Nozha Boujemaa qui m'a encadré tout au long de ma thèse. J'ai particulièrement été sensible à la cohérence et la pertinence dans le choix des directions scientifiques, tout en laissant une importante place à l'initiative personnelle. Elle a su aussi me faire partager ses reflexions sur les tendances de l'état de l'art et particulièrement sur les incontournables questions d'usage en recherche d'images par le contenu. Je lui suis sincèrement reconnaissant d'avoir fait que mes années de thèse soient passionnantes.

Ma rencontre avec Georges Stamon lors de mon DEA a été déterminante dans mon choix de poursuivre en thèse dans le domaine de l'image et de la reconnaissance des formes. Je le remercie vivement pour m'avoir ouvert l'esprit à de nombreuses méthodes et applications dans ces domaines. Puis, au long de ma thèse, ses conseils et son support ont été très précieux. Par ailleurs, j'ai beaucoup apprécié sa grande générosité et sa disponibilité.

L'ambience dans l'équipe Imedia a toujours été stimulante, amicale et très positive. J'en remercie dans le désordre chacun des membres : Laurence, Jean-Paul, François, Bertrand, Valérie, Vincent, Jean-Philippe (notamment pour la relecture), Anne, Nizar, Hichem 1 et 2, Minel, Donald, Sabri, Michel, Yuchun, Andreas et aussi Chahab, Alexandre, Sebastien et Nathalie qui m'ont aussi beaucoup apporté au début de ma thèse puis sont partis.

Je remercie vivement les membres du jury de s'être penchés en détail sur mes travaux de thèse. A travers leurs rapports, les discussions et la soutenance, la qualité et la richesse de leurs nombreuses remarques m'ont beaucoup apporté et permis d'améliorer la version finale de ce document.

Je tiens à remercier certains des relecteurs anonymes de nos publications dont les critiques ont été très riches.

Spéciale dédicace à Miss Aicha, DJ Alexis et DJ Jaco pour avoir assuré les coffee mix, les ballades langoureuses dans les prairies cheynaisiennes, les anneaux à la fleur d'oranger et les bombardements de la base de l'OTAN.

Une grande pensée chaleureuse pour ma précieuse famille, mes amis et aussi mes différents collocataires.

Certaines parties de ma thèse sont fortement associées aux lieux où je les ai

rédigées : le cabanon des Jonchiers, la véranda de Champlong et le jardin de Bidouric.

Cette thèse est dédiée à mon cher Arman dont je suis le parain, malgré lui.

Table des matières

1 Introduction : la recherche par le contenu visuel 11
- 1.1 Principe de la recherche d'images par le contenu . . . 12
- 1.2 Comment identifier une composante visuelle d'image ? . . . 14
- 1.3 Approches existantes en recherche par régions . . . 18
- 1.4 Nos contributions . . . 19
- 1.5 Plan du mémoire . . . 20

2 Mesures de similarité et regroupement de primitives visuelles 23
- 2.1 Distances pour les distributions de couleurs . . . 24
  - 2.1.1 Distances cellule-à-cellule ou "daltoniennes" . . . 24
  - 2.1.2 Distances inter-cellules . . . 26
- 2.2 Positionnement du problème de regroupement . . . 31
- 2.3 Algorithmes de type "k-means" . . . 32
  - 2.3.1 Le seuillage, degré 0 de la quantification . . . 32
  - 2.3.2 k-means / GLA . . . 32
  - 2.3.3 C-Moyennes Floues . . . 34
  - 2.3.4 Expectation/Maximization . . . 36
  - 2.3.5 Limitations de ces approches . . . 38
- 2.4 L'algorithme CA . . . 39
  - 2.4.1 Principe de l'algorithme . . . 40
  - 2.4.2 Raffinement pour une mise en oeuvre effective . . . 43
  - 2.4.3 Perspectives . . . 46
- 2.5 Conclusions . . . 46

3 Détection des composantes visuelles par segmentation en régions 49
- 3.1 Aperçu des approches existantes . . . 49
- 3.2 Notre approche . . . 52
- 3.3 Extraction de primitives locales LDQC . . . 54
- 3.4 Regroupement des primitives LDQC . . . 59
- 3.5 Consolidation spatiale . . . 61
- 3.6 Résultats . . . 61

3.7 Perspectives 63
3.8 Conclusions 65

4 Description fine de régions pour le Paradigme 1 67
4.1 Le paradigme 1 68
4.2 L'existant pour la description de l'apparence visuelle 69
4.3 ADCS : description fine, adaptative et compacte de la variabilité couleur 73
4.4 Forme généralisée de la distance quadratique 75
4.5 Intégration de la similarité géométrique 80
4.6 Interface utilisateur 83
4.7 Résultats 85
4.8 Discussion : points ou régions comme description locale ? 98
4.9 Perspectives 101
4.10 Conclusions 102

5 Paradigme 2 : recherche d'images par composition logique de catégories de régions 105
5.1 Introduction 106
5.2 Résumé et déroulement de l'approche 107
5.3 Catégorisation de régions pour la recherche de similarité 109
5.3.1 Représentation des régions dans l'espace de description . . 109
5.3.2 Catégorisation de régions et catégories voisines 110
5.4 Recherche d'images-cible par composition de catégories 114
5.4.1 Principe 114
5.4.2 Implantation algorithmique 119
5.4.3 Indexation symbolique 121
5.5 Interface Utilisateur et résultats 122
5.5.1 Interaction utilisateur 122
5.5.2 Résultats 124
5.5.3 Application à une photothèque 125
5.5.4 Application au journal télévisé 132
5.5.5 Raffinement de requête - interaction sur les résultats . . . 137
5.5.6 Coût de la recherche 138
5.6 Discussions 138
5.7 Perspectives 141
5.7.1 Description de régions 142
5.7.2 Recherche dans de très grandes bases 143
5.7.3 Interaction système-utilisateur 145
5.7.4 Perspectives en recherche d'information visuelle 146
5.8 Conclusions 146

6 Perspectives **149**

7 Résumé des contributions et Conclusions **151**

A Codage d'une image segmentée **153**

B Résultats de segmentation grossière et description fine **159**

Bibliographie **169**

Résumé

Dans le contexte de la recherche d'information par le contenu visuel, lorsque l'utilisateur formule une requête visuelle, sa cible de recherche est rarement représentée par une image entière comme le suppose le paradigme classique de recherche par une image exemple. L'image ne doit pas être traitée comme une unité atomique, car elle est généralement constituée d'un ensemble composite de zones visuelles exprimant une certaine sémantique.

Un système de recherche d'information visuelle doit permettre à l'utilisateur de désigner d'une manière explicite la cible visuelle qu'il recherche se rapportant aux différentes composantes de l'image. Notre objectif au cours de ce travail a été de réfléchir à comment définir des clés de recherche visuelle permettant à l'utilisateur d'exprimer cette cible visuelle, de concevoir et d'implémenter efficacement les méthodes correspondantes.

Les contributions originales de cette thèse portent sur de nouvelles approches permettant de retrouver des images à partir de leurs différentes composantes visuelles selon deux paradigmes de recherche distincts.

Le premier paradigme est celui de la recherche par région exemple. Il consiste à retrouver les images comportant une partie d'image similaire à une partie visuelle requête. Pour ce paradigme, nous avons mis au point une approche de segmentation grossière en régions et de description fine de ces régions ensuite. Les régions grossières des images de la base, extraites par notre nouvel algorithme de segmentation non supervisée, représentent les composantes visuellement saillantes de chaque image. Cette décomposition permet à l'utilisateur de désigner séparément une région d'intérêt pour sa requête. La recherche de régions similaires dans les images de la base repose sur un nouveau descripteur de régions (ADCS). Il offre une caractérisation fine, compacte et adaptative de l'apparence photométrique des régions, afin de tenir compte de la spécificité d'une base de descripteurs de régions. Dans cette nouvelle approche, la segmentation est rapide et les régions extraites sont intuitives pour l'utilisateur. La finesse de description des régions améliore la similarité des régions retournées par rapport aux descripteurs existants, compte tenu de la fidélité accrue au contenu des régions.

Notre seconde contribution porte sur l'élaboration d'un nouveau paradigme de recherche d'images par composition logique de catégories de régions. Ce paradigme présente l'avantage d'apporter une solution au problème de la page zéro. Il permet d'atteindre les images, quand elles existent dans la base, qui se rapprochent de la représentation mentale de la cible visuelle de l'utilisateur. Ainsi aucune image ou région exemple n'est nécessaire au moment de la formulation de la requête. Ce paradigme repose sur la génération non-supervisée d'un thésaurus photométrique constitué par le résumé visuel des régions de la base. Pour formuler sa requête, l'utilisateur accède directement à ce résumé en disposant d'opérateurs

de composition logique de ces différentes parties visuelles. Un item visuel dans ce résumé est un représentant d'une classe photométrique de régions. Les requêtes logiques sur le contenu des images s'apparentent à celles en recherche de texte. L'originalité de ce paradigme ouvre des perspectives riches pour de futurs travaux en recherche d'information visuelle.

Abstract

In the context of information retrieval by visual content, when the user formulates a visual query, his/her query target is rarely represented by a whole image as assumed in the usual paradigm of query by image example. An image should not be considered as an atomic entity since it is generally formed of a composite set of visual parts which express certain semantics.

A visual information retrieval system should allow the user to explicitly point out the visual target using the various image components. In our work the goal was to investigate methods to define visual search keys which allow the user to express this visual target, and to design and efficiently implement these methods.

The original contributions proposed in this thesis are new approaches which allow the retrieval of images from their various visual components using two distinct query paradigms.

The first paradigm is the query by region example. It consists in retrieving images containing an image part similar to a query visual part. For this paradigm we have designed an approach of coarse segmentation into regions followed by a fine description of these regions. Coarse regions, extracted by our new unsupervised segmentation algorithm from images in the database, represent visually salient components in each image. This decomposition allows the user to separately point out a region of interest for his/her query. Query by similar regions in the image database relies on a new region descriptor (ADCS). It provides a fine, compact and adaptive characterization of region photometric appearance, in order to take into account the specificity of a database of region descriptors. In this new approach, segmentation is fast and extracted regions are intuitive for the user. Fine description improves the similarity of retrieved regions compared to existing descriptors, thanks to the increased accuracy of region content description.

Our second contribution concerns the development of a new image query paradigm by logical composition of region categories. This paradigm has the advantage of providing a solution to the "page zero" problem. It allows the user to attain images, if they exist in the database, which are close to the mental representation of the user visual target. No image nor region example is necessary to formulate the query. This paradigm relies on the unsupervised generation of a region photometric thesaurus constituted by the visual summary of regions in the database. To formulate a query the user can access this summary directly by means of logical composition operators on these different visual parts. Note that a visual item in this summary is a representative of a photometric class of regions. Logical queries on image content relate to those in text retrieval. The originality of this paradigm opens rich perspectives for future work in visual information retrieval.

Chapitre 1

Introduction : la recherche par le contenu visuel

Les bases d'images numériques connaissent un essor considérable depuis quelques années. Leur facilité d'acquisition et de stockage les rendent très attractives pour des applications très diverses. Cet essor est directement lié à celui de l'évolution technologique (capteurs numériques, medium de stockage de données, microprocesseurs, périphériques d'affichage) et concerne aussi bien les milieux professionnels les plus consommateurs d'images que le grand public[1]. Il en résulte une production permanente et considérable d'images numériques dans des domaines tels que l'imagerie satellitaire, la santé, l'illustration, l'audiovisuel, l'architecture, la botanique, la télésurveillance, la photographie grand public.

L'accumulation d'images numériques pose rapidement, dès quelques centaines d'unités, le problème de la recherche d'images. L'approche la plus ancienne (antérieure à l'apparition des images numériques) et encore majoritairement employée aujourd'hui est l'annotation de meta-données telles que les mots-clés, le titre, l'auteur, les conditions de prise de vue et des informations variées dépendant du domaine considéré. Dans la limite de l'information que peut porter l'annotation, ce type d'indexation est conçu pour répondre à des types de requêtes spécifiques et prédéfinis. Les inconvénients majeurs de l'annotation sont : la nécessité de l'intervention d'un humain (pénible sur de grandes bases), leur rigidité (l'ajout ou la suppression des champs de meta-données sur une base entière représente un travail colossal), leur subjectivité (deux personnes annoteront-elles une image donnée avec les même mots-clés ?), les contraintes linguistiques (passage d'une langue à une autre, ambiguïté sémantique). De plus, notons que l'annotation ne

[1] L'enquête [3] menée par l'association PMA reporte un taux de pénétration de 20% des appareils photographiques numériques dans les foyers américains en fin 2002. Le nombre de foyers en possédant un appareil numérique passe de 15 millions en 2001 à 23 millions en 2002 et 33 millions en 2003 (prévision).

pourra jamais décrire le contenu d'une image de façon *exhaustive*.

En même temps que l'essor récent des bases d'images numériques, une alternative à l'annotation manuelle est apparue il y a une dizaine d'années : la recherche d'images par le contenu (ou CBIR en anglais pour content-based image retrieval) [110][22][102]. Elle consiste à caractériser le *contenu visuel* des images par des *descripteurs visuels* et d'effectuer des recherches par similarité visuelle à partir de ces descripteurs. Alors que l'annotation d'images conduit à une indexation de nature sémantique, les descripteurs visuels sont "homogènes" au medium qu'ils décrivent, c'est-à-dire de nature visuelle. La première conséquence majeure est qu'il devient possible d'interroger une base d'images directement à partir de leur contenu visuel. La seconde conséquence est que l'indexation est automatique et donc répétable car déterministe.

Cette nouvelle approche permet de répondre à de nouveaux besoins dans le domaine de la recherche dans les bases d'images.

Quelle est alors la pertinence des images retournées en l'absence d'une sémantique explicite dans le processus d'indexation automatique ? Si l'on utilise un descripteur visuel peu informatif ou bien peu fidèle, les images retournées ne présenteront aucun intérêt pour l'utilisateur. Plus généralement, la pertinence est directement liée aux techniques choisies pour les différents éléments qui composent un système de recherche par le contenu, tels que : descripteurs (couleur, texture, forme), mesure de similarité visuelle, mode de requête, mode de réprésentation des images (global, partiel). Bien que de nombreux travaux sur ces différents aspects ont été proposés dans la littérature, la recherche d'images par le contenu demeure encore aujourd'hui un problème ouvert et très actif.

Les contributions présentées dans ce mémoire portent sur la recherche d'images par leurs composantes visuelles. Elles visent à améliorer la satisfaction de l'utilisateur dans l'approche de recherche d'images par le contenu. Dans ce chapitre, nous commencerons par présenter un aperçu des approches existantes. Nous justifierons ensuite la nécessité d'une représentation partielle des images et expliquerons notre choix pour les régions d'intérêt. Puis, nous introduirons les principales contributions de nos travaux. Une partie de ces travaux a été publiée dans [29][28][8][26][30][27].

1.1 Principe de la recherche d'images par le contenu

Le scénario de recherche d'images par le contenu le plus élémentaire, et le premier historiquement, est celui de recherche globale par l'exemple. L'utilisateur choisit une image exemple et le système détermine les images de la base dont l'apparence visuelle globale est la plus similaire. Le coeur de l'approche réside sur

une description visuelle de chaque image et sur une mesure de similarité adéquate pour le descripteur. La figure 1.1 illustre un scénario de recherche globale d'images sur la plateforme IKONA[2] de l'équipe IMEDIA.

Le principe de cette approche a été proposé en 1991 par Swain et Ballard [110]. Il s'agit du principe fondateur de nombreux travaux et de systèmes tels que QBic [79], PhotoBook [85], Virage [39], MARS [48], ImageRover [101], PicToSeek [34], Compass [10], Ikona [7] pour n'en citer que quelques uns. Dans le cadre de la recherche globale, de nombreux descripteurs d'images et mesures de similarité ont été proposés afin de caractériser les informations de couleur, texture, forme.

FIG. 1.1 – Exemple de requête globale d'image avec IKONA : l'écran de gauche affiche aléatoirement des images de la base. L'utilisateur sélectionne une image exemple (le paysage maritime encadré en rouge) qui ressemble aux scènes qu'il recherche. Le système retourne les images qui présentent une apparence globale similaire à l'image exemple (écran de droite). Dans ce scénario de recherche globale, il n'est pas possible d'effectuer une recherche sur une partie d'image (un palmier par exemple).

D'un point de vue de l'interaction utilisateur, le bouclage de pertinence, issu des techniques de recherche de texte [93], a été appliqué avec succès à la recherche d'images [119][72]. Il permet à l'utilisateur de raffiner sa recherche en indiquant itérativement des exemples d'images pertinentes et non-pertinentes pour sa recherche. Quant à la navigation dans les bases d'images, elle offre une approche complémentaire au scénario de recherche : organisation bi- ou tri-dimensionnelle après une réduction de dimensionnalité par Multi-Dimensional Scaling des histogrammes couleurs [94], organisation tridimensionnelle des vignettes d'images dans l'interface interactive de visualisation [45], par catégories d'images similaires après regroupement non-supervisé [98]. Les images étant organisées par similarité selon leur descripteur visuel, l'utilisateur dispose d'un

[2]http ://www-rocq.inria.fr/imedia/ikona/

aperçu global de la base.

La limitation de ces méthodes de recherche et de navigation exploitant la description visuelle globale des images est l'hypothèse implicite que l'*intégralité* du contenu de l'image est pertinente pour les besoins de l'utilisateur. L'image est considérée comme une entité visuelle atomique, alors que l'humain perçoit généralement une *image comme une entité composite d'objets*. L'intérêt de l'utilisateur peut porter sur une ou plusieurs composantes d'images ou bien il peut souhaiter ignorer le fond d'une image qu'il ne jugera pas pertinent pour sa requête. Il est donc important de permettre à l'utilisateur de désigner explicitement les composantes pertinentes dans les images et d'effectuer la recherche à partir de celles-ci.

1.2 Comment identifier une composante visuelle d'image ?

La figure 1.2 illustre un exemple d'image composite dont les différentes composantes visuelles peuvent constituer un intérêt potentiel pour une recherche dans une base d'images.

FIG. 1.2 – Exemple d'image composite : l'utilisateur peut s'intéresser à l'apparence visuelle globale de l'image, ou individuellement à ses différentes composantes (personnage, logo, sous-titre, fond de campagne, vaches), à la composition des différentes composantes (source : INA/France3).

Les composantes visuelles d'images susceptibles de constituer des "clés de requête" pertinentes pour un utilisateur peuvent être de nature diverse : grandes zones saillantes, détails plus ou moins précis. La détection de composantes d'images dans de grandes bases est un problème difficile, particulièrement lors-

qu'elle doit être automatique. Différentes méthodes pour définir et extraire des composantes d'images ont été proposées dans la littérature pour la recherche par le contenu. Elles se différencient par les aspects suivants :

- intervention de l'utilisateur
- définition a priori ou non des composantes
- connaissance a priori ou non du type d'"objets" recherchés
- précision de la définition des composantes
- coût de calcul
- rapidité d'extraction et de recherche

Nous allons voir comment le choix de la segmentation d'image en régions se positionne vis-à-vis d'autres types existants de définition de composantes d'images.

L'intérêt pour la localisation de zones d'intérêt dans la recherche d'images est apparu en même temps que le principe de recherche d'images par le contenu en 1991 avec les travaux de Swain et Ballard [110]. En effet, en même temps qu'ils introduisirent le principe d'indexation et de recherche d'images par histogrammes couleur très largement répandu encore aujourd'hui, ils proposèrent aussi un mode de recherche localisée d'histogramme (*Histogram Backprojection*) visant à répondre à la question : *"Where in the image are the colors that belong to the object being looked for ?"*. Bien que cet algorithme fût très peu repris dans littérature, la notion de recherche d'"objet" dans les images était déjà présente.

Parmi les représentations partielles existantes pour la recherche d'images par le contenu, nous distinguons les suivantes :

Détourage manuel

Le détourage manuel des zones d'intérêt des images de la base (voir [22]) présente l'avantage de définir les régions selon les attentes de l'utilisateur. Cependant l'inconvénient majeur est la pénible intervention de l'utilisateur qui n'est pas viable pour extraire les régions dans des grandes bases d'images.

Blocs

A l'opposé en termes de coût calculatoire et de précision, la subdivision systématique d'images en blocs [66][76] est simple et non-supervisée, mais très approximative : des descripteurs visuels sont calculés sur les cellules d'un quadrillage constant de l'image. La subdivision est tributaire de l'échelle spatiale de la grille qui ne s'adapte pas au contenu de l'image. Dans des zones uniformes, les cellules seront inutilement trop nombreuses et seront non significatives sur des zones à forte variation photométrique locale. Les méthodes de recherche permettent de

comparer une combinaison de plusieurs carrés dans l'image requête avec plusieurs carrés dans les images de la base. La combinatoire induite par cette méthode est très coûteuse en plus d'être imprécise.

Rétroprojection d'histogramme

Proposée par Swain et Ballard [110], la rétroprojection d'histogramme ne nécessite pas de prédéfinir de zones d'images a priori, comme pour les points d'intérêt (voir plus bas). Originale, pionnière mais peu réutilisée, cette méthode procède de la façon suivante : à partir de l'histogramme d'une zone requête éventuellement définie dynamiquement, pour chaque pixel (sa couleur quantifiée est notée c) on définit une intensité égale à la valeur de la cellule c de l'"histogramme quotient" entre l'histogramme de la requête et l'histogramme global de l'image candidate. L'image des intensités créée est moyennée et les lieux de fortes intensités correspondent à la position la plus probable de l'histogramme requête. Cette méthode est très coûteuse puisqu'elle suppose un parcours de chaque image pour chaque couleur d'histogramme au moment de la requête. Smith s'en est inspiré dans VisualSeek [104, 103] pour extraire des régions au moment de l'indexation.

Points d'intérêt

Initialement développés dans un contexte de stereovision, les points d'intérêt [120] détectent et caractérisent les lieux de hautes fréquences photométriques. Ils ont été employés pour des problèmes de mise en correspondance précise de lieux entre deux images correspondant à des vues différentes d'une même scène. Les propriétés de stabilité et de répétabilité de leur détection ont motivé leur application à la recherche d'images par le contenu. Ils ont été initialement appliqués à la recherche *globale* d'images en niveaux de gris [100]. Les points d'intérêt couleur [77] ont ensuite été utilisés pour une approche de recherche par *parties* d'images [36]. La caractérisation de composantes d'images par points est efficace pour retrouver *avec précision* des détails fins, mais leur mise en oeuvre reste aujourd'hui très coûteuse au niveau de la mesure de similarité. De plus, ils ne se prêtent pas à la caractérisation de zones lisses ou uniformes. Une comparaison des approches de recherche par région et par points sera discutée en section 4.8.

Régions d'intérêt

La dernière famille de systèmes de requêtes partielles d'images repose sur la segmentation non-supervisée des images de la base. Chaque image est partitionnée en un ensemble de zones photométriquement homogènes, les "régions", qui sont indexées individuellement par des descripteurs visuels statistiques. Dans une

image requête, l'utilisateur peut sélectionner la région qui l'intéresse et retrouver les images comportant une région similaire. Cette approche a été utilisée, entre autres, dans les systèmes Blobworld [4], VisualSeek [105], Netra [63]. Les composantes d'image étant définies une fois pour toutes par segmentation *off-line* et étant censées représenter les zones d'intérêt pour l'utilisateur, le choix de la segmentation est crucial dans l'approche de recherche par région-exemple. La segmentation, la description de région et la définition de similarité entre régions sont les aspects qui distinguent les différentes approches existantes. Notons que dans le système SIMPLIcity [114] les régions sont aussi extraites par segmentation, mais utilisées pour calculer des similarités *globales* d'images.

Dans le tableau 1.1, nous résumons succinctement les différentes approches de requêtes partielles d'images : détourage manuel, blocs, retro-projection, points d'intérêt et régions d'intérêt. La comparaison est établie sur les étapes de détection des composantes d'une image, de leur description et de recherche. Nous nous plaçons dans un cadre de recherche séquentielle d'images, c'est-à-dire sans utilisation de techniques d'optimisation (telles que les structures d'index). Pour la comparaison des coûts calculatoires, l'ordre du nombre de régions par image est de 5-10 par image et de l'ordre de 300 points d'intérêt par image (selon la méthode proposée par Gouet [36]).

	manuel	blocs	retro-proj.	points	régions
détection des composantes :					
étape	off-line	off-line	on-line	on-line	off-line
coût humain	maximum	nul	×	×	nul
coût machine	nul	très faible	×	×	moyen
indexation des composantes :					
coût machine	moyen (*)	moyen (*)	moyen (*)	moyen	moyen (*)
nature	statistique	statistique	statistique	locale	statistique
recherche :					
coût machine	moyen (**)	moyen (**)	très élevé	très élevé	moyen
précision des détails visuels	moyenne	faible	moyenne	élevée	moyenne

TAB. 1.1 – Comparaison des différentes approches de requêtes partielles d'images. L'approche par régions correspond à un bon compromis en termes d'automatisation, de coût calculatoire et d'adaptation au contenu visuel de l'image. Légende : × : "non-défini", (*) : dépend du descripteur, (**) : dépend du descripteur et de la distance.

Alors que pour les approches manuelles, par blocs et par régions, les composantes d'images sont explicitement définies dans la phase d'indexation (c'est-à-dire "off-line"), la notion de composantes d'image dans le cas des points d'intérêt et de la rétroprojection de primitives ne prend de sens qu'au moment de la recherche (c'est-à-dire "on-line"). La définition *dynamique* des composantes au moment de

la recherche offre une souplesse intéressante, mais entraîne un coût calculatoire de recherche très élevé. De plus, les caractéristiques géométriques (position, taille, forme) sont difficilement intégrables dans ces représentations. La différence et la complémentarité entre les approches régions et points sera détaillée en section 4.8.

La représentation partielle par régions d'intérêt correspond à un bon compromis car elle adaptative au contenu visuel de l'image (contrairement aux blocs), totalement automatique (contrairement au détourage manuel) et rapide pour la recherche (contrairement aux points et à la rétro-projection de primitives). Elle présente l'avantage de réduire considérablement la complexité du système[3], donc de convenir à la recherche dans de grandes bases. De plus, elle se prête naturellement à une modélisation haut niveau des images dans laquelle une image est considérée comme un ensemble structuré d'objets possédant des caractéristiques individuelles et relatives. Le nouveau paradigme de recherche par composition logique de catégories de régions ("paradigme 2") sera un exemple d'utilisation de ce type de modélisation.

1.3 Approches existantes en recherche par régions

Dans la littérature, l'approche la plus développée en recherche d'images par régions correspond au paradigme de recherche par région-exemple. L'utilisateur sélectionne une région exemple dans une images et le système retrouve les images comportant une région visuellement similaire. C'est que nous nommons le "paradigme 1" de recherche d'images par région-exemple.

Les approches existantes (par exemple Blobworld [13], Netra [24]) pour le paradigme 1 diffèrent sur les deux problèmes suivants. Le premier est non-trivial : il s'agit de la détection automatique des régions dans une base d'images qui doivent être significatives pour l'utilisateur. Les régions obtenues sont souvent peu satisfaisantes ou trop petites et donc trop homogènes pour constituer des clés de requête pertinentes. Le second problème est celui de la description et de la similarité visuelle des régions qui doivent prendre en compte la spécificité visuelle des régions. Les descripteurs couleur de régions qui ont été proposés sont dérivés des descripteurs *globaux* d'images.

Une technique différente impliquant la mise en correspondance de régions pour la recherche d'images est proposée par Wang, Li et Wiederhold dans le système SIMPLIcity [116] et plus tard par Sridhar, Nascimento et Li [106, 108]. La si-

[3]car une région peut être caractérisée par un seul index

milarité entre deux images est mesurée comme combinaison des similarités entre toutes les régions constituantes de chaque image. Bien qu'utilisant les régions, cette approche correspond en fait au paradigme de recherche par *image*-exemple. En effet, le système retourne les images dont l'apparence visuelle *globale* est similaire à l'image exemple. L'utilisateur n'a pas la possibilité de désigner explicitement des régions requête. La qualité de la segmentation n'est pas leur objectif.

1.4 Nos contributions

Les contributions de nos travaux portent sur différents aspects de la recherche d'images par régions d'intérêt. Nous proposons, d'une part, une nouvelle approche pour le paradigme 1 de recherche par région exemple. D'autre part, nous présentons un nouveau paradigme de recherche d'images par régions, le "paradigme 2".

Pour le paradigme 1, nous proposerons l'approche de *segmentation en régions grossières et description fine de régions* qui tient compte des spécificités du problème de recherche d'images par région exemple. La nouvelle méthode de détection de régions visera à isoler les composantes visuellement saillantes, les "régions d'intérêt", susceptibles de constituer des clés de requête pertinentes pour l'utilisateur. Les régions extraites devront intégrer une certaine variabilité photométrique intrinsèque afin d'être caractéristiques entre elles dans la base. L'extraction sera grossière dans le sens où les détails fins de l'image seront englobés dans des régions plus importantes. Pour l'aspect de caractérisation visuelle des régions, nous proposerons un nouveau descripteur de la variabilité couleur des régions extraites. Une représentation *fine* des couleurs présentes dans chaque région permettra de tenir compte de la spécificité du contenu visuel des régions au sein d'une base. Comparée à des approches moins fines de description de régions, nous verrons que notre descripteur offre une meilleure représentation de la variabilité photométrique des régions grossières. La combinaison de l'extraction grossière et de la description fine des régions résultera en un système de recherche par régions plus intuitif. Les résultats seront présentés sur la plateforme IKONA dans un scénario de recherche dans une photothèque généraliste.

La dernière contribution de ces travaux portera sur la présentation d'un nouveau paradigme de recherche d'images : le **paradigme 2** de *recherche par composition logique de catégories de régions*. Contrairement aux paradigmes existants en recherche d'images, aucune image ou région exemple n'est nécessaire. L'*image mentale* recherchée par l'utilisateur est suffisante. Un thesaurus photométrique des régions de la base donne à l'utilisateur un aperçu des régions

présentes dans la base. Il fournit à l'utilisateur les "briques d'images" disponibles dans la base pour spécifier une composition typique des images recherchées. Les régions ne sont plus considérées individuellement, mais par types de régions similaires, ce qui diffère de la recherche à partir d'un exemple précis. Le système peut répondre à des requêtes aussi complexes que : "trouver les images composées de régions de ce type et de ce type mais pas de ce type". Ce nouveau paradigme fait le lien avec des mécanismes existants en recherche d'information et ouvre de nombreuses perspectives en recherche d'information visuelle.

Les paradigmes 1 et 2 présentés correspondent à deux scénarios d'usage différents. Dans le paradigme 1, nous nous intéressons à la recherche de similarité visuelle par rapport à une région donnée, grâce à une description fine du contenu photométrique des régions combinée à une mesure de similarité pertinente. Dans le paradigme 2, la composition logique de catégories de régions permet d'exploiter une information de plus haut niveau sur le contenu des images. L'autre différence notable se situe au niveau du mode d'interaction. Alors que le paradigme 1 correspond à une recherche à partir d'un exemple précis, le paradigme 2 se contente d'une *image mentale*.

1.5 Plan du mémoire

Le prochain chapitre sera consacré aux problèmes de similarité et de regroupement de primitives visuelles qui se posera sous différentes formes dans nos travaux. Nous nous intéresserons aux différentes distances existantes pour les distributions de couleur afin d'identifier les plus adéquates. Concernant le regroupement de primitives, nous justifierons notre choix pour un algorithme de classification en le positionnant par rapport à d'autres algorithmes couramment utilisés.

Dans le chapitre 3, nous présenterons la méthode de détection de régions grossières basée sur le nouvel algorithme de segmentation par classification des distributions locales de couleurs quantifiées (les $LDQC$). Les régions extraites seront utilisées par les deux paradigmes présentés dans les chapitres suivants.

Le paradigme 1 de recherche d'images par région exemple sera étudié dans le chapitre 4. Nous détaillerons les prérequis pour une caractérisation visuelle pertinente des régions pour la recherche dans une base de régions. Nous introduirons le nouveau descripteur de variabilité couleur de distribution adaptative de nuances de couleurs (ou $ADCS$) qui fournit une description fine du contenu photométrique de chaque région. Nous lui associerons une mesure de similarité dédiée. Des résultats de recherche par ce paradigme seront présentés avec le nouveau descripteur sur une base d'images de type photothèque.

Dans le chapitre 5, nous présenterons le nouveau paradigme 2 de recherche

d'images par composition logique de catégories de régions. Des résultats sur des scénarios de recherche seront présentés sur une base de type photothèque et sur une base spécialisée de journaux télévisés. Nous discuterons des améliorations et des perspectives envisageables avec cette nouvelle approche.

Finalement, nous proposerons des perspectives plus larges (chapitre 6) concernant la recherche d'images par régions d'intérêt et nous conclurons au chapitre 7.

Chapitre 2

Mesures de similarité et regroupement de primitives visuelles

La méthode d'extraction de régions (chapitre 3) ainsi que les deux approches de recherche d'images par régions (chapitres 4 et 5) reposent largement sur les mesures de similarité et les méthodes de regroupement de primitives visuelles (couleurs moyennes, distributions de couleur). La performance globale de notre système sera étroitement liée au choix de celles-ci.

Dans nos travaux, la caractérisation visuelle par différents types de distributions de couleurs occupera un rôle majeur. Ce type de primitive porte une information riche qui nécessitera l'usage de mesures de similarité évoluées. Dans la première partie de ce chapitre, nous donnerons un aperçu des principales distances existantes permettant de comparer des distributions de couleur. L'usage restreint de ces distances à certains types de distributions motivera l'introduction de la *forme généralisée de la distance quadratique* qui s'applique au cas général des distributions de couleur adaptatives (basées sur des ensembles de couleurs différents).

Concernant le regroupement, les problèmes qui nous concernent sont la quantification couleur d'images, la segmentation, l'indexation couleur et le regroupement de descripteurs similaires. Pour effectuer les regroupements de primitives visuelles, nous ramenons le problème à celui de *classification non-supervisée*. Ce domaine est très vaste et de nombreuses approches ont été proposées dans la littérature. Nous nous intéressons en particulier aux algorithmes de la famille des "k-means". En les replaçant dans un contexte historique à la fois de quantification vectorielle et de classification non-supervisée, nous adopterons la meilleure variante pour notre problème : l'algorithme d'*agglomération compétitive*. Nous nous le situerons par rapport aux algorithmes General Lloyd Algorithm/k-means, Fuzzy

C-Means et Expectation/Maximization largement utilisés pour des problèmes similaires au nôtre, notamment dans des systèmes de recherche par régions. Nous rappellerons les particularités de cet algorithme et nous en proposerons une mise en oeuvre effective pour notre problème.

2.1 Distances pour les distributions de couleurs

Dans cette partie, nous présentons les principales distances existantes pour établir la similarité entre des distributions de couleurs.

Les distances couramment utilisées pour comparer des distributions de couleurs, distances de type Minkowski ou intersection d'histogrammes, reposent sur une simple comparaison cellule à cellule des distributions (on entend par *cellule* la cellule d'histogramme ou *bin*). Cette approche suppose implicitement que les informations associées à chaque cellule (c'est-à-dire les couleurs ici) sont décorrélées les unes des autres ; or il existe une similarité variable entre toute paire de couleurs. Nous qualifions de "daltoniennes" ces distances de distributions dans la mesure où elles ne prennent pas en compte l'information de couleur associée à chaque cellule. En effet, elles n'intègrent pas la métrique de l'espace couleur. Bien que ces distances donnent des résultats satisfaisants dans le cas, par exemple, de la recherche d'images par distributions globales, elles s'avèrent trop imprécises pour notre problème. Un exemple de leurs limitations est qu'elles considèrent à distance maximale les distributions dont les histogrammes ne s'intersectent pas. Or des zones de pixels dont les histogrammes de couleur ne s'intersectent pas peuvent aussi bien paraître visuellement proches que très éloignées. Plus les distributions sont piquées[1] (car les données sont homogènes), plus ce phénomène se produit facilement, en particulier dans le cas de la description de régions.

2.1.1 Distances cellule-à-cellule ou "daltoniennes"

Dans cette première partie, nous présentons les distances qui se basent sur des différences cellule-à-cellule des histogrammes sans tenir compte de la couleur associée à chaque cellule. En conséquence, pour comparer deux histogrammes de couleurs X et Y, ces distances supposent qu'ils soient calculés sur le *même* partitionnement de l'espace couleur en N cellules. Les valeurs de cellules d'histogrammes sont respectivement notées $\{x_i\}$ et $\{y_j\}$. Elles correspondent aux population (ou pourcentage) des pixels ayant la couleur i. Certaines de ces distances sont présentées dans [87].

[1] Nous qualifions de "piquée" une distribution comportant des modes prononcés et peu nombreux.

- Minkowski / L^p

Pour $p \geq 1$, les distances de type Minkowski, notées L^p, s'écrivent :

$$d(X,Y) = \left[\sum_{i=1}^{N} \mid x_i - y_i \mid^p\right]^{1/p}$$

Les distances L^1 (ou "distance de Manhattan") et L^2 (ou "distance euclidienne") sont couramment utilisées pour la recherche d'images par le contenu.

- Intersection d'histogrammes

L'intersection d'histogramme a été proposée par Swain et Ballard [110] :

$$d(X,Y) = \sum_{i=1}^{N} \min \mid x_i - y_i \mid$$

Notons que les mesures suivantes (test du χ^2, divergence de Kullback Leibler et Jensen Difference Divergence) doivent être qualifiées de "mesures de similarités" plutôt que de distance car elles n'en vérifient pas toutes les conditions.

- Test du χ^2

Le test statistique du χ^2 teste l'hypothèse que les échantillons observés $\{x_i\}$ sont tirés de la population représentée par les $\{y_j\}$. On en déduit la distance suivante entre les distributions :

$$d(X,Y) = \sum_{i=1}^{N} \frac{(x_i - \hat{z}_i)^2}{\hat{z}_i}$$

où $\hat{z}_i = (x_i + y_i)/2$.

- Divergence de Kullback Leibler

Issue de la théorie de l'information, la divergence de Kullback Leibler exprime l'entropie relative de la distribution X par rapport à Y :

$$d(X,Y) = \sum_{i=1}^{N} x_i \log \frac{x_i}{y_i}$$

- "Jensen Difference Divergence"

Par rapport à la divergence de Kullback Leibler, la *Jensen Difference Divergence* [57] présente l'avantage d'être symétrique :

$$d(X,Y) = \sum_{i=1}^{N} \left[x_i \log \frac{x_i}{\hat{z}_i} + y_i \log \frac{y_i}{\hat{z}_i} \right]$$

où $\hat{z}_i = (x_i + y_i)/2$.

Si on note $d_{KL}(X,Y)$ la divergence de Kullback Leibler, et $d_{JD}(X,Y)$ celle de Jensen, on a la relation suivante : $d_{JD}(X,Y) = d_{KL}(X,\hat{Z}) + d_{KL}(Y,\hat{Z})$, où $\hat{Z} = (X+Y)/2$.

2.1.2 Distances inter-cellules

Par opposition aux distances présentées précédemment, les "distances inter-cellules" présentent l'avantage d'*intégrer la distance couleur* entre les cellules comparées. Elles ne supposent donc plus que les cellules soient décorrelées.

Parmi ces distances, nous distingons deux types selon si elles nécessitent que les distributions reposent sur des quantifications identiques de l'espace couleur ou si elles permettent l'usage plus général de quantifications différentes donc adaptatives.

Quantifications identiques

Nous présentons trois distances inter-cellules qui supposent des quantifications identiques : la distance quadratique dans sa forme originale, l'histogramme cumulé et l'histogramme flou. Dans le cas des deux histogrammes, la similarité entre les cellules est intégrée dans la primitive histogramme elle-même.

- Histogramme cumulé

Présenté dans [109], l'histogramme cumulé consiste à associer à chaque cellule, non pas la valeur de population couleur correspondante comme dans l'histogramme classique, mais la somme des populations associées aux cellules "précédentes". La distance L^1 est utilisée pour comparer les histogrammes cumulés :

$$d(X,Y) = \sum_{i=1}^{N} \mid \hat{x}_i - \hat{y}_i \mid$$

où $\hat{x}_i$ et $\hat{y}_i$ désignent les valeurs des histogrammes cumulés : $\hat{x}_i = \sum_{j=1}^{i} x_j$ et $\hat{y}_i = \sum_{j=1}^{i} y_j$.

Dans cette approche, la similarité inter-cellule est induite par la largeur de cellule. Cette représentation a été jugée plus robuste que l'histogramme classique [110] car plus tolérante à des changements dans l'affectation des couleurs aux cellules. Cependant elle suppose un ordre dans les cellules pour permettre l'intégration des valeurs de cellules *précédentes*. Pour des histogrammes de niveaux de gris, l'ordre est naturel, mais dans le cas de la couleur, il n'existe pas d'ordre total et cette représentation n'est pas exploitable.

- Histogramme flou

L'histogramme flou [12] constitue une approche simple et naturelle pour intégrer la similarité couleur dans l'histogramme couleur classique. Chaque pixel de couleur de l'image participe à chaque cellule proportionnellement à la similarité entre la couleur du pixel et à la couleur de la cellule. C'est une approche intéressante pour l'indexation d'images globales car elle tient compte de la corrélation entre les couleurs d'un même histogramme, tout en étant couplée à une distance classique de type L^p. Cependant, les résultats présentés ne montrent pas de gain en performance significatif par rapport à l'histogramme classique associé à une distance de type Minkowski.

- Distance Quadratique (forme originale)

Proposée dans [41], la distance couleur quadratique compare toutes les valeurs de cellules des deux histogrammes qu'elle pondère par leur similarité couleur :

$$\begin{aligned} d_q(X,Y)^2 =& (X-Y)^T A (X-Y) \\ =& \sum_{i=1}^{n}\sum_{j=1}^{n}(x_i - y_i)(x_j - y_j)a_{ij} \end{aligned} \tag{2.1}$$

où $A = [a_{ij}]$ est la matrice de similarité couleur a_{ij} entre les couleurs c_i et c_j :

$$a_{ij} = 1 - d_{ij}/d_{max}$$

où d_{ij} est la distance dans l'espace couleur considéré et d_{max} le maximum global de cette distance. Notons que si l'on remplace A par la matrice identité, nous retrouvons la distance euclidienne, i.e. $d_q(X,Y) = ||X - Y||_{L^2}$.

Quantifications différentes

Dans cette partie, nous considérerons le cas de deux distributions de couleur X et Y calculées sur deux ensembles *différents* de n_X et n_Y couleurs, respectivement.

Les distributions considérées précédemment constituent des cas particuliers de ces distributions. On les note ainsi :

$$\{ (c_1^X, x_1), ..., (c_{n_X}^X, x_{n_X}) \} \text{ et } \{ (c_1^Y, y_1), ..., (c_{n_Y}^Y, y_{n_Y}) \}$$

où les c_i^X et c_j^Y sont les triplets de couleurs pour chaque distribution et $\{x_i\}$ et $\{y_j\}$ leurs populations de pixels associées.

Les distances présentées ici sont celles qui permettent la comparaison de distributions de couleurs dans leur expression la plus générale.

- Earth Mover Distance

La distance Earth Mover Distance (EMD) [95] mesure le coût de transformation d'une distribution à une autre. Si le coût de déplacement d'une unité dans l'espace couleur est donné par la distance couleur, alors la distance entre deux distributions est le minimum des sommes des coûts induits par le déplacement de chaque couleur. Le calcul de la distance EMD se ramène à la solution d'un problème de transport résolu par optimisation linéaire :

$$d(X,Y) = \frac{\sum_{i=1}^{n_X} \sum_{j=1}^{n_Y} g_{ij} d_{ij}}{\sum_{i=1}^{n_X} \sum_{j=1}^{n_Y} g_{ij}}$$

où d_{ij} indique la distance entre les couleurs c_i^X et c_j^Y, et g_{ij} le flot optimal entre les deux distributions tel que le coût total $\sum_{i=1}^{n_X} \sum_{j=1}^{n_Y} g_{ij} d_{ij}$ est minimal sous les contraintes suivantes :

$$g_{ij} \geq 0, \ \forall i, j$$

$$\sum_{i=1}^{n_X} g_{ij} \leq y_j, \ \forall j$$

$$\sum_{j=1}^{n_Y} g_{ij} \leq x_i, \ \forall i$$

$$\sum_{i=1}^{n_X} \sum_{j=1}^{n_Y} g_{ij} = \min(x_i, y_j)$$

Cette méthode de comparaison est très coûteuse car elle requiert la résolution d'un problème d'optimisation linéaire soluble de façon itérative.

- "Weighted Correlation"

Plus récemment, la distance "Weighted Correlation" [55] a été introduite conjointement à un algorithme de quantification couleur adaptative. Elle définit la similarité entre deux distributions comme leur corrélation. L'expression de

cette distance pour des distributions discrètes (cas des histogrammes) fait apparaître les poids w, w', w'' :

$$d(X,Y) = 1 - \sum_{i=1}^{n_X} \sum_{j=1}^{n_Y} w'_{ij} x_i y_j$$

avec la normalisation suivante :

$$\sum_{i=1}^{n_X} \sum_{j=1}^{n_Y} w_{ij} x_i x_j = \sum_{i=1}^{n_X} \sum_{j=1}^{n_Y} w''_{ij} y_i y_j = 1$$

Les poids w, w', w'' correspondent aux volumes des intersections entre les cellules couleur, respectivement, de l'histogramme X avec lui-même, de Y avec lui-même et de X avec Y. Dans le cas de leur quantification couleur, les cellules produites sont sphériques dans l'espace couleur. Ces poids expriment une forme de similarité entre les cellules de l'histogramme. Cette distance dépend étroitement de la quantification choisie, mais cet aspect n'est pas évoqué dans l'article.

La complexité algorithmique de comparaison est du même ordre que pour la distance quadratique.

- Distance Quadratique, Forme Généralisée

Exprimée à partir de la distance quadratique (voir formule 2.1), la forme généralisée de la distance quadratique permet, contrairement à la majeure partie des distances employées à ce jour, de comparer des distributions représentées sur des ensembles de couleur *différents*. La mise en oeuvre de cette distance sera développée en section 4.4.

Contrairement à ce qui a été évoqué dans des publications d'origines différentes [95][87][55][35], la distance quadratique permet de comparer des histogrammes basés sur des quantifications *différentes*. Nous allons le voir grâce à une simple ré-écriture de son expression qui nous mènera à la forme généralisée de la distance quadratique.

Nous cherchons à calculer la quantité $d_{quad}(X,Y)$ où X et Y sont exprimées sur n_X et n_Y *couleurs différentes*. Nous allons réécrire l'expression de la distance quadratique afin de supprimer les termes impliquant des différences de cellules. L'idée est d'exprimer X et Y comme des distributions sur un ensemble commun de couleurs de telle sorte que les différences cellule à cellule soient calculables, puis de développer et finalement réécrire l'expression avec les distributions X et Y dans leur forme d'origine.

Nous définissons X' et Y' comme les extensions des distributions discrètes X et Y sur l'*ensemble* de l'espace couleur (Luv ici) de la façon suivante : X'

prend les mêmes valeurs que X sur l'ensemble $\{c_1^X, ..., c_{n_X}^X\}$, c'est-à-dire les valeurs $\{x_1, ..., x_{n_X}\}$, et 0 sur le reste de l'espace. Nous définissons Y' à partir de Y de la même façon. Donc nous avons $d_{quad}(X', Y') = d_{quad}(X, Y)$. X' et Y' étant définies pour le même ensemble de couleur (l'ensemble complet), $d_{quad}(X', Y')$ peut s'exprimer. En notant A la matrice des similarités entre toutes les couleurs de l'espace, on obtient :

$$\begin{aligned} d_{quad}(X,Y)^2 =& d_{quad}(X',Y')^2 \\ =& (X'-Y')^T A (X'-Y') \\ =& X'^T A X' - X'^T A Y' - Y'^T A X' + Y'^T A Y' \end{aligned}$$

La symétrie de la matrice A de similarité couleur entraîne :

$$d_{quad}(X', Y')^2 = X'^T A X' + Y'^T A Y' - 2X'^T A Y'$$

Par construction de X' et Y', nous avons :

$$\begin{aligned} X'^T A X' =& X^T A^X X \\ Y'^T A Y' =& Y^T A^Y Y \\ X'^T A Y' =& X^T A^{XY} Y \end{aligned}$$

où les matrices A^X, A^Y et A^{XY} sont les restrictions de la matrice A qui donnent les similarités de couleur entre, respectivement, les nuances de couleur de X avec elles-mêmes (de dimension $n_X.n_X$), celles de Y avec elles-mêmes (de dimension $n_Y.n_Y$) et celles de X avec celles de Y (de dimension $n_X.n_Y$). On en déduit la formule suivante pour $d_{quad}(X, Y)^2$ dans laquelle n'apparaît plus de différence cellule-à-cellule :

$$d_{quad}(X,Y)^2 = X^T A^X X + Y^T A^Y Y - 2X^T A^{XY} Y$$

et sous forme scalaire, nous obtenons :

$$d_{quad}(X,Y)^2 = \sum_{i,j=1}^{n_X} x_i x_j a_{c_i^X c_j^X} + \sum_{i,j=1}^{n_Y} y_i y_j a_{c_i^Y c_j^Y} - 2\sum_{i=1}^{n_X}\sum_{j=1}^{n_Y} x_i y_j a_{c_i^X c_j^Y} \qquad (2.2)$$

L'expression de la similarité $a_{c_i^X c_j^Y}$ entre deux couleurs c_i^X et c_j^Y de l'espace Luv est donnée plus loin par l'expression (3.1). L'expression (2.2) est la **forme généralisée de la distance quadratique** entre les distributions X et Y déterminées sur des ensembles quelconques de couleurs $\{c_1^X, ..., c_{n_X}^X\}$ et $\{c_1^Y, ..., c_{n_Y}^Y\}$.

Note : la distance proposée dans Netra [24] et reprise dans MPEG7 pour la comparaison des distributions adaptatives du descripteur DCD est une expression particulière de la forme généralisée de la distance quadratique :

$$d(X,Y)^2 = \sum_{i=1}^{n_X} x_i^2 + \sum_{i=1}^{n_Y} y_i^2 - 2\sum_{i=1}^{n_X}\sum_{j=1}^{n_Y} x_i y_j a_{c_i^X c_j^Y}$$

Par rapport à la forme généralisée, elle impose, pour chaque paire de couleur (c_i, c_j) de X et Y, que $a_{c_i c_j} = 1$ si $i = j$ et 0 sinon. Cela signifie, pour chaque distribution, que les couleurs sont à distance maximale les unes des autres, car la quantification sous-jacente est *imprécise*.

En résumé, la distance quadratique permet de prendre en compte la distance couleur dans le calcul de la distance entre distributions. En plus de cet avantage, la forme généralisée que nous venons de présenter permet de comparer des distributions déterminées sur des ensembles de couleurs différents donc reposant sur des quantifications *adaptatives à chaque région*.

2.2 Positionnement du problème de regroupement

Dans les différents problèmes de regroupement que nous recontrerons, nous n'aurons aucune information a priori susceptible de guider le regroupement telle que le nombre de classes ou leurs prototypes, du fait de la nature hétérogène des images traitées ainsi que leur grand nombre. L'algorithme choisi devra pouvoir estimer au mieux par rapport aux données (les *primitives visuelles*) le nombre de classes, leurs prototypes et l'association finale entre données et classes. Par ailleurs, le temps de calcul est un facteur important car l'algorithme sera utilisé de façon intensive sur la base, souvent plusieurs fois pour chaque image de la base. De plus l'algorithme devra pouvoir opérer dans des espaces de dimensions élevées (parfois supérieure à 20).

Introduisons les notations suivantes qui seront utilisées dans cette partie :
- N : nombre de primitives
- $\{x_j, j = 1, ..., N\}$: ensemble des primitives à regrouper ($x_j \in \mathbb{R}^p$, où $p \geq 3$)
- C : nombre de classes ($C < N$)
- $\{\beta_i, i = 1, ..., C\}$: ensemble des prototypes de classes ($\beta_i \in \mathbb{R}^p$, où $p \geq 3$)
- $U : \{1, ..., N\} \mapsto \{1, ..., C\}$: association entre primitives et classes
- $\mathbf{P} = \{C, \beta_i, U\}$: partition à estimer

Le problème se résume en ces termes :
Etant donné un ensemble de primitives $x_i, i = 1, ..., N$ trouver la partition $\mathbf{P} = \{C, \beta_i, U\}$ optimale selon un critère (qui sera à définir).

2.3 Algorithmes de type “k-means”

Notre problème est lié à la fois aux domaines de quantification vectorielle (théorie de la communication et de l'information) et de classification non-supervisée (analyse de données, reconnaissance des formes). Ces deux domaines sont activement étudiés depuis plusieurs dizaines d'années. Nous nous intéressons en particulier à la famille reconnue des algorithmes de type “k-means” qui ont été développés parallèlement dans les deux domaines depuis les années cinquante à nos jours. Ce paragraphe se base principalement sur les articles de synthèse [37] et [49].

Dans un contexte de théorie de la communication et de l'information, les méthodes de quantification (voir l'article de synthèse de Gray et Neuhoff de 1998 [37]) visent à représenter une source (ou “signal”, “données”) en un ensemble fini et réduit de symboles, appelé “codebook”. Dans ce contexte, les applications sont principalement la conversion analogique/numérique et la compression de données. Quant aux techniques de classification, elles sont utilisées dans le contexte de l'analyse de données pour le regroupement de données et exploitent les similarités dans l'espace de représentation des celles-ci. Nous suggérons au lecteur de se référer à l'étude de Jain et Murty de 1999 [49] sur les principales familles de classification.

2.3.1 Le seuillage, degré 0 de la quantification

Considérons l'ensemble de vecteurs $\{x_j, j = 1, ..., N\}$ de l'espace $\mathbb{R}^p$ que nous souhaitons représenter par un nombre réduit de vecteurs $\{\beta_i, i = 1, ..., C\}$ du même espace. La façon la plus simple est de définir les $\{\beta_i\}$ comme les sommets (ou les centres) des cellules d'une subdivision systématique de l'espace $\mathbb{R}^p$ et d'associer chaque x_j à la cellule dans laquelle il se trouve. Autrement dit, il s'agit simplement de seuiller chaque composante réelle de chaque vecteur x_j. Dans le cas de la description couleur d'images, le seuillage des valeurs de pixel est la technique généralement utilisée pour produire les histogrammes couleur [110]. La plupart des descripteurs couleur repose sur ce principe élémentaire.

Dans cette approche, les classes sont de simples hypercubes de l'espace $\mathbb{R}^p$ et les prototypes sont fixés à l'avance en nature et en nombre. Il s'agit naturellement de la technique la plus simple et la plus rapide, mais la moins précise en terme de représentation des données.

2.3.2 k-means / GLA

K-means et GLA figurent parmi les algorithmes les plus souvent rencontrés dans la littérature. Bien que développés dans des domaines distincts (quanti-

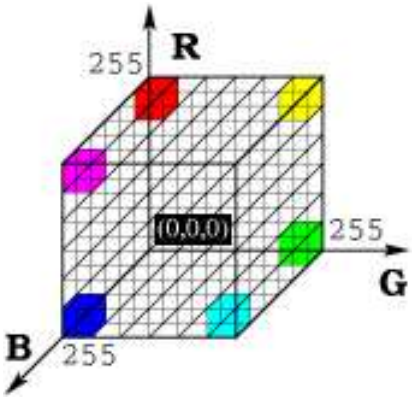


FIG. 2.1 – Exemple de seuillage (ou *quantification systématique*) d'un espace couleur : l'espace RGB (256^3 = 16 millions de couleurs) est partitionné selon 6 valeurs par composante ($6^3 = 216$ cellules). 216 couleurs sont alors utilisées pour représenter les 16 millions de couleurs.

fication vectorielle pour GLA et classification de données pour k-means), leur principe est identique. Il se résume de la façon suivante : l'algorithme suppose donnée une partition initiale. Chaque itération comprend trois étapes : détermination de l'association entre les données et les classes, mise à jour des centres de classes puis mise à jour du critère de convergence. Les itérations se déroulent jusqu'à convergence vers un optimum local, éventuellement global.

Historiquement, c'est en 1957 qu'est proposé pour la première fois ce principe, par Lloyd [60], dans un contexte de quantification pour des données monodimensionnelles[2].

Au milieu des années soixante, le principe de l'algorithme de Lloyd apparaît dans un cadre de classification avec l'algorithme des "**k-means**" présenté par Mac Queen en 1967 [65]. Il permet de classifier des données vectorielles. Pour les problèmes de classification, cet algorithme est le plus simple et le plus courant qui utilise le critère de l'erreur quadratique moyenne [49].

En 1980, Linde, Buzo et Gray [58] proposent une généralisation l'algorithme de Lloyd à la *quantification vectorielle* et introduisent différentes mesures de distorsion (erreur de quantification). Il s'agit de l'algorithme **GLA** (pour "Generalized Lloyd Algorithm") aussi nommé **LBG** d'après les initiales de ses auteurs. Cet algorithme est très largement usité et constitue un algorithme de référence en quantification vectorielle. Le principe de GLA est celui des k-means.

[2]Afin de quantifier les données scalaires en un nombre donné de "niveaux de reproduction" qui partitionnent l'axe réel en intervalles, Lloyd définit un critère d'optimalité en terme de distorsion minimale et propose l'algorithme de descente itérative qu'il nomme "Méthode I". C'est cet algorithme qui est l'ancêtre des algorithmes de types GLA/k-means.

Dans [37], Gray et Neuhoff font le lien terminologique entre l'approche de classification de données par rapport à celle de quantification vectorielle de la façon suivante : *[...] These algorithms were developed for statistical clustering applications, the selection of a finite collection of templates that well represent a large collection of data in the MSE (Mean Squared Error) sense, i.e., a fixed-rate Vector Quantizer with an MSE distortion measure in quantization terminology.*

Déroulement de l'algorithme des k-means/GLA

Soient les données $\{x_j, j = 1, ..., N\}$ et les prototypes (ou "codewords") $\{\beta_i, i = 1, ..., C\}$.

1. A l'itération k=0, les prototypes $\{\beta_i\}$ sont initialisés.
2. Nouvelle itération $k = k + 1$.
3. Association aux classes : à tout prototype β_i on associe la donnée x_j la plus proche. Une classe C_i est définie comme : $\forall i \in \{1, ..., C\},\ C_i = \{x \mid i = \text{argMin}_{i'}(d(x, \beta_{i'}))\}$, où $d(.,.)$ est la distance entre données et prototypes.
4. Mise à jour des prototypes : chaque prototype β_i pour l'itération suivante est défini comme le centroïde de la classe C_i.
5. Calcul de l'erreur moyenne quadratique E^k de la partition.
6. Reprendre à l'étape 2 jusqu'à ce que $\mid E^{k+1} - E^k \mid$ soit faible.

Utilisation dans les systèmes existants

L'algorithme des k-means est utilisé dans la segmentation proposée par Wang [17] et dans le système SIMPLIcity [56]. Leur algorithme détermine le nombre de classes en testant différentes valeurs. Cette heuristique du choix de nombre de classes est lourde car nécessite d'effectuer une classification pour chaque nombre de classes testé. Dans le système Netra [63], l'algorithme GLA est utilisé à deux fins : d'une part pour constituer la palette de 256 couleurs, ou "color codebook", servant à indexer les régions de la base (l'algorithme est appliqué à l'ensemble des couleurs des images de la base). D'autre part, il est utilisé pour sélectionner les couleurs dominantes de chaque région parmi les 256 couleurs de la palette. Les couleurs dominantes du descripteur DCD de la norme MPEG7 [68] sont aussi obtenues par GLA.

2.3.3 C-Moyennes Floues

Dans les algorithmes k-means/GLA le problème majeur est le risque de la convergence du critère vers un optimum local non-global. La partition finale est

sensible à l'initialisation. Afin de réduire ce risque, la classification floue introduit des degrés d'appartenance floue entre les données et les classes. Plutôt que de considérer que l'association entre une donnée et une classe est binaire (appartient ou n'appartient pas), les données sont autorisées à appartenir à plusieurs classes avec des degrés variables. Aux extrêmes, un degré nul exprime la non-appartenance à une classe et un degré de 1 signifie l'appartenance totale. L'algorithme de classification floue le plus usité est le Fuzzy C-Means (**FCM**) [5] introduit par Bezdek en 1981. Il correspond littéralement à la version floue des k-means. Par opposition, les méthodes k-means et GLA sont alors appelées algorithmes "k-means exclusifs" ou "hard k-means". Moyennant un surcoût raisonnable de stockage et de calculs, FCM est plus précise que les hard k-means dans l'estimation des prototypes.
La partition optimale des données est obtenue par minimisation de la fonction objectif suivante :

$$J = \sum_{i=1}^{C} \sum_{j=1}^{N} (u_{ij})^q d^2(x_j, \beta_i)$$

sous la contrainte :

$$\sum_{i=1}^{C} u_{ij} = 1, \ i = 1, ..., N$$

où $u_{ij} \in [0; 1]$ désigne le degré d'appartenance de la donnée x_j à la classe de prototype β_i. $q > 1$ est le paramètre flou (plus q est grand, moins les valeurs d'appartenance sont marquées). La minimisation de cette fonctionnelle par rapport aux u_{ij} est obtenue, comme pour les k-means, itérativement. Elle conduit à la convergence de la partition floue donnée par les u_{ij}.

Déroulement de l'algorithme de FCM

1. A l'itération k=0, les prototypes $\{\beta_i, i = 1, ..., C\}$ et la matrice d'appartenance $U = [u_{ij}]$ sont initialisés.
2. Nouvelle itération $k = k + 1$.
3. Calcul des degrés d'appartenance :

$$u_{ij} = \frac{\left[\frac{1}{d^2(x_j, \beta_i)}\right]^{1/(q-1)}}{\sum_{k=1}^{C} \left[\frac{1}{d^2(x_j, \beta_k)}\right]^{1/(q-1)}}, \forall i, j$$

4. Mise à jour des prototypes :

$$\beta_i = \frac{\sum_{j=1}^{N} u_{ij}^q x_j}{\sum_{j=1}^{N} u_{ij}^q}, \forall i$$

5. Mise à jour du critère de convergence
6. Reprendre à l'étape 2 jusqu'à la stabilité des degrés d'appartenance (i.e. convergence vers zéro de $|| U^(k+1) - U^k ||$).

Utilisation dans les systèmes existants

La segmentation proposée par Grecu et Lambert [38] utilise FCM pour générer une surestimation du nombre de classes qui sont ensuite fusionnées selon des critères flous de compacité et d'isolation.

2.3.4 Expectation/Maximization

Il est important d'évoquer ici la classification par l'algorithme Expectation-Maximization (**EM** [23]). Bien qu'issue des probabilités, l'approche est similaire à FCM. Les données sont considérées comme des observations issues d'une mixture de C distributions (typiquement des gaussiennes) dont EM va chercher à estimer les paramètres selon le maximum de vraisemblance. Dans un contexte de classification, chaque gaussienne modélise une classe et fournit la partition des données. Dans le cas de gaussiennes multivariées, les paramètres de chaque distribution i consistent en leur moyenne μ_i et leur matrice de covariance Σ_i. Les distributions sont mélangées selon les pondérations $\{\gamma_i, i = 1, ..., C\}$. Le paramètre global de la mixture est $\theta = \{ \{\mu_i\}, \{\Sigma_i\}, \{\gamma_i\} \}$. La densité de probabilité de la mixture s'exprime alors ainsi :

$$\begin{aligned} p(x;\theta) &= \sum_{i=1}^{C} \gamma_i p_i(x \mid i;\theta) \\ &= \sum_{i=1}^{C} \gamma_i . \frac{1}{2\pi^{(p/2)} \mid \Sigma_i \mid^{1/2}} \; exp\left(-\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1} (x-\mu_i) \right) \end{aligned}$$

où p est la dimension des données x avec la contrainte $\sum_{i=1}^{C} \gamma_i = 1$. Le critère d'optimalité est le maximum de log-vraisemblance, qui s'écrit dans ce cas :

$$\mathcal{E}(\theta) = \sum_{j=1}^{N} ln \sum_{i=1}^{C} \gamma_i p(x_j \mid i)$$

Déroulement de l'algorithme EM

1. A l'itération k=0, le paramètre global θ est initialisé à $\theta^0 = \{ \{\mu_i^0\}, \{\Sigma_i^0\}, \{\gamma_i^0\} \}$.

2. Nouvelle itération $k = k + 1$.
3. Etape "Expectation" : estimation de l'appartenance des données aux distributions.
$$p^k(i \mid x_j) = \frac{\gamma_i^k p^k(x_j \mid i)}{p^k(x_j)}$$
4. Etape "Maximization" : la maximisation de log-vraisemblance donne les formules suivantes de mise à jour des paramètres θ^{k+1} à partir de θ^k :

$$\begin{aligned}
\mu_i^{k+1} =& \frac{\sum_{j=1}^N p^k(i \mid x_j) x_j}{\sum_{j=1}^N p^k(i \mid x_j)} \\
\Sigma_i^{k+1} =& \frac{\sum_{j=1}^N p^k(i \mid x_j)(x_j - \mu_i^k)(x_j - \mu_i^k)^T}{\sum_{j=1}^N p^k(i \mid x_j)} \\
\gamma_i^{k+1} =& \frac{1}{N} \sum_{j=1}^N p^k(i \mid x_j)
\end{aligned}$$

5. Calcul de la variation du critère de convergence :
$$\begin{aligned}
\Delta =& \mathcal{E}(\theta^{k+1}) - \mathcal{E}(\theta^k) \\
=& -\sum_{j=1}^N ln(\frac{p^{k+1}(x_j)}{p^k(x_j)})
\end{aligned}$$
6. Reprendre à l'étape 2 jusqu'à ce que Δ soit faible.

A la convergence de EM, le modèle des C mixtures est estimé optimalement par rapport au maximum de vraisemblance, avec le risque d'obtenir un minimum *local* selon l'initialisation des données choisie. Les probabilités $p^k(i \mid x_j)$ donnent l'association finale entre chaque donnée et l'une des C gaussiennes.

Issus respectivement de la logique floue et des probabilités, les algorithmes FCM et EM adoptent des approches très similaires. Dans les deux cas, l'association des données aux classes est estimée optimalement (au sens du critère d'erreur) au fil des itérations de façon floue dans FCM et probabiliste dans EM et fournit, à la convergence, la partition finale. L'association entre les données x_j et les classes i pour FCM et les distributions i pour EM s'expriment ainsi :

– dans EM, $p^k(i \mid x_j)$ exprime la probabilité que x_j soit générée par une distribution i, avec la contrainte $\sum_{i=1}^C p^k(i \mid x_j) = 1$ pour toute distribution i

– dans FCM, u_{ij} exprime le degré d'appartenance de x_j à une classe i, avec la contrainte $\sum_{i=1}^{C} u_{ij} = 1$ pour toute classe i

Concernant la forme des classes détectées, l'usage de la distance de Mahalanobis [40] dans FCM modélise des classes hyperellipsoïdales et s'exprime à partir de la matrice de covariance Σ_i (voir expression (2.6)) que l'on retrouve dans les paramètres des gaussiennes estimées dans EM.

Utilisation dans les systèmes existants

La segmentation du système Blobworld [4] utilise l'algorithme EM pour le regroupement de primitives de couleur et de texture. Les moyennes de gaussiennes sont initialisées aléatoirement et les matrices de covariances par la matrice identité (ils ont abandonné une stratégie d'initialisation plus évoluée qui n'améliorait pas leurs résultats). Pour le choix du nombre de distributions (donc de classes), les valeurs 2 à 5 sont testées et le nombre retenu est celui qui vérifie le critère de Minimum Description Length [91]. Comme pour les algorithmes précédents, le choix automatique du nombre de classes ne peut être fait qu'en exécutant plusieurs fois l'algorithme.

2.3.5 Limitations de ces approches

Nous venons de voir que les algorithmes k-means/GLA, FCM et EM procèdent de façon très similaire : initialisation, mise à jour puis convergence d'une partition qui minimise un critère de fidélité aux données. Elles visent à estimer de façon plus ou moins fine et plus ou moins rapide les centres de classes ainsi que l'association des données à ceux-ci en minimisant un critère. Elles présentent deux problèmes intrinsèques :

Sensibilité à la partition initiale

Si ces algorithmes peuvent être initialisés avec une "bonne" partition initiale, ils permettront de trouver efficacement l'optimum global même sur de grands ensembles de données [49]. En l'absence de "bonne" partition initiale, même avec une gestion floue (ou probabiliste avec EM) de l'association entre données et prototypes, la convergence de ces algorithmes vers un optimum global n'est pas toujours assurée. Ce défaut est fréquent dans les approches paramétriques de classification dont les k-means font partie.

Estimation du nombre de classes

Dans la détermination des partitions que nous venons de voir, le nombre de classes est toujours supposé donné. En quantification vectorielle, la contrainte d'un nombre de niveaux de représentation constant peut-être justifié par des raisons techniques (par exemple : nombre de couleurs que peut afficher un écran dans le cas de la quantification d'une image). Pour notre problème (générique) de regroupement de primitives visuelles, le nombre de classes n'est pas donné a priori et nous pensons qu'il doit plutôt refléter la nature des données. Nous souhaitons que le nombre de classes soit aussi un paramètre à estimer de façon optimale.

En théorie, le nombre optimal de classes peut être défini comme celui qui optimise le critère de classification sur les différentes partitions obtenues pour tous les nombres possibles de classes. La mise en oeuvre de cette stratégie n'est absolument pas viable car trop coûteuse, même pour des ensembles réduits de données.

Autres techniques de classification

De conception très différente de celles citées, les techniques de classification non-paramétrique ont une vision locale des données. Souvent basée sur les fenêtres de Parzen, elles exploitent les similarités entre les données elles-mêmes. Nous citerons l'estimation de gradient de densité [18] qui a été appliquée efficacement à la sur-segmentation et quantification d'image couleur par Comaniciu [19]. Leur inconvénient est de supposer que les données soient denses et que l'espace de représentation soit de faible dimension (de l'ordre de 6 pour [19]). Elles deviennent vite inadéquates au regroupement de primitives visuelles de primitives de dimension supérieure à trois, c'est-à-dire plus évoluées que les simples triplets de couleur.

Dans la suite de ce chapitre, nous allons présenter l'algorithme d'agglomération compétitive (CA) dont la conception est basée sur celle de FCM. Contrairement aux approches présentées précédemment, CA présente l'avantage majeur d'estimer le nombre de classes au même titre que les prototypes. De plus, la surestimation du nombre de classes à l'initialisation permettra d'atténuer fortement le problème de l'initialisation de la partition.

2.4 L'algorithme CA

Notre choix pour regrouper les primitives par classification non-supervisée se porte sur l'algorithme d'Agglomération Compétitive (dit **CA** pour "Competitive

Agglomération") proposé par Frigui et Krishnapuram en 1997 [32].

Dans un contexte plus général de classification non-supervisée, il présente à la fois les avantages des algorithmes hiérarchiques (typiquement les non-paramétriques) et ceux des algorithmes à partitionnement (typiquement les k-means) :

(i) le nombre de classes est déterminé automatiquement

(ii) la sensibilité à l'initialisation est réduite grâce au grand nombre de classes initiales

(iii) l'appartenance floue des données aux classes limite le risque du minimum *local* de la partition, car les données ne sont pas définitivement associées à une classe

(iv) il n'y a pas de contrainte a priori sur la faible dimensionnalité des données

Les avantages de CA par rapport aux algorithmes de classification détaillés précédemment sont les suivants : par rapport à k-means/GLA, CA présente les avantages (i)(ii)(iii), par rapport à FCM et EM les avantages (i)(ii) et par rapport aux non-paramétriques principalement l'avantage (iv). Notons par ailleurs si k-means/GLA, FCM, EM sont appliqués un grand nombre de fois pour déterminer en plus le nombre optimal de classes, leur temps d'exécution peut être beaucoup plus long qu'avec CA.

2.4.1 Principe de l'algorithme

En utilisant les notations de [32], nous appellerons $\{x_j, \forall j \in \{1, ..., N\}\}$ l'ensemble des N données que nous souhaitons classifier et C le nombre de classes. $\{\beta_i, \forall i \in \{1, ..., C\}\}$ représentent les prototypes à déterminer. Les C prototypes seront initialisés à partir des données comme nous le verrons plus tard lors de la mise en oeuvre de la classification. La distance entre une donnée x_j et un prototype β_i est $d(x_j, \beta_i)$ et doit être choisie efficacement en fonction des données à classifier. La distance euclidienne permettra de détecter des classes sphériques tandis que, plus générale, la distance de Mahalanobis [40] détectera des classes ellipsoïdales.

La classification CA est obtenue par minimisation de la fonction objectif J suivante :

$$J = J_1 + \alpha J_2, \tag{2.3}$$

où

$$J_1 = \sum_{i=1}^{C} \sum_{j=1}^{N} u_{ij}^2 d^2(x_j, \beta_i)$$

et

$$J_2 = -\sum_{i=1}^{C}[\sum_{j=1}^{N} u_{ij}]^2$$

sous la contrainte d'appartenance :

$$\sum_{i=1}^{C} u_{ij} = 1, \forall j \in \{1, ..., N\}$$

où u_{ij} représente le degré d'appartenance de la donnée x_j au prototype β_i.

Le terme J_1 est la fonction objectif de FCM dont la minimisation vise à réduire les distances intra-classes. Ce terme peut être interprété comme une mesure de l'erreur quadratique faite si l'on représente les données x_j par les prototypes β_i. La motivation de CA réside dans l'ajout du terme J_2 qui est minimal lorsque le nombre de classes est faible. α est le poids de compétition entre les deux termes J_1 et J_2 dont les effets sont opposés. Ainsi, en cherchant à minimiser la fonctionnelle J, CA vise à optimiser à la fois à minimiser les distances intra-classes ainsi qu'à réduire le nombre de classes. De façon plus générale, l'idée maîtresse de l'algorithme est d'intégrer au sein même du processus de classification le terme J_2 qui est un critère de validité de partition (voir [6]).

En effet, dans [6], le rapprochement a été établi entre le terme J_2 et les critères de validité de partition. En testant des partitions avec différents nombres de classes, le nombre optimal de classes est défini comme étant celui qui optimise le critère de validité. Comme nous l'avons vu précédemment, c'est l'approche notamment adoptée dans Blobworld avec l'algorithme EM et le critère MDL testé sur 2 à 5 classes. En optimisant le critère J_2 dans la classification elle-même, CA permet de déterminer automatiquement le nombre de classes en une seule passe.

Nous allons voir dans le paragraphe suivant que la minimisation de J est effectuée de manière itérative. Etant donnée une partition initiale des données, les paramètres suivants seront mis à jour au fil des itérations : le nombre de classes C, la valeur des prototypes $\{\beta_i\}$ et la matrice des appartenances $U = [u_{ij}]$. A la convergence, C correspondra au nombre de classes déterminé automatiquement, les $\{\beta_i\}$ fourniront les centres optimaux de classes et la matrice d'appartenance floue U nous permettra de décider de l'association finale entre données et classes.

La résolution du problème de minimisation de J [32] conduit à l'expression suivante des degrés d'appartenance u_{ij} :

$$u_{ij} = u_{ij}^{FCM} + u_{ij}^{Biais} \tag{2.4}$$

avec :

$$\begin{aligned} u_{ij}^{FCM} =& \frac{1/d^2(x_j, \beta_i)}{\sum_{k=1}^{C} 1/d^2(x_j, \beta_k)} \\ u_{ij}^{Biais} =& \frac{\alpha}{d^2(x_j, \beta_i)}(N_i - \overline{N}_j) \end{aligned}$$

où

$$\begin{aligned} N_i =& \sum_{k=1}^{N} u_{ik} \\ \overline{N}_j =& \frac{\sum_{k=1}^{C} 1/d^2(x_j, \beta_k).N_k}{\sum_{k=1}^{C} 1/d^2(x_j, \beta_k)} \end{aligned} \tag{2.5}$$

Notons que u_{ij}^{FCM} correspond à l'expression des degrés d'appartenance dans l'algorithme FCM pour un facteur flou $q = 2$. N_i définit la population floue de classe i.

Dans l'expression (2.3), le facteur α définit l'importance relative entre les deux fonctions objectif J_1 et J_2. Celles-ci ont deux effets opposés : J_1 est minimum lorsqu'il y a autant de classes que de données tandis que J_2 l'est avec une seule classe. La conséquence de la compétition entre les deux termes est que les classes peu représentatives auront tendance à se dépeupler (c'est-à-dire N_i faible) au profit des autres classes et seront supprimées. Le nombre de classes converge en décroissant.

A l'iteration k, le poids de compétition α s'exprime de la façon suivante :

$$\alpha(k) = \eta_0 \exp(\frac{-k}{\tau}) \frac{\sum_{i=1}^{C} \sum_{j=1}^{N} u_{ij}^2 d^2(x_j, \beta_i)}{\sum_{i=1}^{C} [\sum_{j=1}^{N} u_{ij}]^2}$$

Au fil des itérations, α décroît, donc l'importance est d'abord accordée au processus d'agglomération des classes (minimisation de J_2), puis à l'optimisation de la classification (minimisation de J_1). Ainsi le choix automatique du nombre de classes est effectué principalement lors des premières itérations. Une fois le nombre de classes stabilisé, le comportement de CA devient proche de celui de FCM. α est complètement déterminé par son amplitude η_0 et son taux de décroissance τ.

Les valeurs de prototypes sont initialisées, puis mises à jour selon la formule suivante :

$$\beta_i = \frac{\sum_{j=1}^{N} u_{ij}^2 x_j}{\sum_{j=1}^{N} u_{ij}^2}, \forall i \in \{1, ..., C\}$$

β_i correspond au barycentre de toutes les données pondérées par leur degré d'appartenance à la classe i.

Les degrés d'appartenance u_{ij} sont initialisés à partir de la formule (2.4) en posant $N_i = 1/C$ pour chaque classe i (populations uniformes à l'initialisation).

A chaque itération, les classes à faible population sont supprimées, i.e. les classes i telles que $N_i < \epsilon$, où ϵ est le seuil de population minimum. C'est dans ce cas que le nombre de classes C diminue.

Note : plutôt que d'utiliser directement la distance euclidienne qui détecte des classes sphériques, nous pouvons utiliser la distance de Mahalanobis [40], basée sur la distance euclidienne et la covariance des données. Dans un cadre de segmentation par classification de primitives visuelles, cette distance permet de détecter des classes ellipsoïdales qui fournissent un bon modèle notamment pour les dégradés de couleur. Elle s'exprime ainsi :

$$d^2(x_j, \beta_i) = \mid \Sigma_i \mid^{1/n} (x_j - c_i)^T \Sigma_i^{-1} (x_j - c_i) \tag{2.6}$$

où c_i est le centre de la classe de prototype β_i et Σ_i sa matrice de covariance floue. La matrice de covariance est mise à jour selon la formule suivante :

$$\Sigma_i = \frac{\sum_{j=1}^{N} u_{ij}^2 (x_j - c_i)(x_j - c_i)^T}{\sum_{j=1}^{N} u_{ij}^2} \tag{2.7}$$

2.4.2 Raffinement pour une mise en oeuvre effective

Par rapport à l'article de référence [32], nous proposons les raffinements suivants visant à rendre la classification plus efficace et adaptée à nos problèmes.

Critères de convergence

La condition de convergence proposée dans l'article de référence est la stabilité des prototypes. Pour s'assurer que le comportement de l'algorithme CA converge bien vers celui de FCM, c'est-à-dire que les dernières itérations soient dédiées à l'optimisation de la partition U, nous allons voir qu'un second critère de convergence doit être imposé.

La matrice d'appartenance U est mise à jour à chaque itération. Chaque élément u_{ij} exprime le degré d'appartenance floue dans $[0; 1]$ de la donnée x_j à la classe de prototype β_i. Pondéré par α (voir expression (2.4)), le terme u_{ij}^{Biais} tend vers 0 au fil des itérations et u tend alors vers u_{ij}^{FCM}. Dans l'implantation de CA, en plus de la stabilité des prototypes, on impose alors comme critère de convergence supplémentaire que les u_{ij}^{Biais} soient négligeables afin que les u_{ij} convergent bien vers les u_{ij}^{FCM}.

A la convergence, une étape de décision sera nécessaire pour obtenir la partition finale : pour chaque donnée x_j, on associera la classe i qui maximise l'appartenance u_{ij}. On aura alors, quand u_{ij}^{Biais} est négligeable, pour chaque donnée j $\in \{1, ..., N\}$:

$$u_{ij} = u_{ij}^{FCM}, i^* = argMax(u_{ij}) \Rightarrow i^* = argMax(u_{ij}^{FCM})$$

D'après l'expression de u_{ij}^{FCM} (formule (2.4)), on a :

$$i^* = argMin(d(x_j, \beta_i))$$

Cela se traduit par le fait que, à la convergence, la classe i^* qui maximise u_{ij} pour une donnée x_j est celle dont le prototype est le plus proche de x_j. En résumé, l'ajout du critère d'arrêt basé sur des u^{Biais} négligeables entraîne que les données seront correctement associées aux prototypes les plus proches.

Proximité inter-classes

Afin de garantir une distance minimale inter-classes à l'issue des itérations, on peut effectuer une fusion des classes proches, c'est-à-dire celles dont les prototypes sont à distance inférieure à un seuil donné. Ce seuil est choisi selon les données, la distance adoptée et le problème de classification. Par exemple, dans un but de quantification couleur d'images (pour les LDQC, voir chapitre 3) ou de régions (pour les ADCS, voir chapitre 4), on imposera une distance minimale entre les couleurs prototypes qui correspond à un seuil de discrimination perceptuelle entre les couleurs, établi empiriquement. Cette étape évite un nombre inutilement élevé de classes.

Notons que ce seuil est à définir avec précaution, car un seuil trop élevé nuit à la qualité de la classification.

Influence des paramètres sur la granularité de classification

La granularité de la classification, c'est-à-dire l'effet d'agglomération des données, est contrôlée principalement par le seuil ϵ de population minimum de classes et par le nombre C de classes initiales. Plus ϵ est faible, plus grand sera le nombre de classes et plus des petites classes pourront "survivre". Quant au nombre de classes initiales il a une certaine incidence sur le nombre de classes finales. A l'usage, nous avons observé qu'une modification de quelques unités du nombre de classes initiales ne modifiait pas le nombre final, mais qu'une modification par un facteur multiplicatif (deux ou plus) augmentait sensiblement le nombre de classes finales. Un grand nombre de classes initiales tend à augmenter dans une certaine mesure le nombre de classes finales. Concernant les paramètres η_0 et τ, ils influent

aussi sur la granularité de la classification puisqu'ils permettent de renforcer ou réduire la minimisation du second terme J_2 (formule (2.3)). Au fil des itérations, le nombre de classes C décroît plus fortement avec η_0 et plus rapidement avec τ. Cependant, pour nos différents problèmes, les meilleurs résultats ont été obtenus avec les paramètres $\eta_0 = 5$ et $\tau = 10$ proposés dans l'article de référence [32].

Quant au seuil de proximité inter-classes (voir plus haut), il doit être choisi en rapport avec la granularité de la classification, mais ne doit pas constituer un critère de granularité en tant que tel.

Initialisation de l'algorithme

Nous choisissons le nombre initial de classes C comme le maximum estimé de nombre de classes attendues sur un grand nombre de cas de classifications.

Les prototypes $\{\beta_i\}$ étant supposés représenter au mieux les données, il est a priori préférable de les initialiser à partir des données elles-mêmes plutôt qu'à partir d'une initialisation aléatoire. Lorsque nous classifierons des primitives issues d'une image, nous verrons que la répartition spatiale de ces primitives nous aidera à affiner le choix des prototypes initiaux.

La matrice U étant définie récursivement au fil des itérations, son initialisation est aussi nécessaire. En effet, dans la formule (2.4), u^{Biais} dépend des populations N_i dépendant elles-même de U. Pour le premier calcul de U, nous considérerons par défaut que toutes les classes auront la même population, c'est-à-dire $N_i = N/C, \forall i = 1, ..., C$.

Déroulement global de l'algorithme

Le déroulement de l'algorithme détaillé ci-dessous résume l'ensemble des étapes de la mise en application de CA à notre problème de regroupement de primitives visuelles.

Initialisation :

- du compteur k=0
- des prototypes $\{\beta_i^0\}$ à partir des données en certains sites d'images
- du nombre de classes C comme le nombre maximum
- de la matrice U^0 par la formule (2.4), en posant $N_i = 1/C, \forall i$
- des $N_i, \forall i$ par la formule (2.5)

Répéter

- k=k+1
- mise à jour :
  - des distances $d^2(x_j, \beta_i),\ \forall i, j$
  - de α_k (formule (2.4.1))

  - de la matrice d'appartenance U^k (formule (2.4))
  - des populations $N_i, \forall i$ (formule (2.5))
- suppression des classes vides : $\forall i$, si $N_i < \epsilon$, supprimer la classe i
- mise à jour :
  - du nombre de classes C
  - des prototypes $\beta_i^k, \forall i$ (formule (2.4.1))
- calcul du critère de convergence :
  - $\kappa_1 = max_{i,j}(u_{ij}^{Biais})$
  - $\kappa_2 = max_i \left(|| \ \beta_i^{k+1} - \beta_i^k \ || \right)$

Jusqu'à ce que κ_1 et κ_2 soient faibles, pour garantir la convergence de l'algorithme vers FCM et la stabilité des prototypes

Fusion d'éventuelles classes proches : tant qu'il existe deux classes très proches (i.e. $i \neq j$ tels que $|| \ \beta_i - \beta_j \ ||$ soit faible), fusionner la classe de population la plus faible parmi $\{i, j\}$ dans la plus grande en additionnant les degrés de la plus petite à ceux de la plus grande

Partition finale : chaque donnée x_j est associée à la classe i qui maximise u_{ij}

2.4.3 Perspectives

Récemment Le Saux et Boujemaa ont proposé une amélioration de l'algorithme CA, dans un but de catégorisation de bases d'images : l'algorithme de classification non-supervisée ARC (Adaptive Robust Competition)[99, 98]. Le processus de regroupement s'adapte aux densités variables des groupes naturels de données. De plus, l'ajout d'une classe-bruit [21] permet à ARC de regrouper les données isolées (dites *outliers*) et ambiguës dans une classe distincte, la classe-bruit, afin d'améliorer le partitionnement des données qui présentent un regroupement plus évident.

Une étude est actuellement en cours, au sein de l'équipe IMEDIA sur l'utilisation de l'algorithme ARC pour l'extraction de régions [46].

2.5 Conclusions

Nous avons, dans ce chapitre, commencé par donner un aperçu des distances existantes permettant de mesurer la similarité entre les distributions de couleur. Les distances existantes ont été présentées selon si elles peuvent ou non tenir compte de quantifications couleur adaptatives. La distance quadratique (pour des quantifications identiques) et sa forme généralisée (pour des quantifications différentes) s'avèreront être des choix pertinents dans la suite de nos travaux pour les raisons suivantes : temps de calcul raisonnable comparée à EMD, plus fidèles

en termes de perception visuelle que les distances cellules-à-cellule et adaptées à tout type de quantification couleur.

Dans la suite du chapitre, nous avons traité le problème de regroupement de primitives. Nous avons d'abord présenté les méthodes de classification non-supervisée les plus couramment usitées dans la famille des k-means : k-means/GLA, FCM, et EM. Nous avons vu que, bien qu'issues des domaines de quantification vectorielle, d'analyse de données ou de probabilités, ces techniques fonctionnent de façon très similaire. Ces techniques sont celles majoritairement utilisées dans la littérature, comme nous le verrons dans les prochains chapitres, pour des problèmes de regroupement de primitives dans un cadre de recherche d'images par régions. Cependant elles souffrent du problème de l'initialisation de la partition qui se traduit par le risque de convergence vers un optimum local ainsi que de l'estimation du nombre de classes. Il s'agit d'aspects critiques pour des problèmes telles que la segmentation, l'indexation, la quantification couleur d'image dans des bases hétérogènes.

Nous avons détaillé l'algorithme d'agglomération compétitive (*CA*) qui présente l'avantage majeur d'estimer automatiquement le nombre de classes, de réduire la sensibilité à l'initialisation et de réduire le risque de l'optimum local), tout en optimisant la partition des données et les centres de classes. En complément à l'algorithme initial, nous avons présenté la mise en oeuvre effective de cet algorithme pour le regroupement de primitives visuelles (initialisation, choix des paramètres et critère de convergence).

A travers quatre étapes distinctes dans nos travaux, nous allons montrer dans la suite comment utiliser efficacement les propriétés de CA pour le regroupement de primitives visuelles.
Les avantages de CA, en particulier celui de la détermination automatique du nombre de classes, en font un algorithme privilégié pour les problèmes de regroupement non-supervisé de primitives visuelles. Dans nos travaux, il nous permettra d'effectuer efficacement des regroupements de primitives de différentes natures pour des problèmes distincts :

- quantification couleur de l'image (classification de triplets de couleur)
- regroupement des primitives LDQC (classification de distributions couleur)
- description des régions avec ADCS (classification de triplets de couleur)
- regroupement des régions similaires (classification des descripteurs des régions de la base)

Chapitre 3

Détection des composantes visuelles par segmentation en régions

Les paradigmes 1 et 2 de recherche d'images présentés dans les chapitres suivants reposent sur les parties constituantes d'une image, plutôt que sur les images vues comme un tout. Ces parties doivent représenter des zones d'intérêt pour l'utilisateur. Dans le chapitre introductif, nous avons évoqué différents modes de représentation partielle d'image. Parmi ceux-là, la segmentation automatique en régions est une approche naturelle et simple pour représenter l'ensemble des composantes d'une image susceptibles d'intéresser l'utilisateur.

Dans ce chapitre, nous proposons une nouvelle méthode de détection automatique des régions adaptée au problème de requête partielle dans les bases d'images. Rapide et non-supervisée, cette méthode permet de détecter les *régions grossières* des images, c'est-à-dire visuellement saillantes pour l'utilisateur. Les régions extraites intègrent une variabilité photométrique qui les rendra caractéristiques les unes des autres dans la base de régions.

3.1 Aperçu des approches existantes

Nous rappelons la définition générale de la segmentation d'images en régions.

Définition 1. La **segmentation d'une image** I est son partitionnement en sous-ensembles de pixels adjacents, notés $\{R_i\}$ et appelés **régions**, présentant des attributs photométriques similaires, tel que :

- $R_i \neq \emptyset$
- $\cup R_i = I$
- $R_i \cap R_j = \emptyset, \forall i \neq j$

Etudiée depuis plus de 30 ans, la segmentation d'image est utilisée dans des domaines très variés où l'imagerie est présente : médical, militaire, surveillance, satellitaire, robotique, contrôle qualité. Elle consiste à détecter les zones d'intérêt d'une image selon un critère d'homogénéité photométrique. Elle dépend donc souvent du domaine d'application. La prise en compte, lorsqu'elle est possible, de connaissance a priori du domaine dans l'algorithme de segmentation permet un gain de performance.

La segmentation demeure aujourd'hui encore un problème ouvert et complexe. On pourra se référer à l'article de synthèse sur les principales familles de segmentation par Pal et Pal [81], ou à celui de Haralick et Shapiro [43], ou de Fu et Mui [33]. Elles sont généralement caractérisées par la nature de la primitive photométrique et par la méthode permettant de les regrouper. Différents critères peuvent les distinguer : nature des primitives utilisées, nature de l'image (niveaux de gris, couleur, multi-spectrale), précision de la détection, robustesse, répétabilité, rapidité, supervision (intervention de l'utilisateur), intégration de connaissances spécifiques. Historiquement, la majeure partie des techniques ont été développées pour les images en niveaux gris. Dans ce cas les pixels sont simplement caractérisés par un scalaire. L'évolution des moyens informatiques (en terme de traitement, stockage et affichage) ont permis relativement récemment de généraliser l'usage de l'image couleur là où elle est désirable ou nécessaire. Les techniques de segmentation couleur se sont réellement développées à partir de ce moment. Les images en niveaux de gris peuvent toujours être ramenées au cas particulier d'une image couleur quelque soit l'espace de représentation (RGB, Luv, Hsv, ...). Nous nous intéresserons donc au cas général de l'image couleur (qui est majoritaire dans les bases utilisées) en gardant à l'esprit que tous les algorithmes présentés dans nos travaux peuvent être aussi appliqués à des images en niveaux de gris.

Dans le contexte spécifique de la recherche d'images par le contenu sur des bases hétérogènes de type photothèque, la segmentation est un problème particulièrement difficile étant donnée la nature composite des images. Dans la littérature, différentes méthodes d'extraction de régions ont été proposées avec des contraintes de qualité de régions extraites très variables. Une comparaison objective de ces méthodes est délicate du fait de l'impossibilité d'établir une base de référence d'images segmentées. L'évaluation repose généralement sur la présentation d'exemples d'images segmentées. Elle est d'autant plus difficile lorsque les exemples sont très peu nombreux (parfois même absents) ou non significatifs (typiquement scènes de coucher de soleil, fleurs se détachant sur un fond vert). De plus, les régions extraites sont rarement rendues disponibles à l'utilisateur dans l'interface graphique du système de recherche par régions.

Dans cette partie, nous allons brièvement citer les différentes approches qui ont été proposées pour extraire les régions dans les bases d'images. Nous résumerons les approches par leurs points principaux à savoir la caractérisation photométrique

et les méthodes de regroupements de primitives.

Une première famille d'approches se base sur une caractérisation photométrique simple : les couleurs, sans intégrer d'information de variation locale. Parmi celles-ci, certaines extraient les régions directement dans l'espace image par regroupement de pixels de couleurs similaires par "split and merge" [78] ou par une variante de l'algorithme "single-link" [107], ou par des heuristiques de fusion de régions adjacentes utilisant la couleur moyenne et une réduction de complexité de contours [44].

A partir de la couleur, d'autres approches effectuent une quantification systématique (voir section 2.3.1) a priori de l'espace couleur puis effectuent un filtrage dans l'espace image [51], ou une fusion de blocs fixes basée sur les couleurs quantifiées [50], ou encore une extraction de région par retro-projection d'histogramme couleur suivie d'un filtrage spatial [104, 103].

Plus adaptatif que la quantification systématique, la classification non-supervisée des pixels couleurs a été proposée avec FCM sur les composantes (a,b) de l'espace Lab [38], avec k-means sur la composante Hue de Hsv [48], ou encore avec k-means sur les composantes (L,a,b,x,y) (couleur et position) [53].

Dans ces approches, notons que la seule prise en compte de l'information de couleur peut difficilement identifier des zones homogènes en texture et ne peut produire de segmentation satisfaisante sur des images de nature diverse.

Une seconde famille d'approches effectue un regroupement sur des primitives photométriques plus riches tenant compte de la couleur et de la texture : dans Windsurf [2], regroupement par k-means de vecteurs mixtes constitués des composantes de la transformée en ondelettes calculée sur chaque composante couleur de l'espace Hsv ; dans SIMPLIcity [116], regroupement par k-means de vecteurs mixtes constitués des composantes couleurs de l'espace Luv et des coefficients de la transformée en ondelettes de Daubechies calculées sur la luminance ; dans Blobworld [4, 15], regroupement par EM de vecteurs mixtes constitués des composantes couleurs de l'espace Lab, de coefficients d'anisotropie, polarité et contraste (calculé à partir du gradient des intensités) puis des positions (x,y) des pixels. Notons que dans ces trois approches, la sélection du nombre de classes se fait par répétition de l'algorithme de k-means sur une fourchette de différents nombres de classes probables.
Bien qu'exploitant des primitives plus riches, ces approches reposent sur les algorithmes k-means/GLA, EM qui présentent les inconvénients évoqués au chapitre 2. De plus, l'usage de primitives de natures différentes (typiquement couleur et texture) pose le problème de leur combinaison efficace en une seule primitive.

Une approche différente de celles citées est utilisée dans le système Netra [24, 63]. Elle repose sur l'approche de diffusion d'énergie "Edgeflow" [64, 61] pour

la détection de contours utilisant des primitives de couleur et texture. Conçue dans un but de détection précise de contours, cette méthode produit des régions satisfaisantes mais trop fine pour notre approche et au prix d'un temps de calcul élevé.

Parmi les approches régions, celle proposée récemment par Chen et Wang (*Unified Feature Matching* [17]) qui soulève le problème de la difficulté d'obtenir des régions satisfaisantes par segmentation. Leur solution se rapproche d'une segmentation par classification de primitives couleur/texture dans laquelle serait omise l'étape de construction des régions comme ensemble de pixels connexes. A l'issue de la classification, une région est définie comme l'ensemble des pixels appartenant à la même classe de primitives couleur/texture. Chaque région, dont les pixels ne sont alors pas nécessairement connexes, est caractérisée par un descripteur de couleur/texture et un autre de "forme". La question se pose alors de la pertinence de la caractérisation de la "forme" d'une région si elle composée par exemple de 2 zones non-connexes qui n'ont aucun lien entre elles (par exemple un ballon rouge et un camion rouge dans une même image). Cependant, le problème de cet aspect est moins gênant pour leur approche dans la mesure où ils se situent en effet dans un scénario de recherche *globale* d'images. La similarité entre deux images est définie à partir de la similarité entre toutes les régions des deux images.

Pour le paradigme qui nous intéresse de recherche par région exemple, la cohérence visuelle des régions détectées est importante. Il ne s'agit pas d'aspirer à une segmentation sémantique d'objets (au sens de "object segmentation" [102]). Nous souhaitons plutôt obtenir des régions qui correspondent au mieux aux composantes visuellement significatives pour l'utilisateur. Ces pourront doivent constituer des clés pertinentes pour les requêtes visuelles.

Contrairement aux approches que nous venons d'évoquer, la méthode de segmentation que nous allons présenter se base sur une primitive *unique* qui intègre une information photométrique *riche* (LDQC) et l'algorithme de regroupement CA qui présente l'avantage de déterminer automatiquement le nombre de régions (voir section 2.4). Cette méthode visera à produire une détection *grossière* des régions d'image.

3.2 Notre approche

Dans ce paragraphe, nous allons expliquer la motivation pour une détection *grossière* de régions dans le cadre de la recherche d'images par le contenu. Nous introduirons ensuite l'approche de segmentation pour la détection grossière.

Détection grossière

Dans le contexte de la recherche d'images, les enjeux de la segmentation pour la détection de régions diffèrent de nombreux problèmes de segmentation rencontrés en vision par ordinateur. Dans ce dernier cas, le but est d'estimer le plus précisément possible les contours des objets pour des problèmes, par exemple, de recalage d'images. Quant au problème de recherche d'images, nous ne chercherons pas à effectuer de la "reconnaissance d'objets" dans chaque image de la base. En effet, il est important de noter qu'il n'est pas nécessaire qu'une région détectée corresponde à un objet entier pour retrouver des objets similaires. Il s'agit de détecter les régions qui constituent des *clés de requête* pertinentes pour l'utilisateur. Par exemple, pour rechercher des images de tigres dans différents contextes (forêt, brousse), il suffit d'effectuer la recherche sur une zone qui couvre une partie visuellement caractéristique du tigre pour retrouver d'autres tigres.

Les régions que nous cherchons à détecter doivent donc, d'une part, correspondre aux zones visuellement *cohérentes et saillantes* de chaque image susceptibles d'être sélectionnées par l'utilisateur comme requête. D'autre part, elles doivent couvrir des zones suffisamment *caractéristiques* pour permettre une recherche par le contenu efficace. Les régions doivent donc présenter une certaine variabilité photométrique et correspondre aux zones principales. C'est pour cette raison qu'il sera nécessaire par la suite de proposer une *description fine des régions* pour représenter leur variabilité photométrique (chapitre 4). Les détails fins seront ignorés ou intégrés au sein de régions adjacentes plus importantes. Afin d'obtenir de telles régions, nous allons proposer une méthode de **détection grossière de régions**.

Choix de la primitive

Le contenu des images étant très varié, il n'est a priori pas possible d'exploiter de connaissance d'un domaine donné. La nature composite de ces images présente des niveaux de détails variables et l'homogénéité photométrique des zones saillantes de l'image doit donc être mesurée au niveaux des *voisinages* de pixels plutôt qu'au niveau de chaque pixel. Pour une détection grossière, nous estimons que l'information photométrique contenue dans la primitive doit être plus riche que pour une détection de nombreuses régions très homogènes. Ceci écarte les techniques n'exploitant que l'information couleur ponctuelle de chaque pixel qui s'apparentent à la quantification couleur. Il nous semble nécessaire d'exploiter l'information photométrique des voisinages de l'image pour détecter des régions aussi bien texturées qu'uniformes à l'aide d'*une seule primitive intégrée*. On évite ainsi le problème de la combinaison de primitives de natures différentes posé par les descripteurs mixtes cités précédemment. Par ailleurs, la particularité

du problème de recherche d'images par régions est que chaque région isolée dans l'image sera destinée à être comparée aux milliers de régions extraites des autres images de la base. Nous souhaitons donc que la primitive utilisée lors de la segmentation soit cohérente avec celle utilisée pour la description et la comparaison des régions de la base.

Regroupement de primitives

La segmentation et le regroupement sont deux domaines étroitement liés. Le choix de la méthode de regroupement de primitives influence directement la qualité de la segmentation. Il est nécessaire d'utiliser une méthode performante de regroupement, comme nous l'avons vu au chapitre précédent.

Rapide et non-supervisée

Le grand nombre d'images, donc de régions, traitées nécessite que l'algorithme de segmentation soit *rapide* et *non-supervisé*. Mais la rapidité ne doit pas être synonyme de qualité médiocre : les régions obtenues doivent être significatives pour l'utilisateur.

Ces considérations sur les spécificités du problème de recherche d'images par régions nous conduisent à proposer l'approche de segmentation basée sur la **classification des distributions locales de couleur (les "LDQC")**. Les LDQC sont les Distributions Locales de Couleurs Quantifiées (ou *Local Distribution of Quantized Colors*, en anglais). Nous allons expliquer comment elle permet de répondre aux problèmes posés.

Nous commencerons par présenter les primitives LDQC qui permettent de caractériser les voisinages des images, puis la mise en oeuvre de leur regroupement. Suivra l'étape d'intégration de l'information d'adjacence de régions afin d'améliorer la cohérence de la segmentation finale. Ce chapitre s'achèvera sur la présentation de résultats de segmentation et sur une discussion sur son efficacité.

Le déroulement général de la méthode de segmentation est présenté à la figure 3.1.

3.3 Extraction de primitives locales LDQC

Notre approche de segmentation est basée sur la classification CA des distributions locales de couleurs de l'image. Ces primitives intègrent naturellement la diversité des couleurs dans les voisinages des pixels. La classification de telles

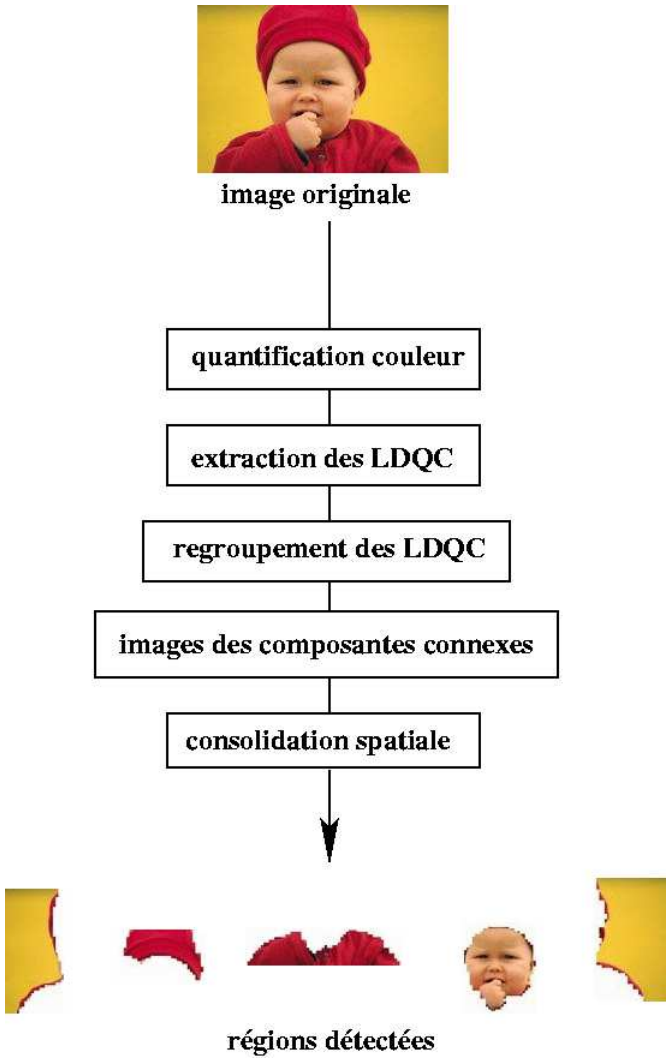


FIG. 3.1 – Résumé du déroulement de la méthode de segmentation.

primitives représentatives de larges voisinages donne des régions cohérentes plus naturellement. En effet, la formation de régions par regroupement de larges voisinages similaires conduit à des régions naturellement plus grossières et nécessite moins de post-traitement spatial. Ces primitives sont cohérentes avec le descripteur couleur ADCS qui sera présenté au chapitre 4.

Le choix de l'ensemble de couleurs pour calculer ces distributions locales de couleurs est crucial : il doit être compact pour gagner en vitesse de classification et représentatif d'un voisinage de pixels. Les images traitées comportent, en moyenne, 60.000 couleurs différentes, alors qu'environ une trentaine peut être suffisante avec la quantification proposée. Nous définissons cet ensemble de couleurs comme l'ensemble adaptatif obtenu par quantification couleur de l'image pour réduire considérablement le nombre de couleurs à traiter sans perte importante de l'information perceptuelle. Puis, pour chaque voisinage de l'image, la distribution locale est évaluée sur cet ensemble de couleurs réduit et pertinent.

Dans les méthodes de segmentation citées précédemment, les primitives sont soit simplement les pixels couleurs, qui posent problème pour la détection de zones texturées, soit des vecteurs mixtes de couleur et texture qui imposent une combinaison imparfaite de deux informations de natures différentes. Contrairement à ces primitives, les LDQC contiennent en elles-mêmes l'information locale de variabilité couleur permettant de détecter aussi bien des zones uniformes que texturées.

Après classification, chaque distribution LDQC sera remplacée dans l'image par l'étiquette de sa classe. Les étiquettes connexes formeront les régions et les petites régions seront fusionnées ou supprimées à l'aide d'un graphe d'adjacence de régions.

En résumé, les quatre phases principales de la segmentation sont les suivantes :
- quantification couleur (produit les *couleurs quantifiées*)
- extraction des distributions locales de couleurs quantifiées (*LDQC*)
- regroupement des LDQC similaires
- fusion et suppression des petites régions

Quantification couleur de l'image

Cette étape de quantification couleur de l'image vise à réduire le nombre de couleurs sur lesquelles les distributions locales (LDQC) vont être déterminées, tout en conservant suffisamment d'information colorimétrique pour distinguer les régions saillantes les unes des autres.

Différentes approches de quantification couleur ont été proposées qui varient en complexité de calculs et en précision (voir [97]). L'une des méthodes les plus usitées est l'algorithme *GLA*, dit aussi *LBG*. Dans le chapitre précédent, nous avons vu que l'algorithme CA présente, par rapport à GLA/LBG, les avantages

suivants : détermination automatique du nombre de classes (donc de couleurs ici) et réduction du problème de convergence vers un optimum local. L'algorithme CA est donc utilisé ici pour effectuer la quantification couleur de l'image par regroupement des pixels couleur.

Définition 2. Les **couleurs quantifiées de l'image** (ou "Quantized Colors") sont définies comme les prototypes de classes obtenus par regroupement CA des pixels couleurs de l'image.

Concernant l'espace de représentation des couleurs, la classification reposant fortement sur la métrique, il est nécessaire de choisir un espace perceptuellement uniforme, tels que Luv et Lab [97] [118]. Ces espaces ont été conçus de sorte que les différences de couleurs jugées égales par l'oeil humain soient aussi égales en distance euclidienne L^2 dans ces espaces. Par contre, les espaces RGB et Hsv ne sont pas perceptuellement uniformes [92]. Le modèle Hsv est intuitif, mais souffre d'irrégularités (la composante H cyclique et H et v n'ont pas de sens pour s petit). RGB présente généralement l'avantage de ne nécessiter aucune transformation, mais sa topologie ne rend pas compte des distances couleurs perçues par l'oeil humain. C'est l'espace Luv qui est adopté ici dont la transformation est donnée dans [118].

L'usage de la distance L^2 dans le regroupement des couleurs produit des classes hypersphériques dans l'espace couleur. Afin de tenir compte de dégradés de couleur et d'intensité, la distance de Mahalanobis [40] est utilisée pour détecter de manière plus générale les classes de couleur hyperellipsoïdales.

La quantification est donc obtenue par regroupement des pixels couleurs avec CA et la distance de Mahalanobis dans l'espace Luv. Le déroulement de l'algorithme CA est détaillé en 2.4.2.

L'initialisation des prototypes couleur dans l'algorithme CA est effectuée à partir d'une sélection des couleurs présentes en différents sites de l'image. Ces couleurs appartiennent aux données que nous cherchons à classifier et sont éloignées les unes des autres dans l'image car choisies sur une grille (voir figure 3.2). Cette heuristique d'initialisation constitue une approximation raisonnable pour la détermination des couleurs quantifiées finales. Comparée, par exemple, à une initialisation aléatoire, cette heuristique permet à CA de converger plus rapidement et plus certainement vers l'optimum global de la partition des couleurs.

La granularité de la classification (voir section 2.4) a été choisie de sorte que les grandes zones d'images avec une forte texture soient représentées par plus d'une couleur. A la convergence de la classification, les prototypes des différentes classes obtenues définissent l'ensemble des n couleurs quantifiées. Comme CA détermine automatiquement le nombre de classes, le nombre de couleurs quantifiées n sera représentatif de la diversité couleur des images naturelles. La figure 3.3 illustre le résultat de la quantification de l'image "enfant". Bien que l'ensemble de

FIG. 3.2 – Lieux des pixels définissant les prototypes couleur initiaux. Ces prototypes sont les pixels situés aux intersections intérieures d'une grille 7×7 afin de couvrir les différents lieux de l'image (à l'exception des bords de la grille). Nous définissons ici 36 prototypes initiaux.

couleurs de l'image soit considérablement réduit (de 43.217 à 27), l'image quantifiée de l'enfant conserve l'essentiel de l'information colorimétrique nécessaire à la segmentation en régions.

FIG. 3.3 – A gauche : l'image originale (43.217 couleurs). A droite : l'image quantifiée (27 couleurs)

Extraction des LDQC

La quantification couleur de l'image fait déjà apparaître visuellement des zones saillantes. Mais ces zones ne peuvent pas être extraites automatiquement à ce stade, car elles comportent différentes couleurs quantifiées dès qu'elles sont texturées. Afin de saisir les caractéristiques visuelles des zones (uniformes ou texturées), il est nécessaire d'extraire des primitives sur des *voisinages*. Les Distributions Locales de Couleurs Quantifiées (LDQC) vont permettre de détecter les variations locales des couleurs dans l'image quantifiée et le regroupement des LDQC de l'image formera la segmentation en régions.

Pour extraire toutes les LDQC de l'image, nous déplaçons une fenêtre F_k sur les pixels de l'image quantifiée et évaluons la distribution locale de couleurs quantifiées au sein de la fenêtre.

Définition 3. Etant donné l'ensemble des couleurs quantifiées $\{q_i\}$ de l'image, en chaque fenêtre F_k de l'image quantifiée, la **LDQC** est l'histogramme local $LDQC_k$ défini par :

- $LDQC_k = \{LDQC_k(q_i)\}$, où
- $\forall q_i, LDQC_k(q_i) =\mid \{(x,y) \in F_k / couleur(x,y) = q_i)\} \mid$

La figure 3.3 montre trois exemples de LDQC sur une image en des lieux de couleurs et de variations différentes. Pour une image 500×400, la largeur de fenêtre est de 31 pixels et le pas de déplacement, défini comme une demi-largeur de fenêtre, vaut 16 pixels. Le nombre de LDQC extraits de cette image se calcule ainsi : $(500/16) \times (400/16)=31 \times 25=775$. La surface de fenêtre S_F définit la résolution spatiale de la segmentation : plus elle est grande et plus les motifs extraits seront grands. Définissant le niveau minimum de détail détecté par notre segmentation, une largeur de fenêtre r_f 31 pixels a été jugée satisfaisante pour des images de 500×400 pixels. Dans le cas général, cette largeur de fenêtre est déterminée proportionnellement à la taille de l'image.

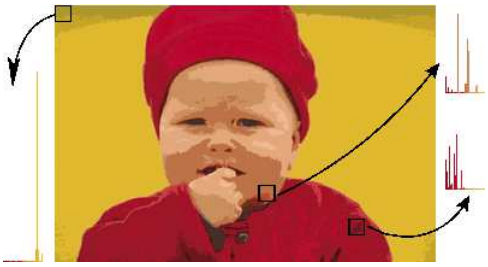


FIG. 3.4 – Illustration des LDQC : exemple de calcul sur trois fenêtres. Nous observons que plus le voisinage est uniforme et plus les LDQC sont piquées.

3.4 Regroupement des primitives LDQC

Les LDQC extraites dans l'image vont être regroupées par classification avec CA. La classification CA dépend notamment de la définition d'une distance dont la capacité à mesurer les similarités entre les LDQC influera sur la qualité de la segmentation. Notons que la particularité des LDQC est de reposer sur une quantification des couleurs adaptative à l'image.

En section 2.1, nous avons présenté une synthèse des principales distances permettant de mesurer la similarité entre les distributions de couleurs. Nous y justifions notre choix sur la distance couleur quadratique [41], qui est adaptée à différents types de distributions couleur en intégrant la similarité couleur inter-cellules. Nous rappelons ici son expression dans sa forme d'origine. On définit X

et Y deux distributions de couleurs LDQC sur les n couleurs quantifiées. On peut les écrire comme paires de couleur/pourcentage de population :
$X = \{(c_1, p_1^X), ..., (c_n, p_n^X)\}$ et $Y = \{(c_1, p_1^Y), ..., (c_n, p_n^Y)\}$.
Alors la distance quadratique est définie par la forme quadratique suivante :

$$\begin{aligned} d_q(X,Y)^2 =& (X-Y)^T A (X-Y) \\ =& \sum_{i=1}^{n}\sum_{j=1}^{n}(p_i^X - p_i^Y)(p_j^X - p_j^Y)a_{ij} \end{aligned}$$

où $A = [a_{ij}]$ est la matrice de similarité couleur a_{ij} entre les couleurs c_i et c_j :

$$a_{ij} = 1 - d_{ij}/d_{max} \tag{3.1}$$

où d_{ij} est la distance euclidienne dans l'espace Luv et d_{max} le maximum de cette distance dans l'espace couleur.

L'algorithme de classification CA tel qu'il est présenté à la section 2.4 est utilisé pour classifier les LDQC extraites de l'image avec la distance quadratique. La bonne initialisation des prototypes initiaux participe à la qualité de la segmentation puisque ces prototypes convergeront vers les prototypes finaux. Pour la classification des LDQC, comme pour la quantification couleur que nous venons de voir, les prototypes LDQC sont initialisés à partir des données elles-mêmes en considérant des LDQC en différents lieux de l'image. Dans la figure 3.5, la grille 6×6 définit les lieux où sont choisis les 25 prototypes LDQC initiaux. Cette initialisation à partir des données elles-mêmes, vise de nouveau à faciliter la convergence vers l'optimum global de la partition des données.

FIG. 3.5 – Lieux des LDQC définissant les prototypes initiaux. Ces LDQC-prototypes sont choisies comme celles centrées aux intersections intérieures d'une grille 6×6 afin de couvrir les différents lieux de l'image (à l'exception des bords de la grille). Nous avons donc ici 25 prototypes initiaux.

A la convergence de la classification, on obtient un ensemble de classes de LDQC ainsi que leurs prototypes. L'image segmentée est obtenue en attribuant

à chaque pixel l'étiquette de la classe à laquelle appartient la LDQC centrée en ce pixel. Un filtre de vote par focalisation graduelle [9] est ensuite appliqué pour supprimer les étiquettes isolées dans l'espace image.

3.5 Consolidation spatiale

A ce point, nous avons obtenu une partition complète de l'image en régions adjacentes et homogènes en primitives LDQC. Quelques régions peuvent s'avérer trop petites pour constituer des régions d'intérêt ; elles augmentent alors inutilement le nombre total de régions dans la base. De plus, dans des scènes complexes, elles sont souvent situées à la frontière entre deux grandes régions ou à l'intérieur d'une grande région. Elles doivent être fusionnées pour améliorer la topologie des régions d'intérêt.

Nous désirons que chaque région couvre au minimum 1,5% de la surface de l'image. En dessous de ce seuil, une région est fusionnée à sa région voisine la plus proche visuellement si elle en a une. Deux régions sont dites "proches visuellement" si la moyenne des LDQC qu'elles contiennent sont proches, c'est-à-dire si la distance quadratique entre deux LDQC moyennes est inférieure à un seuil fixé. La stratégie itérative de fusion des petites régions est la suivante : tant qu'il existe une région de surface inférieure à 1,5% et possédant une région voisine "proche visuellement", on fusionne la petite région dans sa voisine. A l'issue des itérations, s'il reste des petites régions (i.e. qui n'ont pas été jugées suffisamment proches de leurs voisines pour être fusionnées), elles sont supprimées et ne sont pas indexées.

La fusion de régions s'opère à l'aide d'un Graphe d'Adjacence de Régions [84] (ou RAG pour *Region Adjacency Graph*). Les attributs des régions sont stockés dans les noeuds et l'information d'adjacence de régions dans les arêtes du graphe (adjacence, longueur des contours communs). Les attributs géométriques de régions (surface, position et compacité en particulier) seront utilisés comme compléments de descripteurs géométriques pour la recherche par régions (voir 4.5).

Le codage des images segmentées est détaillé en annexe A.

3.6 Résultats

Nos tests de détection de régions ont été effectués sur une base issue des photothèques Corel[1] et IDS [2]. Elle comporte des scènes de nature très différente : pay-

[1] http ://www.corel.com
[2] http ://www.imagedusud.fr

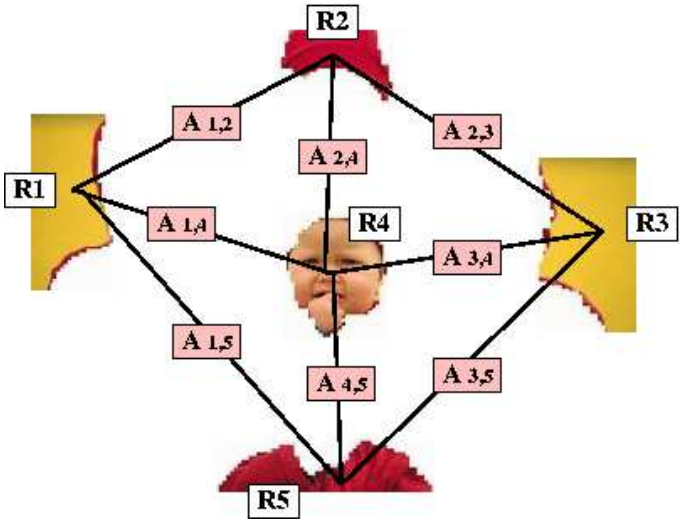


FIG. 3.6 – Structure RAG de l'image partitionnée. Le RAG représente les attributs des régions R_i dans les noeuds et les informations d'adjacence entre toutes les paires (i, j) de régions adjacentes dans les arêtes du graphe. La segmentation exploite ces informations pour la fusion des régions.

FIG. 3.7 – Image segmentée (chaque région est illustrée par sa couleur moyenne)

sages, portraits, objets, dessins, peintures, architecture, jardins, animaux, fleurs,...

Le processus de segmentation est rapide (une moyenne de 2.1 secondes par image) ce qui est convenable pour de grandes bases d'images. 56.374 régions ont été automatiquement extraites des 11.479 images (on obtient une moyenne de 5.2 régions par image sur les bases Corel et IDS).

nombre d'images	11.479
nombre de régions	56.374
nombre de régions par image	5.2
temps de segmentation par image	2.1 s

Des exemples d'images segmentées sont présentés dans la figure 3.8. De plus nombreux exemples ont été regroupées en annexe B.

Même dans les scènes complexes, les régions extraites présentent une cohérence visuelle et sont généralement intuitives pour l'utilisateur. La segmentation grossière a montré sa capacité à intégrer au sein des régions des zones formées de différentes nuances d'une même teinte, de fortes textures, de détails spatiaux isolés. Une telle variabilité perceptuelle rend chaque région plus spécifique vis à vis des autres régions de la base. Les régions supprimées (petites et grisées dans les exemples de la figure 3.8) représentent une très faible proportion des surfaces d'images.

3.7 Perspectives

Traitements des cas difficiles

La détection a été définie pour extraire des régions grossières. Les granularités photométrique (lors du regroupement des distributions LDQC) et spatiale (choix de la surface de motif minimum, fusion de régions adjacentes dans les régions) ont été choisies dans le processus de segmentation dans le but de détection des régions saillantes visuellement. Ce choix s'est montré efficace pour la majorité des images traitées dans nos bases, cependant il existe toujours des cas d'images très difficiles à segmenter (même pour un humain) dont le contenu visuel est complexe. Pour résoudre les cas de segmentation difficile, on pourrait envisager, par exemple, d'ajouter un degré d'interactivité au moment de la requête pour que l'utilisateur définisse ou raffine dynamiquement les régions d'intérêt, par une représentation multi-échelle de la segmentation en régions, par exemple. Cependant, la conséquence serait inévitablement un surcoût calculatoire important, de plus, la méthode ne serait plus complètement automatique.

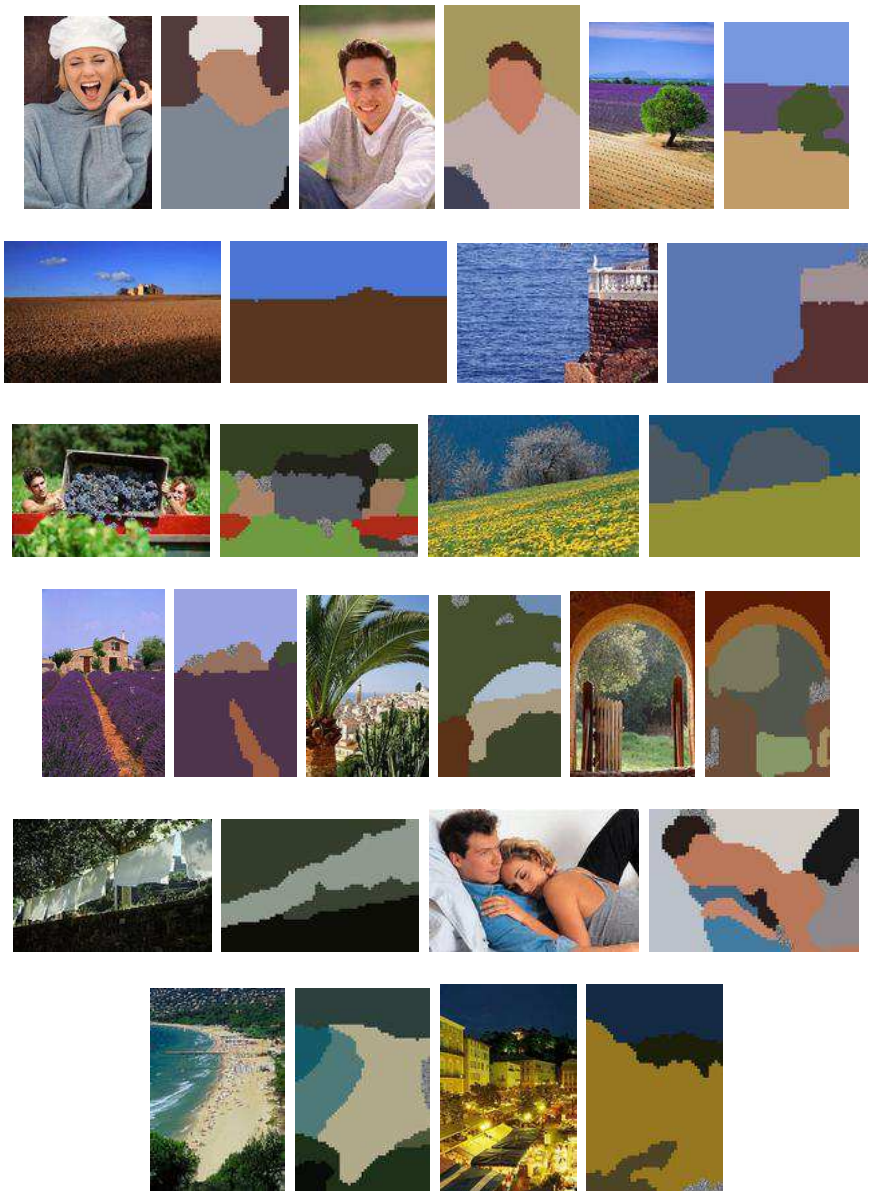


FIG. 3.8 – Illustration de la segmentation grossière. Chaque image originale est suivie de l'image des régions détectées représentées par leur couleur moyenne. Les petites régions ignorées sont grisées. Plus d'exemples sont montrés en annexe B.

Accélération de la méthode

Le regroupement par CA des distributions LDQC avec la distance quadratique est l'étape la plus critique en temps de calcul. Une solution consisterait à simplifier l'expression de cette distance en diagonalisant la matrice A afin de ramener le calcul à une distance L^2 pondérée entre les LDQC.

Classification non-supervisée

La classification non-supervisée par CA intervient dans deux étapes du processus de segmentation (quantification couleur et regroupement des LDQC). Dans le chapitre précédent, nous avons étudié les avantages de l'agglomération compétitive par rapport aux algorithmes largement utilisés dans la littérature pour la quantification vectorielle et plus généralement le regroupement de données. Nous avons aussi évoqué l'algorithme ARC (Adaptive Robust Clustering) qui permet en plus la gestion du bruit et des densités variables des données. L'utilisation de cet algorithme offrirait des avantages certains pour la segmentation. Des travaux vont actuellement dans ce sens au sein de l'équipe IMEDIA [46].

3.8 Conclusions

Rapide et non-supervisée, la méthode de segmentation présentée repose sur la classification CA des primitives LDQC qui caractérisent la variabilité locale de la couleur. Les avantages de CA sur les algorithmes existants permet de détecter plus efficacement le bon nombre de régions d'une image.

Les régions détectées par cette méthode constituent les composantes des images de la base qui sont visuellement saillantes et intuitives pour l'utilisateur. La détection étant grossière, les régions englobent une variabilité photométrique qui les rend plus caractéristiques les unes des autres dans la base.

Chapitre 4

Description fine de régions pour le Paradigme 1

Dans ce chapitre, nous allons nous intéresser au **paradigme 1** de la recherche d'images par régions : celui de la recherche par région exemple. Dans ce scénario, l'utilisateur peut sélectionner explicitement une partie d'image préalablement définie et retrouver automatiquement les images comportant une partie d'apparence visuelle similaire.

Le problème posé est celui de la similarité d'apparence visuelle entre une région requête et les régions de la base. Ce paradigme utilise les régions détectées par la méthode de détection grossière présentée au chapitre précédent. Les régions identifient les zones saillantes de chaque image de la base. Par construction, elles comportent une certaine variabilité du contenu visuel qui les rend caractéristiques les unes des autres dans la base.

La similarité visuelle entre deux régions repose sur deux aspects étroitement liés : celui de la caractérisation visuelle individuelle de chaque région de la base, appelée *description*, et celui de la mesure de similarité entre les descripteurs associés à toute paire de régions de la base. La pertinence des régions retournées dépend directement de ces deux aspects que nous abordons dans ce chapitre.

Dans un premier temps, nous introduirons l'intérêt de ce paradigme 1 en termes d'usage, mais aussi par rapport au problème plus général de la recherche par régions. Ensuite nous proposerons le descripteur ADCS pour caractériser *finement* la variabilité couleur d'une région ainsi qu'une mesure de similarité adéquate à cette finesse de description. Finalement, les résultats de recherche de régions seront présentés sur une base d'images de type photothèque sur la plate-forme IKONA. Nous verrons que la combinaison de la détection grossière de régions combinée à la description fine de l'apparence visuelle est une approche

privilégiée pour ce paradigme de recherche.

4.1 Le paradigme 1

L'intérêt de l'étude de ce paradigme est double.

D'un point de vue des besoins de l'utilisateur, il permet de répondre à la requête du type : "retrouver les images de la base comportant une zone d'apparence visuelle similaire à celle-ci, indépendamment du reste de l'image". Un tel paradigme correspond à un besoin concret pour rechercher des gens (à partir de zones de peau), des scènes d'extérieur (à partir d'une zone de ciel), des zones de végétations pour ne citer que quelques exemples évidents. De façon plus générale, ce paradigme est utile dans toute base d'images composites dans laquelle l'utilisateur souhaite retrouver des parties d'images dont l'apparence visuelle est pertinente pour la requête. Comme nous le verrons, la possibilité d'interaction dans l'interface graphique peut permettre à l'utilisateur de moduler la similarité visuelle en fonction de l'objet de sa recherche. Dans ce contexte, il est important de garder à l'esprit que les procédés de représentation et de similarité des régions sont totalement non-supervisés et n'intègrent pas d'informations de nature sémantique, mais uniquement fondées sur l'*apparence visuelle*.

Du point de vue des avancements dans le domaine de la recherche d'images par le contenu, ce paradigme peut être considéré comme le *paradigme canonique* de la recherche d'images par régions. Plus précisément, il pose les problématiques élémentaires de tout système de recherche utilisant les régions : à savoir, d'une part, la détection de régions (chapitre précédent), et d'autre part leur description et leur similarité visuelle (chapitre présent). A partir d'une analyse automatique du contenu des images, ces problèmes sont non-triviaux et il est donc important de les étudier de manière approfondie. L'évaluation des techniques visant à répondre à ces problèmes doit passer par l'implantation effective de ce paradigme dans un système : sélection par l'utilisateur d'une région requête, comparaison entre le descripteur de celle-ci et les descripteurs des autres régions de la base et finalement, visualisation des contours des régions similaires[1]. Il est alors possible de mieux comprendre les enjeux de la représentation d'images en régions, ses limitations, sa capacité à répondre aux attentes de l'utilisateur. Ces considérations permettent d'améliorer et de valider les techniques adoptées qui pourront servir de base à des approches plus élaborées telles que, par exemple, le bouclage de pertinence sur les régions, l'association de mots-clés aux régions, la catégorisa-

[1]Dans les systèmes existants de recherche par régions, les contours des régions retrouvées sont rarement présentés dans l'interface graphique, alors qu'ils sont nécessaires à l'utilisateur pour formuler sa requête et apprécier les résultats.

tion de régions d'une base, les requêtes sur la composition logique de catégories de régions (voir paradigme 2 présenté au chapitre 5).

Notons que la plupart des contributions existantes porte sur des aspects particuliers de la recherche par région en délaissant les autres. Pour la sélection partielle de régions, par exemple, des approches préconisent le simple usage de blocs d'images ou bien des techniques de segmentation peu performantes. Concernant la description, très peu de systèmes posent réellement la question de la spécificité photométrique d'une région et dérivent directement les descripteurs utilisées pour comparer des images entières.

Dans notre approche, nous nous penchons individuellement sur chacune des problématiques d'extraction, de description et de comparaison de régions pour satisfaire au mieux à l'implantation du "paradigme canonique".

4.2 L'existant pour la description de l'apparence visuelle

Nous nous intéresserons en particulier au problème de la caractérisation de la variabilité de la couleur. Il s'agit de la primitive la plus discriminante perceptuellement pour les bases généralistes (de type photothèque) sur lesquelles nous travaillons. Notons que de nombreux travaux ont par ailleurs porté sur la description de la texture (modèles MRSAR [69][111], *Wold features* [59], chaînes de Markov cachées [117], ondelettes de Daubechies [20], bancs de filtres de Gabor [67], matrices de co-occurrence [42], descripteurs de la norme MPEG7 [68]).

En recherche d'image par le contenu, les travaux sur la description *globale* de l'apparence visuelle des images sont plus anciens et plus avancés que pour la description de régions. Dans cette section, nous commencerons donc par présenter un aperçu de l'existant pour la description *globale* et ensuite pour la description de régions. Finalement, nous mettrons en avant les limitations dans la représentation de la couleur dans les descripteurs existants et nous exposerons les motivations pour une nouvelle approche de description fine et adaptative des régions.

L'existant pour la description globale des images

Les premiers travaux de description visuelle pour la recherche d'images par le contenu ont été présentés en 1991 par Swain et Ballard [110] avec l'histogramme couleur. L'article de Smeulders et al. [102] offre un aperçu des familles de descripteurs proposés jusqu'en 2000. Récemment aussi, l'élaboration de la norme MPEG7 [68], dédiée au cadre plus général de la description de documents, préconise une large variété de descripteurs visuels.

Le pouvoir discriminant de la couleur pour comparer l'apparence visuelle des images a été mis en avant par Swain et Ballard [110]. Ils proposèrent l'histogramme couleur pour en décrire le contenu. L'accumulation des couleurs des pixels d'une image constitue un descripteur très simple et robuste à la translation et à la rotation de la scène étudiée. L'histogramme couleur constitue la base des descripteurs couleur utilisés aujourd'hui à partir d'une quantification systématique de l'espace couleur. Quant aux descripteurs de régions, les plus évolués reposent sur ce type d'histogramme.

C'est pour la description *globale* d'images que des variantes de l'histogramme couleur ont été proposées. Elle ont majoritairement porté sur l'intégration de l'information spatiale des couleurs. Une famille d'approches exploite la disposition spatiale des pixels pour chaque couleur quantifiée de l'histogramme. L'*autocorrélogramme* représente le nombre d'occurrences de paires de pixels ayant des couleurs (quantifiées) données et à distance donnée l'un de l'autre [47]. Pour chaque couleur quantifiée, les *color coherent vectors* représentent séparément le nombre de pixels appartenant à une composante connexe de taille supérieure à un seuil donné et de ceux appartenant à une composante de taille supérieure au seuil [82]. Les *color tuples* comptent le nombre d'occurrences de pixels ayant une configuration triangulaire donnée [90]. Les *spatial color histograms* ou *color density maps* accumulent les pixels de même couleur appartenant à une même zone (en forme d'anneau ou de secteur) [88]. Les *geometric histograms* sont présentés dans [89] comme une unification des approches précédentes dans la mesure où elles consistent à accumuler, pour chaque couleur quantifiée, les pixels présentant une configuration géométrique donnée. Présenté dans [73] et adopté dans MPEG7 [74, 68], le *colour structure descriptor* compte, pour chaque couleur quantifiée, le nombre d'éléments structurants (typiquement un voisinage 3×3) contenant un pixel de cette couleur. Le *joint histogram* est un histogramme couleur auquel une ou plusieurs autres dimensions sont ajoutées pour caractériser une mesure locale telle que texture, gradient, forme. Une cellule de cet histogramme multidimensionnel contient le nombre de pixels ayant la même couleur et la même valeur de propriété locale [83]. Les *weighted histograms* consistent à pondérer la participation de chaque pixel dans la cellule de couleur correspondante, non par 1 comme dans l'histogramme classique, mais par une mesure caractéristique du voisinage du pixel (le Laplacien par exemple) [11]. Nous évoquerons une dernière variante intéressante de l'histogramme couleur, le *fuzzy histogram* [12, 113], dans lequel chaque pixel participe à toutes les cellules de l'histogramme classique avec une pondération correspondant à la similarité entre la couleur du pixel et celle de chaque cellule.

Notons que l'histogramme classique peut être vu comme un cas particulier pour chacune de ces variantes.

Ces variantes portent sur la repartition spatiale des couleurs dans l'image,

mais reposent sur une représentation *imprécise* des couleurs. En effet, ces distributions sont calculées à l'issue d'une *quantification systématique* de l'espace couleur basé sur un seuillage selon chaque composante (voir section 2.3.1). A notre connaissance, seuls deux travaux ([95] et [55]) utilisent une quantification couleur *adaptative à chaque image*.

L'existant pour la description des régions

Dans cette partie, nous résumons les descripteurs couleur qui ont été proposés pour la recherche d'images par région. Ils sont majoritairement inspirés de l'histogramme couleur classique qui a été initialement proposé pour la description globale d'image.

La couleur moyenne des pixels d'une régions constitue la caractérisation la plus simple de la couleur (MARS [48], SIMPLIcity [56], FRIP [51], Kompatsiaris et al. [53]). Elle peut être justifiée sur des régions très homogènes en couleur (dans le cas d'une sur-segmentation) ou si le but recherché est la simplicité de description.

La caractérisation par *plus d'une* couleur par région repose sur l'histogramme et plus généralement sur le concept de "couleurs dominantes". Dans [50], [106] et dans le système Blobworld [14], les régions sont décrites par des histogrammes sur une quantification systématique des espaces couleur considérés (respectivement : 216 cellules dans RGB, 166 dans HSV et 218 dans LAB). Dans le cas de Blobworld, le nombre de cellules est en fait ramené à 5 par projection dans un sous-espace. Dans le système VisualSeek, les régions sont décrites par les *binary color sets* qui peuvent être assimilés à des histogrammes ; cependant, l'implémentation effective n'utilise qu'une seule couleur pour décrire une région parmi une quantification systématique en 166 cellules de l'espace HSV [104].

La notion de couleurs dominantes peut être vue comme la sélection d'un sous-ensemble de couleurs dans une *quantification imprécise* de l'espace couleur. Deux couleurs dominantes au maximum sont extraites dans [44]. Dans Netra [24], le nombre de couleurs dominantes, choisies parmi une palette de 256 couleurs déterminée pour une base d'images donnée, est de 3.5 par région en moyenne. La norme MPEG7 propose de façon plus générale le descripteur DCD (pour *dominant colour descriptor*) dont le nombre maximal de couleurs dominantes est fixé à 8. Les couleurs dominantes présentent l'avantage de constituer un descripteur plus compact que l'histogramme couleur classique (seules les couleurs pertinentes sont considérées). Cependant elles sont peu nombreuses et sont toujours sélectionnées parmi une palette fixe de couleur.

Dans [38], les régions sont obtenues par classification non-supervisée des chrominances des pixels de l'image (composantes (a,b) de l'espace LAB). La similarité entre deux régions est alors définie comme l'intersection des volumes de leurs

classes associées. Dans Windsurf [2], la couleur est décrite, implicitement, par les coefficients moyens des ondelettes de Daubechies déterminées indépendamment sur chaque composante couleur.

Nécessité d'une description fine et adaptative pour les régions

Dans notre approche, les régions, issues de la détection grossière, comportent une variabilité colorimétrique. Afin de caractériser au mieux cette spécificité visuelle, nous proposons une nouvelle approche pour la description fine des régions.

Les descripteurs que nous avons cités, qu'il s'agisse des histogrammes ou des couleurs dominantes, reposent tous sur la définition d'une palette de couleurs intermédiaire. Cette palette comporte autour de 200 couleurs (selon les systèmes) et est utilisée pour représenter les millions de couleurs potentiellement présentes dans une région. Il s'agit donc d'une réduction d'information colorimétrique importante. La comparaison de distributions calculées sur une palette *commune* de couleurs permet d'utiliser, dans la phase de recherche, des distances de types "cellule-à-cellule" comme les distances L^p (voir section 2.1.1) qui sont simples à calculer.

Cependant, la représentation de la couleur dans un descripteur visuel doit reposer sur un ensemble *compact et pertinent* de couleurs représentatives. Pour décrire l'apparence visuelle de deux régions, il est donc naturel d'avoir plutôt recours à une *quantification couleur adaptative à chaque région*. Les couleurs quantifiées fournissent un ensemble compact de couleurs choisies parmi l'espace de couleur *entier*. Elles sont donc plus fidèles à chaque région que les 200 couleurs de la palette commune à toute la base de région. Le descripteur région résultant est alors la distribution de ces couleurs quantifiées. De telles distributions adaptatives nécessitent l'introduction de distances plus élaborées, comme nous allons le voir avec la forme généralisée de la forme quadratique.

De plus, l'élaboration d'un descripteur de régions doit tenir compte de la nature de la formation des régions au sein d'une image. D'une part, une image définit plusieurs régions, donc le nombre d'entrées dans la base augmente. D'autre part, une région étant formée par un critère d'homogénéité de primitives visuelles (les LDQC dans notre segmentation), la distribution couleur extraite d'une région sera plus "piquée" que celle extraite d'une image entière. Autrement dit, nous devons tenir compte du fait que, par rapport à la recherche globale d'images, *la recherche par régions (paradigme 1) nécessite de comparer plus de distributions de couleurs qui sont plus piquées*.

Dans ces conditions, le choix des couleurs pour représenter les distributions doit permettre de discriminer visuellement les régions de la base tout en produisant un descripteur compact. Ces couleurs, que nous nommerons "nuances de couleurs", seront extraites par *quantification fine et adaptative* des couleurs de

chaque région. Contrairement aux approches citées, l'ensemble de couleur ne sera ni commun à toute la base, ni à une image, mais propre à chaque région[2].

Ce nouveau descripteur (ADCS) est destiné à caractériser la variabilité photométrique des régions de façon *adaptative, fine et compacte*. Notre approche de **segmentation grossière et description fine des régions** pour la recherche d'images par région exemple repose d'une part sur la méthode de détection grossière présentée au chapitre précédent et d'autre part sur le descripteur ADCS.

4.3 ADCS : description fine, adaptative et compacte de la variabilité couleur

Nous introduisons le descripteur de région **ADCS**, pour *Adaptive Distribution of Colors Shades* ou *Distribution Adaptative de Nuances de Couleurs*.

Principe

L'extraction du descripteur ADCS repose sur la détermination d'un ensemble de couleurs pertinentes pour chaque région : les *nuances de couleurs* ou *color shades*. Elles sont obtenues par quantification adaptative des couleurs contenues dans chaque région par l'algorithme CA. Nous rappelons que l'avantage majeur de cet algorithme, comparé en section 2.4 à l'algorithme de quantification reconnu GLA, réside dans l'estimation automatique du nombre de classes, c'est-à-dire, du nombre de nuances de couleurs.

Définition 4. Soit R une région et $\{c_i\}$ l'ensemble des couleurs quantifiées de cette région. Le descripteur **ADCS** de la région R est l'histogramme $ADCS_R$ défini par :

- $ADCS_R = \{(c_i, p_i)\}$, où
- $\forall$ color shade c_i, $p_i = \mid \{(x, y) \in R / \text{ couleur}(x, y) = c_i)\} \mid$

Un descripteur ADCS est donc une distribution adaptative, c'est-à-dire constitué d'un ensemble de paires couleur/population (c_i, p_i) propre à chaque région.

Après une transformation dans l'espace Luv, choisi pour son uniformité perceptuelle, les pixels de chaque région sont regroupés avec CA ; la distance de

[2]En effet, l'usage des couleurs quantifiées d'une image entière pour décrire une région introduit un biais. La participation des couleurs d'une région dans la quantification de l'image entière est proportionnelle à la surface de la région. Ainsi les couleurs obtenues par une telle quantification sont plus représentatives des grandes régions que des petites. Cette propriété n'est pas désirable.

Mahalanobis est utilisée. Les prototypes sont initialisés à partir des données et leur nombre initialisé à 40. La granularité est définie finement : la population minimum de classe est fixée à un facteur de 0.005 par rapport au nombre total de pixels de la région. La distance minimum entre deux nuances de couleurs et fixée à 6 (distance jugée minimum pour la distinction de deux couleurs différentes dans l'espace LUV). Elle permet de fusionner les nuances inutilement proches.

Nous rappelons que cette quantification diffère de celle présentée pour le prétraitement de la segmentation en section 3.3 : la quantification est, ici, adaptative et plus fine. On opère autant de regroupements de pixels par classification CA qu'il y a de régions dans l'image.

Pour chaque région, les centres des classes obtenues à la convergence du regroupement définissent l'ensemble de nuances de couleurs propres à la région. Contrairement aux représentations de la couleur dans les descripteurs existants, les nuances de couleurs prennent leur valeur dans l'*ensemble* de l'espace considéré (LUV ici) et ne sont pas tributaires d'une palette de 200 couleurs prédéfinies pour toute la base. Par ailleurs leur nombre est estimé automatiquement et il illustrera la variabilité couleur de la région : par exemple, dans une région de ciel le nombre de nuances de couleur extraites sera nettement inférieur à une zone correspondant à une foule.

Dans une distribution ADCS, la population de chaque nuance de couleurc_i présente dans une région est obtenue à la convergence de CA : il s'agit des populations de classe N_i (voir formule (2.5)) calculée après la décision finale d'appartenance entre les données (pixels couleur) et les classes dont les prototypes définissent les nuances de couleurs. Le descripteur ADCS d'une région est la distribution de l'ensemble de ces nuances de couleurs et de leur population respective.

La finesse de description couleur par les *nuances de couleur* est le résultat de trois facteurs :

- les nuances de couleur sont choisies dans l'*espace couleur entier* (et non pas à partir d'une palette fixée de 200 couleurs)
- la quantification pour sélectionner les nuances de couleur est *adaptative à chaque région*
- la quantification est obtenue par une granularité *fine* de l'algorithme de classification (CA)

Représentation compacte

La syntaxe de l'index ADCS d'une région donnée est la suivante :

$$\text{index ADCS} : n, (l_1, u_1, v_1, p_1), ..., (l_n, u_n, v_n, p_n)$$

où n désigne le nombre de nuances de couleurs et chaque triplet (l_i, u_i, v_i) les composantes dans l'espace LUV de la i^e nuance de couleur. p_i désigne le nombre

de pixel ayant cette couleur i normalisé par rapport au nombre total de pixels de la région. L'index est donc un vecteur de $1 + 4n$ composantes. Un octet étant nécessaire à la représentation de chaque composante, l'index ADCS est représenté sur $1+4n$ octets. A titre de comparaison, si l'on utilise un octet pour représenter la valeur de chaque cellule d'un histogramme classique de 200 couleurs, nous verrons que l'index ADCS est en moyenne (car n dépend de la complexité couleur de la région) 3 fois plus compact que celui d'un histogramme classique.

Notons que cet index est "autonome" car il ne dépend pas d'une représentation sous-jacente de l'espace couleur : les composantes de chaque couleur sont stockées dans l'index même. L'ordre de représentation des nuances (l_i, u_i, v_i) dans l'index n'importe donc pas.

Interprétation

La figure 4.1 illustre les distributions ADCS extraites sur chacune des cinq régions extraites de l'image "enfant". Nous remarquons que les régions homogènes telles que le fond jaune, sont représentées par des distributions ADCS très piquées, avec un nombre réduit de nuances. Sur les trois autres régions plus texturées, les distributions ADCS sont moins piquées et comportent plus de nuances de couleur.

La figure 4.2 compare les distributions classiques (espace Luv partitionné en 6 valeurs par composantes) et les distributions ADCS d'une région mauve et texturée de lavande et d'une région de ciel présentant un dégradé de bleu. Les distributions classiques (au centre de la figure) sont quasiment identiques du fait d'un manque de discrimination entre les teintes bleues et mauves du fait de la quantification imprécise et systématique de l'espace couleur. Ce problème de discrimination est un obstacle à la bonne mesure de similarité visuelle des régions. A l'inverse, les nuances de couleurs des distributions ADCS sont plus *fidèles* au contenu visuel de chaque région et sont plus compactes.

4.4 Forme généralisée de la distance quadratique

Issue d'une quantification couleur adaptative au niveau région, une distribution ADCS représente un ensemble de couleurs dont le nombre et surtout la nature peuvent être très variables. Nous souhaitons comparer deux régions d'index ADCS X et Y caractérisés par n_X et n_Y nuances de couleurs $\{c_i^X\}$ et $\{c_j^Y\}$:

$$\text{index X} : n_X, c_1^X, x_1, ..., c_{n_X}^X, x_{n_X}$$
$$\text{index Y} : n_Y, c_1^Y, y_1, ..., c_{n_Y}^Y, y_{n_Y}$$

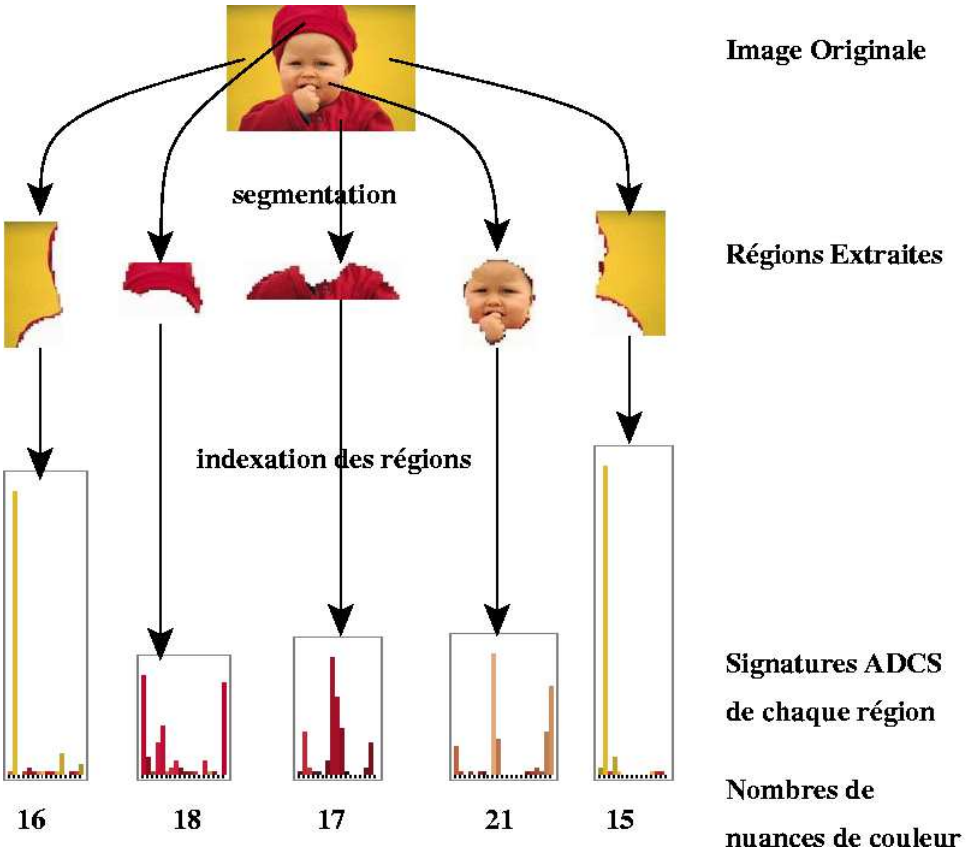


FIG. 4.1 – L'image originale, ses régions détectées et leurs index ADCS respectifs. Notons que l'ordre des couleurs dans la représentation des ADCS est sans importance. Nous remarquons que les zones texturées du bonnet et du manteau sont représentées par des nuances différentes de rouge et que le fond jaune quasiment uniforme est décrit par un pic majeur de jaune et par quelques autres couleurs mineures. L'ensemble des couleurs sélectionnées est pertinent visuellement pour chaque région associée.

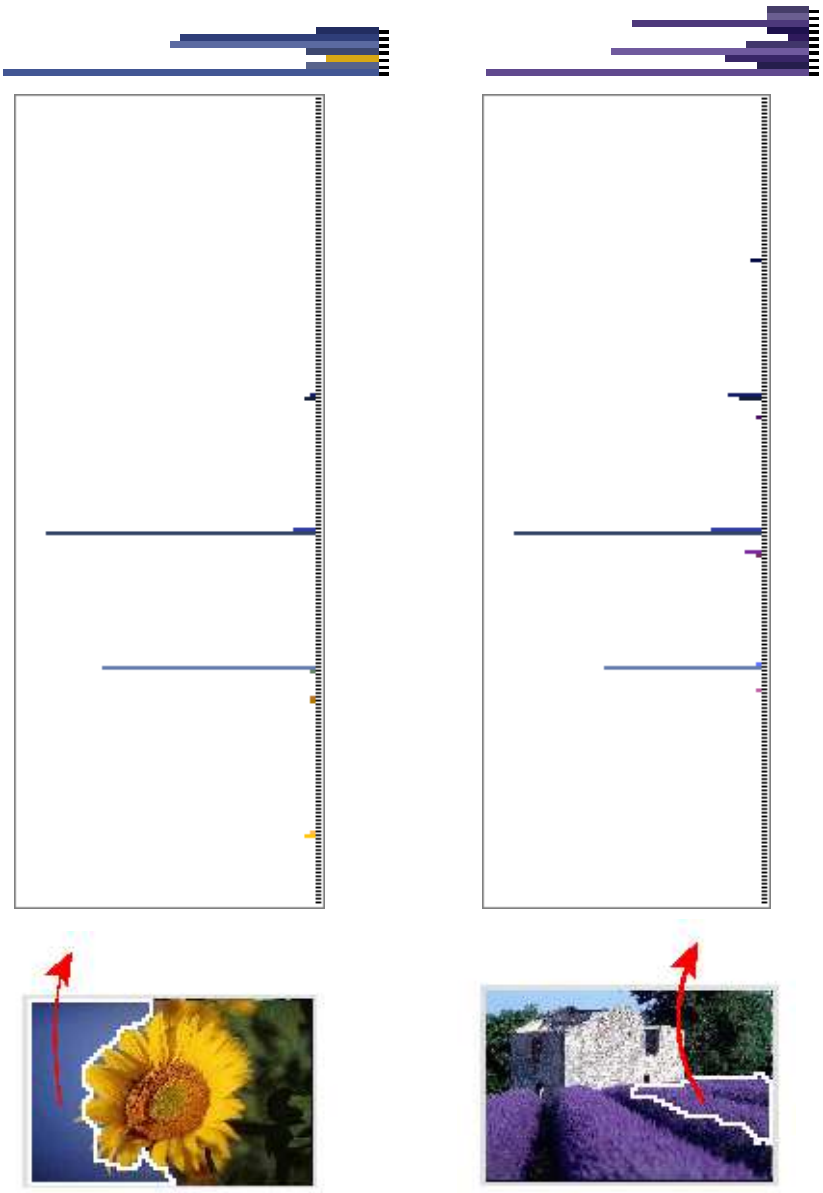


FIG. 4.2 – Limitation de l'histogramme couleur classique : le degradé de ciel bleu et la région mauve texturée de lavande sont deux régions perceptuellement différentes. Leurs histogrammes couleur classiques sont quasiment identiques (au milieu). Leur faible résolution couleur regroupent des couleurs perceptuellement différentes dans des mêmes cellules, tandis que les nuances de couleur du descripteur ADCS offrent une description plus fidèle (en haut) et permettent de distinguer plus efficacement ces régions.

où les couleurs notées $\{c_i^X\}$ et $\{c_j^Y\}$ correspondent à des triplets (l, u, v) de composantes dans l'espace Luv.

Les distances de type Minkowski ou d'intersection d'histogrammes, habituellement utilisées pour comparer des histogrammes calculés sur des partitionnements identiques de l'espace couleur (voir section 2.1), ne sont pas applicables aux distributions ADCS. Pour comparer deux régions, il est nécessaire de disposer d'une distance capable de mesurer la similarité entre deux index ADCS en intégrant à la fois les informations quantitatives (comparaison des ensembles de populations x et y), mais aussi qualitatives (comparaison des ensembles de couleurs c^X et c^Y). En section 2.1, nous avons résumé les principales distances existantes pour comparer des histogrammes couleur sur des quantifications identiques et différentes.

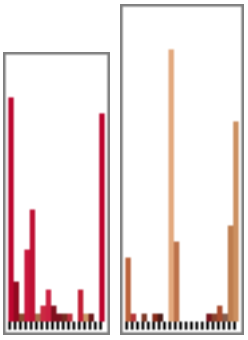


Fig. 4.3 – Exemple de deux index ADCS à comparer : ils sont exprimés sur deux ensembles différents de couleurs.

A notre connaissance, seules deux distances ont été proposées pour comparer des histogrammes basés sur des quantifications différentes : les distances *Earth Mover Distance* [95] et *Weighted Correlation* [55]. Dans les deux cas, elles ont été appliquées à la recherche globale d'images et avec une quantification propre à chaque image. La première approche se ramène à un problème d'optimisation linéaire qui se résoud itérativement ; elle est donc complexe et coûteuse en temps de calcul. La seconde, plus rapide (se calcule en $\mathcal{O}(NN')$ où N et N' sont les nombres de cellules couleur de chaque distribution), est définie spécifiquement pour leur algorithme de quantification basé sur les k-means. Leurs expressions sont données en section 2.1.

Notre choix se porte sur la *forme généralisée de la distance quadratique* qui a été introduite en section 2.1). Le principal avantage de la distance quadratique généralisée est son adéquation à tout type de quantification couleur et la prise en compte de la métrique de l'espace couleur à travers la matrice de similarité. Son expression pour comparer une distribution ADCS X avec une distribution ADCS Y de nuances de couleurs respectives $\{c_1^X, ..., c_{n_X}^X\}$ et $\{c_1^Y, ..., c_{n_Y}^Y\}$ est la suivante :

$$d_{quad}(X,Y)^2 = \sum_{i,j=1}^{n_X} x_i x_j a_{c_i^X c_j^X} + \sum_{i,j=1}^{n_Y} y_i y_j a_{c_i^Y c_j^Y} - 2\sum_{i=1}^{n_X}\sum_{j=1}^{n_Y} x_i y_j a_{c_i^X c_j^Y}$$

L'expression de la similarité $a_{c_i^X c_j^Y}$ entre deux nuances de couleur c_i^X et c_j^Y de l'espace LUV est donnée par l'expression (3.1).

Lors de la recherche de régions similaires à une région d'index ADCS X, nous utiliserons cette formule pour déterminer les régions de la base dont l'index Y minimise la distance quadratique. Dans cette formule, le troisième terme dépend de X et Y, tandis que les deux premiers ne dépendent que chaque index X,Y *individuellement*. Nous pouvons écrire $d_{quad}(X,Y)^2 = F(X,X) + F(Y,Y) - 2.F(X,Y)$, où $F(X,Y) = \sum_{i=1}^{n_X}\sum_{j=1}^{n_Y} x_i y_j a_{c_i^X c_j^Y}$. Afin de réduire le coût calculatoire au moment de la comparaison, nous précalculons $F(X,X)$ et $F(Y,Y)$: au moment de l'indexation d'une région par son index ADCS X, nous calculerons et lui associerons la quantité $F(X,X)$. Ainsi, lors de la phase de recherche de régions similaires, la comparaison se réduit principalement au calcul du terme croisé $F(X,Y)$. Le précalcul des termes $F(X,X)$ et $F(Y,Y)$ permet de réduire le temps de recherche par un facteur 3.

Dans Netra [24], la mesure de comparaison de leur descripteur de couleurs dominantes est une approximation la distance quadratique : le terme inter-cellules n'existe pas étant donnée leur hypothèse que toutes les couleurs (qu'ils nomment "couleurs dominantes") sont 2 à 2 à distance maximale dans l'espace couleur. Cette forme de distance quadratique approximée est reprise dans le standard MPEG7 [68] en association au descripteur DCD évoqué précédemment. Elle revient, dans les termes $F(X,X)$ et $F(Y,Y)$ de notre formule (2.2), à poser $a_{c_i c_j} = 1$ si $i = j$ et 0 sinon. En effet, ils imposent que les couleurs au sein d'un même index soient à distance maximales, autrement dit de similarité nulle. Nous ne faisons pas de telle hypothèse car nos nuances de couleurs peuvent être proches entre elles à des degrés variables. Il en résulte une description couleur des régions moins précise et donc moins fidèle que la nôtre.

Dans Blobworld [15], une approximation différente de la distance quadratique est utilisée pour comparer les régions. Ces dernières sont indexées par des distributions de couleurs exprimées sur une *quantification systématique* de l'espace LAB en 218 couleurs. Après projection des index dans l'espace des valeurs propres de la matrice A, la distance quadratique est approximée par la distance euclidienne sur les premières composantes du nouvel espace. Cette double approximation de la distance conduit à une perte de performance qui n'est pas étudiée. Notons que cette optimisation n'est pas applicable à notre cas sachant que deux ADCS sont exprimées sur des ensembles de couleurs propres à chaque région.

Nous rappelons qu'une synthèse des principales distances existantes pour comparer des distributions de couleur est présentée en section 2.1.

4.5 Intégration de la similarité géométrique

Lors d'une requête par région exemple, la description photométrique ADCS est privilégiée. Cependant, comme nous le verrons dans l'interface, l'utilisateur a la possibilité de la combiner avec des descripteurs élémentaires de nature géométrique, s'il juge ces primitives pertinentes pour le type de régions qu'il recherche.

Les attributs géométriques de régions utilisés sont : surface, position et compacité.

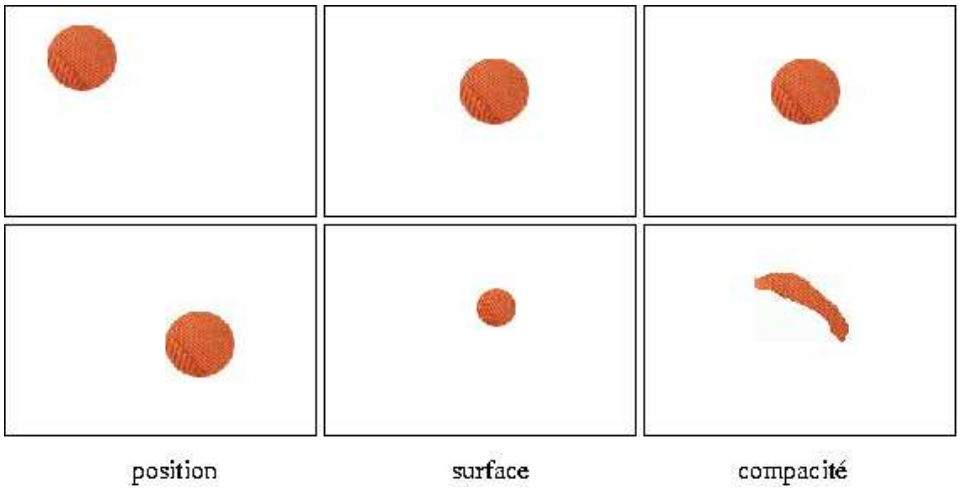


FIG. 4.4 – Illustration des 3 descripteurs géométriques utilisés. Bien que photométriquement similaires, les régions de la ligne du haut diffèrent de celles du bas selon, de gauche à droite, la position, la surface et la compacité.

Le descripteur de surface est définie comme le rapport entre le nombre de pixels contenus dans la région et le nombre total de pixels dans l'image. La surface d'une région représente son importance au sein de l'image qui la contient.

$$S = \frac{\mid Region \mid}{hauteurImage \ \times \ largeurImage}$$

où $\mid Region \mid = \ [\sum_{(x,y) \in \ Region} 1 \ ]$ désigne le nombre de pixels de la région.

Le descripteur de position est défini comme le centre de masse des pixels d'une région dont les coordonnées sont normalisées par rapport aux dimensions de l'image. Cette normalisation rend l'attribut de position indépendant de la taille

globale de l'image mais aussi de son rapport hauteur/largeur (format portrait, paysage, carré...) :

$$P = (P_X, P_Y) \;=\; \left(\frac{\sum_{(x,y)\in\ Region} x/\mid Region \mid}{largeurImage} \;,\; \frac{\sum_{(x,y)\in\ Region} y/\mid Region \mid}{hauteurImage} \right)$$

Le descripteur de compacité est une caractérisation très simple de forme de régions. Elle est calculée comme le rapport du carré de la somme des longueurs des contours de la région par la surface de la région :

$$C = \frac{\left[\sum longueur\ des\ contours\right]^2}{\mid Region \mid}$$

Notons qu'une region peut comporter plusieurs contours si elle est "trouée" : un contour extérieur et autant de contours intérieurs qu'elle comporte de trous. La compacité est maximale pour un disque et très faible pour une région allongée, fine ou de forme irrégulière.

Ces informations géométriques sont caculées à l'issue de la segmentation et stockées dans la structure du graphe d'adjacence de régions (voir annexe A).

Concernant la description de forme, nous ne jugeons pas nécessaire de calculer de descripteur plus précis (voir par exemple ceux proposées dans MPEG7 [68]) car la forme des régions obtenues par segmentation sur une base d'image généraliste n'est pas jugée suffisamment pertinente[3]. Le critère de compacité a été jugé suffisant à l'usage sur les bases généralistes.

Les descripteurs S de surface et C de compacité sont des scalaires et la position $P = (P_X, P_Y)$ est une paire de scalaires. Pour la recherche, la distance L^1 sera utilisée pour S et C et L^2 pour P.

Codage de l'index : la syntaxe de l'index général (descripteurs ADCS et géométriques) pour une région donnée est la suivante :

$$\text{index} \;: n, (l_1, u_1, v_1, p_1), ..., (l_n, u_n, v_n, p_n), S, C, P_X, P_Y$$

Chaque composante est codée sur un octet. La taille de l'index pour une région est donc de $4n + 5$ octets.

En résumé, pour la recherche par régions, on utilise les combinaisons suivantes de descripteurs avec leurs distances respectives :

[3]Par contre, pour la recherche dans des bases de logos, de timbres, de graphiques, par exemple, des descripteurs de contours plus sophistiqués s'avèrent très discriminants.

– ADCS, distance quadratique (forme généralisée)
– Position, distance L^2
– Surface, distance L^1
– Compacité, distance L^1

Nous allons définir la distance globale $D(R_r, R_c)$ entre une région requête R_r et une région candidate R_c qui combine les distances entre ces quatre descripteurs. Dans un cadre de recherche dans des bases hétérogènes d'images de type photothèque, l'importance relative de ces descriptions est variable selon la nature de l'objet ou de la zone recherchée. Une pondération par défaut sera proposée à l'utilisateur qu'il pourra modifier dans l'interface graphique (voir figure 4.6). On note A_r, P_r, S_r, C_r et A_c, P_c, S_c, C_c les descripteurs ADCS, position, surface et compacité associés respectivement à la région requête R_r et à la candidate R_c. La distance globale $D(R_r, R_c)$ est exprimée comme combinaison linéaire des quatre distances :

$$\begin{aligned} D(R_r, R_c) = \; & \alpha_A.d_{quad}(A_r, A_c) + \\ & \alpha_P.\, ||\, P_r - P_c\, ||_{L^2} \; + \\ & \alpha_S.\, |\, S_r - S_c\, | \; + \\ & \alpha_C.\, |\, C_r - C_c\, | \end{aligned} \tag{4.1}$$

Les poids α_A, α_P, α_S et α_C sont initialisés avec des valeurs par défaut (respectivement 1.0, 0.2, 0.2 et 0.2) et sont ajustables dans l'interface. On impose que $\alpha_A \in]0;1]$ et $\alpha_P, \alpha_S, \alpha_C, \in [0;1]$. Dans ces conditions, la démonstration que D est une distance est directe avec la vérification des quatre conditions : identité, symétrie, inégalité triangulaire et non-négativité. La dynamique des valeurs des quatres sous-distances ainsi que celles des 4 pondérations $\alpha_A, \alpha_P, \alpha_S, \alpha_C$ sont normalisées entre elles. Un coefficient α élevé accroît l'importance accordée au descripteur correspondant. Un coefficient mis à zéro annule le calcul de la distance du descripteur.

Dans la formule (4.1), les distances géométriques étant beaucoup simples à évaluer que la distance quadratique, une heuristique d'accélération du calcul de $D(R_r, R_c)$ est proposée pour la recherche de régions similaires.
L'ordre croissant de complexité de calcul des quatre termes de $D(R_r, R_c)$ est le suivant : $|\, C_r - C_c\, |$, $|\, S_r - S_c\, |$, $||\, P_r - P_c\, ||_{L^2}$, $d_{quad}(A_r, A_c)$. Au moment de la recherche, ils sont évalués dans leur ordre de complexité. Dès que l'une des distances sera très grande, la région R_c sera immédiatement considérée comme non-similaire en lui associant une distance $D(R_r, R_c)$ pseudo-infinie. Ainsi, le calcul complet de la distance totale (formule (4.1)) ne sera effectué que pour les régions qui sont des candidates potentielles.

Si les poids α_P,α_S, α_C sont fixés à zéro par l'utilisateur, cette stratégie d'optimisation n'est pas mise en oeuvre car $D(R_r, R_c)$ est réduite à $d_{quad}(R_r, R_c)$.

En notant $D = D(R_r, R_c)$, la stratégie de rejet se déroule ainsi :

initialisation : $D = 0$
si ($\alpha_C > 0$ et $\alpha_C . \mid C_r - C_c \mid >$ `SEUIL_COMPACITE`)
  alors $D =$ `VALEUR_PSEUDO_INFINIE`
sinon
  $D = D + \alpha_C . \mid C_r - C_c \mid$
  si ($\alpha_S > 0$ et $\alpha_S . \mid S_r - S_c \mid >$ `SEUIL_SURFACE`)
    alors $D =$ `VALEUR_PSEUDO_INFINIE`
  sinon
    $D = D + \alpha_S . \mid S_r - S_c \mid$
    si ($\alpha_P > 0$ et $\alpha_P . \mid\mid P_r - P_c \mid\mid_{L^2} >$ `SEUIL_POSITION`)
      alors $D =$ `VALEUR_PSEUDO_INFINIE`
    sinon
      $D = D + \alpha_P . \mid\mid P_r - P_c \mid\mid_{L^2} + \alpha_A . d^A_{quad}(R_r, R_c)$

La quantité `VALEUR_PSEUDO_INFINIE` est une valeur de distance arbitrairement grande qui place les régions non désirables en fin de liste. Les seuils `SEUIL_POSITION`, `SEUIL_SURFACE` et `SEUIL_COMPACITE` sont fixés comme la moitié de la distance maximale propre à chaque descripteur.

Nous verrons dans la section 4.7 que cette stratégie de rejet se traduit par une accélération du processus de recherche.

4.6 Interface utilisateur

Le système de recherche de régions est intégré à notre plateforme Ikona, construit sur une architecture client-serveur (serveur écrit en C++ et interface utilisateur client en Java). Les différentes fonctionnalités de recherche et de navigation avec Ikona sont décrites dans [7].

Pour le scénario de recherche de régions (“paradigme 1”), l'utilisateur commence à naviguer aléatoirement dans la base grâce à l'interface client (voir capture d'écran 4.6). Chaque région dans chaque vignette image peut être cliquée pour désigner la région requête. Dans une deuxième fenêtre, l'utilisateur peut ajuster l'importance relative des différents descripteurs (ADCS, surface, position et compacité) selon leur pertinence pour le type de régions (ou objets) recherchés.

Le serveur retourne les images comportant une région similaire à la région requête affichées en ordre croissant de distance finale (formule (4.1)) selon la stratégie de rejet. Dans les images retournées, les contours des régions retournées sont tracés en blanc.

FIG. 4.5 – L'interface utilisateur Ikona pour la recherche de régions (paradigme 1). Chaque région dans chaque vignette est cliquable dans la fenêtre principale. Les régions retrouvées dans les images sont identifiables par leurs contours blancs. La seconde fenêtre de paramétrage permet l'ajustement dynamique de l'importance relative de la diversité couleur, position, surface et compacité, c'est-à-dire les poids les poids α_A, α_P, α_S, α_C.

4.7 Résultats

Protocole de test

Notre système a été testé avec un PC standard à 2.0 GHz (512Mo de mémoire vive). La base de test comporte au total 11.479 images (majoritairement en couleur) réparties de la façon suivante :

- 792 images de textures de la base Vistex [4]
- 552 de la banque d'images *Images Du Sud* [5]
- 10.135 la base *Corel* [6]

Les deux dernières bases sont des images génériques de fleurs, dessins, portraits, paysages, architecture, fractales, gens, fruits, jardins, voitures, cuisines, ...

L'évaluation des performances d'un système de recherche d'images par le contenu est une tâche ardue car elle dépend étroitement de la perception visuelle de l'humain, du domaine d'application (générique ici) et du contenu de la base elle-même. Pour un système de recherche de régions, la difficulté est accrue car l'évaluation requiert l'identification préalable de régions spécifiques pour la création d'une base vérité-terrain de régions.

Nous avons construit notre base vérité-terrain de régions en partie avec la base Vistex qui est construite à partir de parties d'images de scènes réelles. L'autre partie de notre base vérité-terrain de régions est constituée de régions détectées par notre système. Elles ont ensuite été étiquetées manuellement pour définir leur appartenance à l'une des trois classes suivantes : personne (régions de peau), lavande, piscine. Au sein de chaque classe, les régions se réfèrent au même *objet*, mais sont aussi perceptuellement proches.

Description de régions

Dans la figure 4.6, la troisième image de chaque exemple est l'image créée à partir des nuances de couleur utilisées pour indexer chaque région. La forte similarité visuelle entre ces images avec l'image originale correspondante montre la précision de la description de variabilité couleur ADCS. Cette fidélité d'apparence visuelle est due à la quantification fine et adaptative des régions.

Un total de 963.215 nuances de couleur dans l'espace couleur Luv a été automatiquement déterminé pour indexer les 56.374 régions ce qui donne une moyenne de 17 nuances de couleurs par région. Parmi ces couleurs, 690.419 sont uniques, nombre à comparer à la palette de 200 couleurs fixées dans les descriptions couleur

[4] http ://www-white.media.mit.edu/vismod/imagery/VisionTexture/vistex.html
[5] http ://www.imagedusud.fr
[6] http ://www.corel.com

usuelles. Il est important de noter que cette haute résolution couleur est compatible avec une représentation compacte de l'index. La taille de stockage d'un index est 69 octets en moyenne (un scalaire est stocké comme un octet). Celui-ci s'avère trois fois plus compact qu'un histogramme couleur classique à 200 cellules.

L'extraction des index ADCS pour les régions d'une image prend environ 0,8s.

nombre d'images	11.479
nombre de régions	56.374
nombre total de nuances de couleurs	963.215
nombre total de nuances uniques	690.419
nombre de nuances de couleurs par région	17
temps d'indexation par image	0,8s

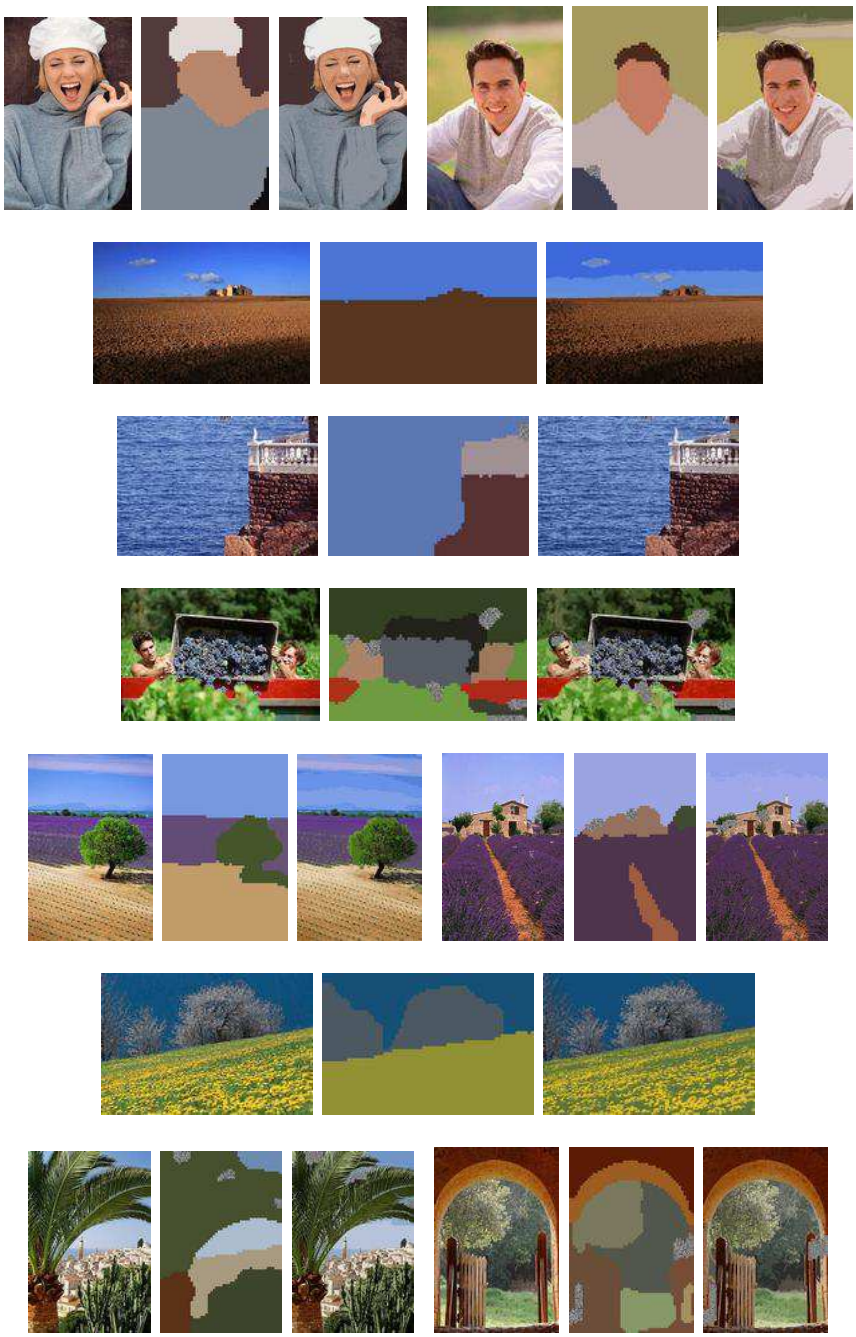
Recherche de régions similaires

Evaluation qualitative

A l'usage, les requêtes à l'aide de notre système ont toujours retourné des régions présentant des distributions de couleurs perceptuellement proches pour différents types de région-requête : uniformes ou texturées ou avec différentes nuances de la même teinte. Les régions retournées donnent une impression de continuum visuel au fil des rangs. La prise en compte de la position, de la surface des régions et de la compacité (c'est-à-dire en attribuant des valeurs non nulles aux poids dans la fenêtre de paramètres d'Ikona) améliore presque très souvent la pertinence des résultats tout en accélérant les temps de requête grâce à la stratégie de rejet de régions décrite précédemment.

Lors de la comparaison des résultats obtenus avec ADCS par rapport à l'histogramme couleur classique, l'amélioration de la similarité perceptuelle est remarquable parmi les régions retournées. Le gain en fidélité de description des régions participe à la réduction du "fossé sémantique" ("semantic gap", en anglais), dans la mesure où les résultats sont plus satisfaisants du point de vue de l'utilisateur.

Selon l'apparence visuelle de la région requête, les résultats peuvent être plus ou moins pertinents. Par exemple, une requête sur une petite région noire n'a pas de sens dans l'absolu puisqu'il peut s'agir de parties d'objets très différents mais aussi d'ombre. Inversement, des "concepts" peuvent comporter des apparences très différentes ; par exemple, les concepts "chien", "vêtement", "voiture" peuvent présenter des teintes et des textures très variées. Dans le tableau 4.7, nous avons relevé des exemples d'objets pour lesquels les requêtes régions ont montré une forte corrélation avec leur apparence visuelle.



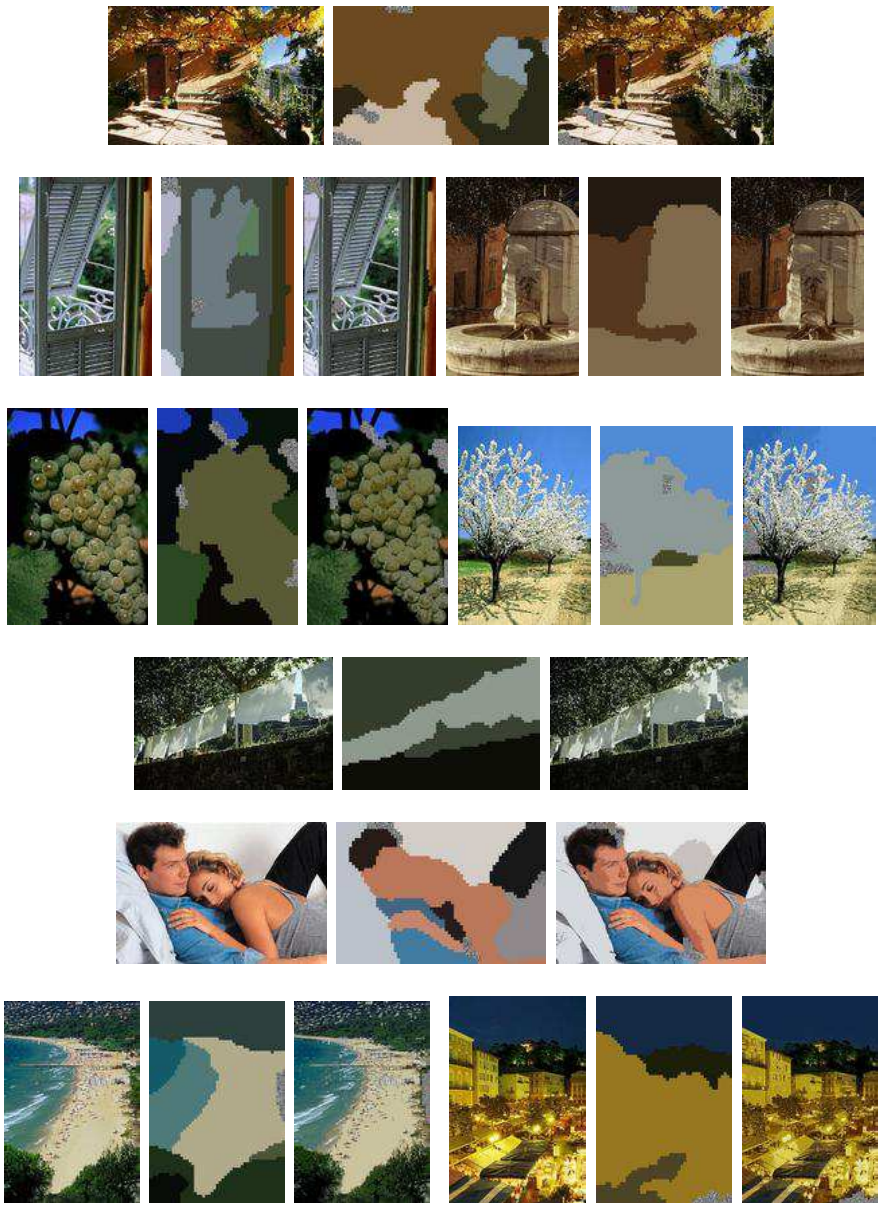


FIG. 4.6 – Illustration de la description fine. Chaque triplet d'images comporte l'image originale, l'image segmentée et l'image des régions avec leur nuances couleur utilisées pour leur indexation. Les petites régions supprimées sont grisées. La forte similarité visuelle entre chaque image originale et l'image des nuances de couleur illustre la précision du descripteur. Plus d'exemples sont montrés en annexe B.

taille	**position dans l'image**	**teinte**	**variabilité couleur**	**"objet" probable**
grand	en bas	blanc	faible	neige
grand	en bas	violet	forte	champ de lavande
grand	en haut	bleu	dégradé	ciel
non-discriminant	*non-discriminant*	cyan	faible	piscine
petit à moyen	centre	rose clair	dégradé	peau

TAB. 4.1 – Exemples de régions présentant une corrélation remarquable entre leur description visuelle et sémantique.

Il est fréquent que l'apparence visuelle d'une région, aussi cohérente visuellement soit-elle, ne soit pas spécifique à une unique classe d'objets. La "valeur sémantique" des régions détectées dépend donc souvent du scénario de recherche et de l'intérêt de l'utilisateur.

Ces considérations nous ont aidés à établir la base vérité terrain de régions, afin de produire une évaluation quantitative de la précision de notre système.

Amélioration du pouvoir discriminant

Les figures 4.7 et 4.8 illustrent l'amélioration apportée par la distance quadratique par rapport à la distance L^1. Nous avons choisi 3 imagettes "ciel", "brique" et "osier". Chacune a été transformée selon 6 facteurs d'intensité. Nous obtenons un ensemble de trois familles d'imagettes d'homogénéités et d'intensités différentes, soit 18 images (figure 4.7). Les distributions de couleur des 3 images les plus claires (correspondant au facteur 2 d'intensité) et celles des 3 les plus sombres (facteur 0.5) ont été comparées à celles des 17 autres imagettes avec la distance L^1 couplée à un histogramme classique LUV en 216 cellules (fig. 4.7) et avec la distance quadratique couplée au descripteur ADCS (fig. 4.8).

On observe la capacité de la distance quadratique à mesurer efficacement le continuum perceptuel entre les différentes transformations d'une même imagette tout en discernant les familles d'imagettes entre elles (ciel, brique, osier), bien que les distributions soient piquées donc d'intersections parfois nulles. Il est important de noter que ces propriétés de continuité de la distance quadratique par rapport aux changements d'intensité est aussi vérifiable pour les changements de couleurs et de manière générale à tout changement qui entraîne une translation de la distribution sur des cellules voisines. Alors que ces changements pourront se traduire

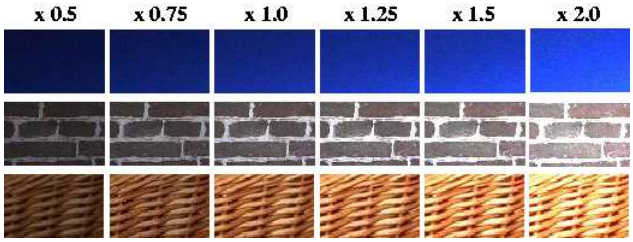


FIG. 4.7 – 3 imagettes extraites de régions d'images ont été transformées selon 6 facteurs d'intensité. Elles diffèrent en couleurs et textures.

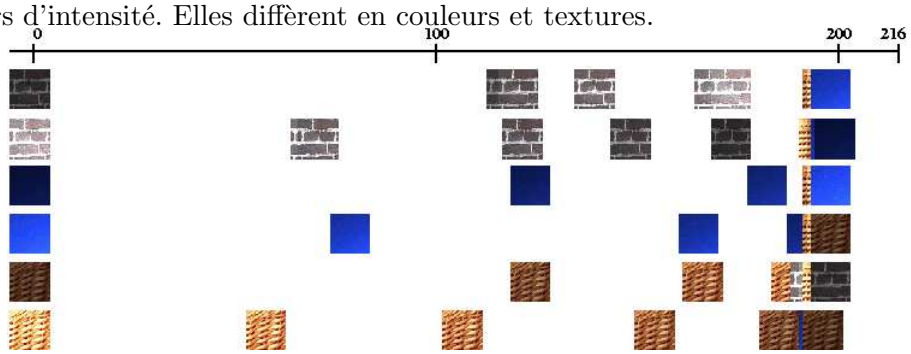


FIG. 4.8 – 6 comparaisons avec L^1. Sur chaque ligne, les imagettes sont positionnées selon la valeur de leur distance L^1 en distribution à la première imagette. La graduation indique les valeurs de ces distances à l'exemple. Les imagettes très similaires sont correctement jugées les plus proches avec L^1, mais les autres sont agglomérées à proximité de la distance maximum (autour de 200) et ne sont plus discernables les unes des autres : en moyenne, 13 imagettes sur 18 se trouvent à une distance dans l'ensemble $\{197, 198, 199, 200, 201\}$. Ceci illustre l'influence de l'homogénéité des données sur l'imprécision des distances de types cellule-à-cellule.

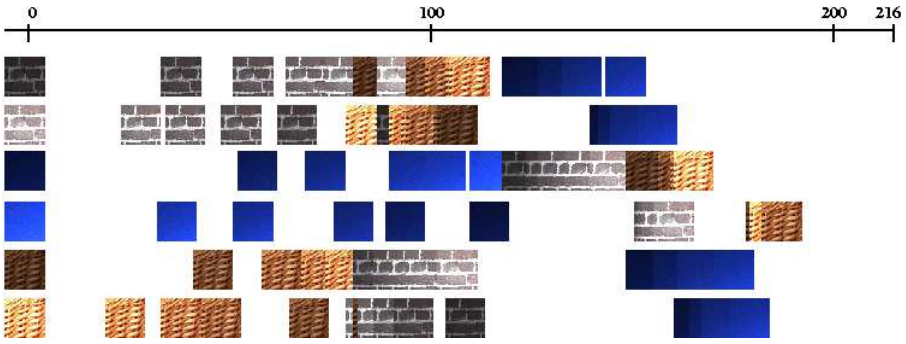


FIG. 4.9 – 6 comparaisons avec la distance quadratique. Les valeurs de distance sont plus étalées qu'avec L^1 et fournissent des résultats nettement plus satisfaisants. Notre perception du *continuum* visuel entre les images est mesurée plus fidèlement et avec plus de discrimination avec la distance quadratique.

par des valeurs maximales avec la distance L^1 (phénomène de saturation), ils correspondront à des variations progressives de la distance quadratique. A l'instar de L^1, les mêmes problèmes devraient être observés avec tout autre type de "distance daltonienne" (voir section 2.1). Il en résulte une meilleure fidélité dans la mesure de similarité visuelle. De plus, la distance quadratique ne présente pas ce phénomène de saturation des valeurs de distance et offre un meilleur pouvoir de discrimination visuelle entre les descripteurs de régions.

Cette propriété de la distance quadratique, combinée au descripteur ADCS, est importante pour permettre une comparaison plus discriminante visuellement entre les différentes régions de la base.

Evaluation Numérique

Les descripteurs régions proposés dans la littérature reposent tous sur une palette de couleurs commune à toute la base, de type histogramme couleur classique. En guise de comparaison à ADCS, nous avons aussi indexé notre base avec un descripteur classique de ce type : quantification de l'espace LUV en 6 cellules par composante et comparaison d'histogramme avec la distance L^1.

L'évaluation numérique de la précision de recherche de notre système a été testée en considérant chaque région de la base vérité terrain comme région requête parmi les classes suivantes :

- 88 classes de Vistex (792 images et 792 régions vérité terrain)
- classe "lavande"(108 images et 134 régions vérité terrain)
- classe "personnes" (371 images et 634 régions vérité terrain)
- classe "piscine" (26 images et 29 régions vérité terrain)

La base vérité-terrain comporte donc au total 1297 images et 1589 régions. Ces annotations de régions ont été effectuées manuellement. Chacune de ces régions a été utilisée comme région requête. Parmi les premières régions retournées (analysées jusqu'au rang 50) la précision au rang k est mesurée comme le quotient par k du nombre de bonnes régions retournées jusqu'au rang k. Les figures 4.10 et 4.11 montrent les courbes de précision obtenues par requêtes automatiques de chacune de ces régions avec les trois configurations de description suivantes : ADCS, histogramme classique et combinaison de ADCS avec les descripteurs géométriques proposés dans la section 4.5. Par souci de clarté, la précision sur la base Vistex est présentée sur la moyenne des 88 classes.

Pour toutes les classes, ADCS améliore la précision avec un gain positif par rapport à l'histogramme classique. Ce gain est variable selon les classes considérées. Les régions décrites par plusieurs nuances de couleurs ont été mises en correspondance avec des régions comportant plusieurs nuances de couleurs et idem pour les régions comportant une (ou très peu) de couleurs. Nous avons ob-

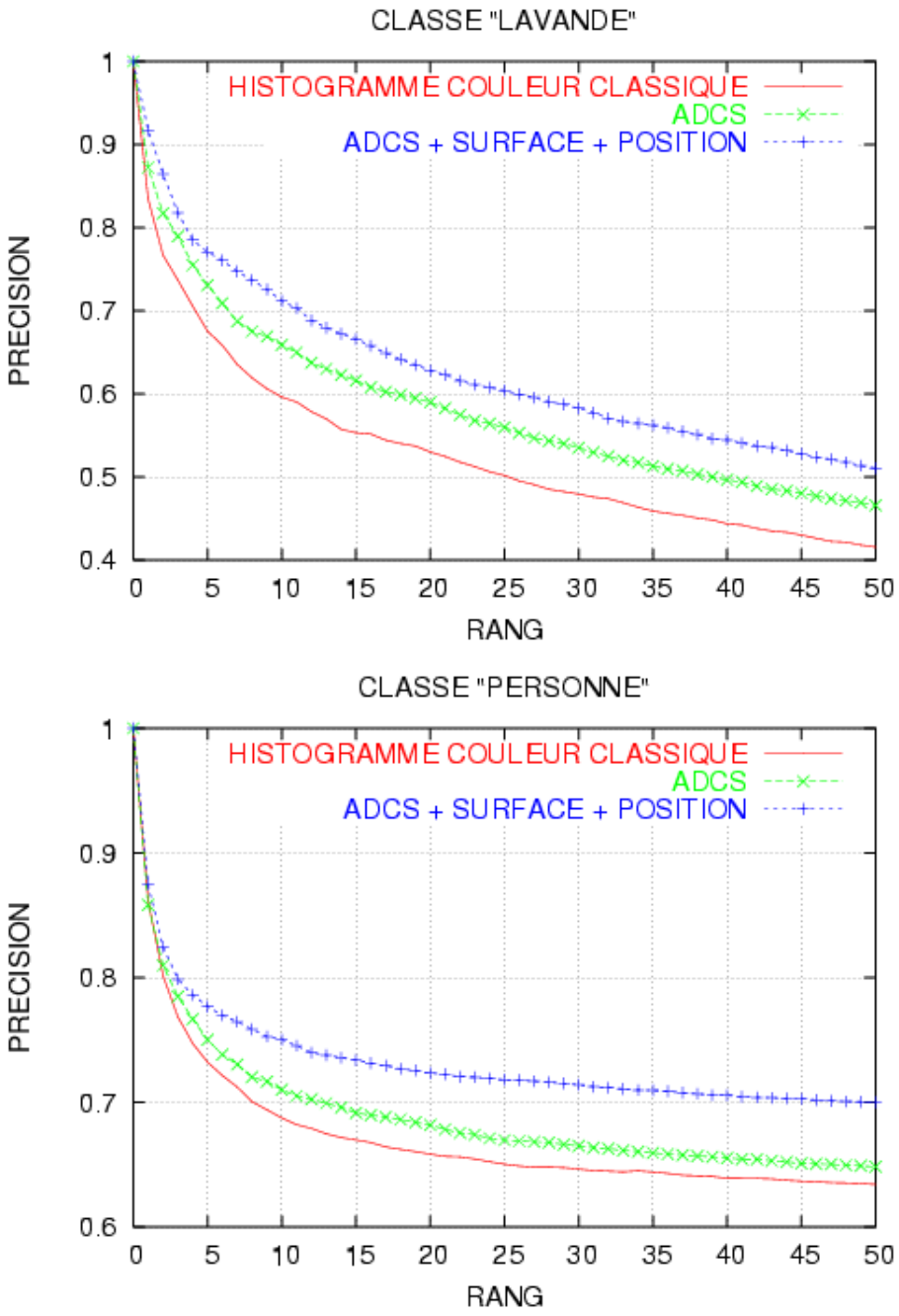


FIG. 4.10 – Courbes de précision sur les classes “lavande” et “personne” en utilisant les 3 modes de recherche : histogramme classique contre ADCS et contre ADCS combiné avec la surface et la position.

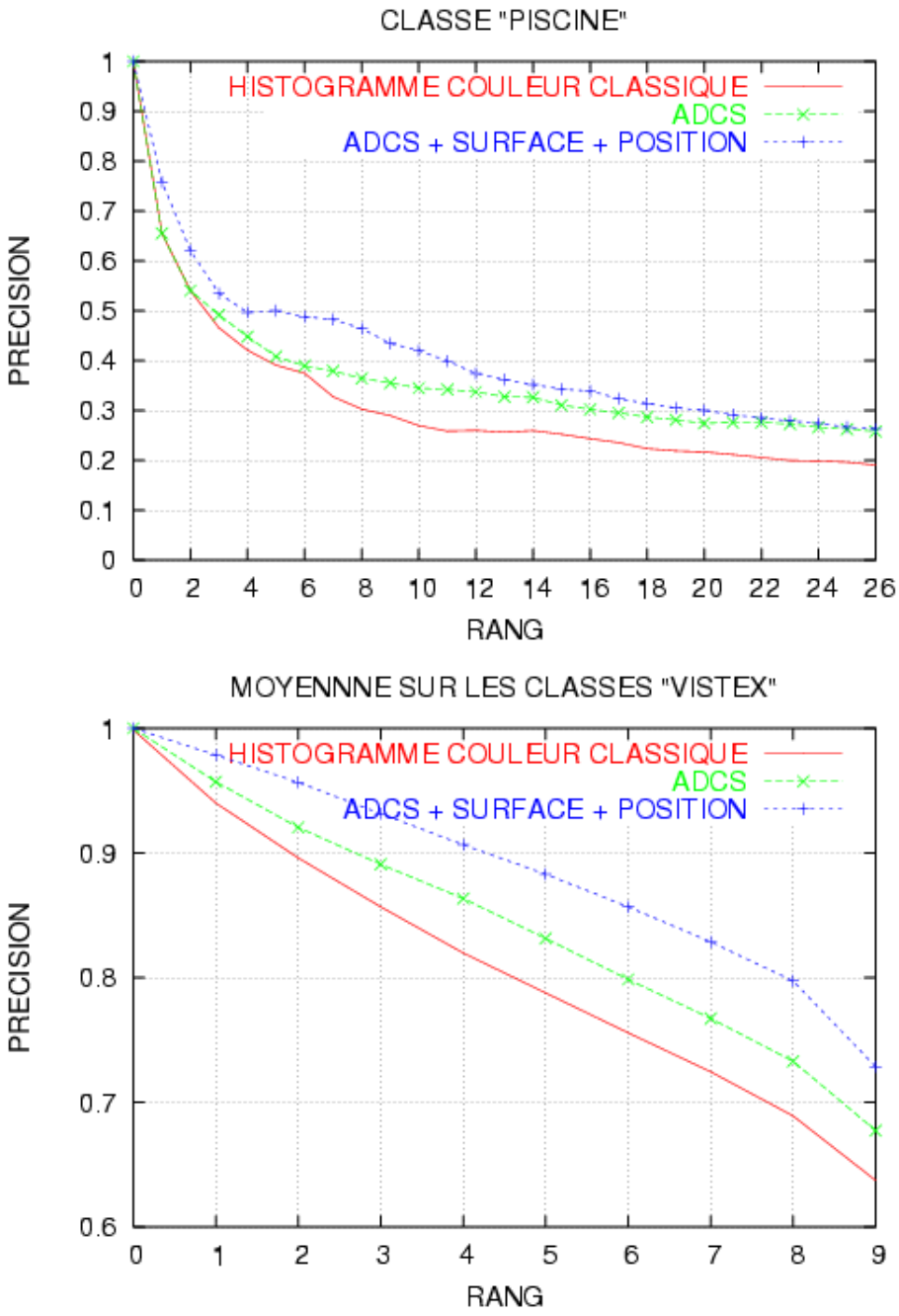


FIG. 4.11 – Courbes de précision sur les classes “piscine” et “Vistex” en utilisant les 3 modes de recherche : histogramme classique contre ADCS et contre ADCS combiné avec la surface et la position.

FIG. 4.12 – Résultats de recherche à partir de la région de lavande en haut à gauche avec l'histogramme couleur classique à 216 cellules. L'histogramme classique ne peut pas distinguer des nuances de mauve de nuances de bleu. Aucun descripteur géométrique n'est utilisé ici.

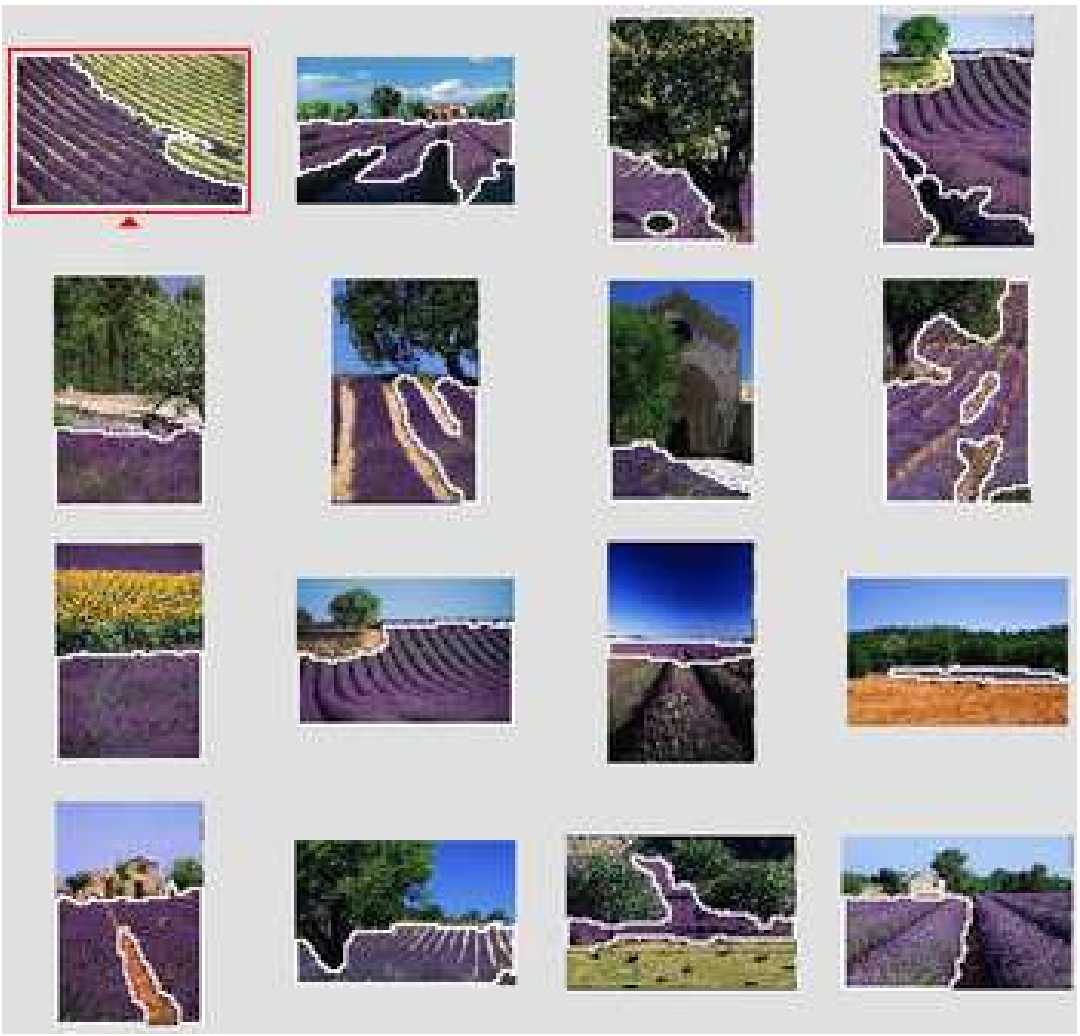


FIG. 4.13 – Résultats de recherche à partir de la région de lavande en haut à gauche avec ADCS. La fine résolution couleur du descripteur photométrique améliore la précision des régions retrouvées.

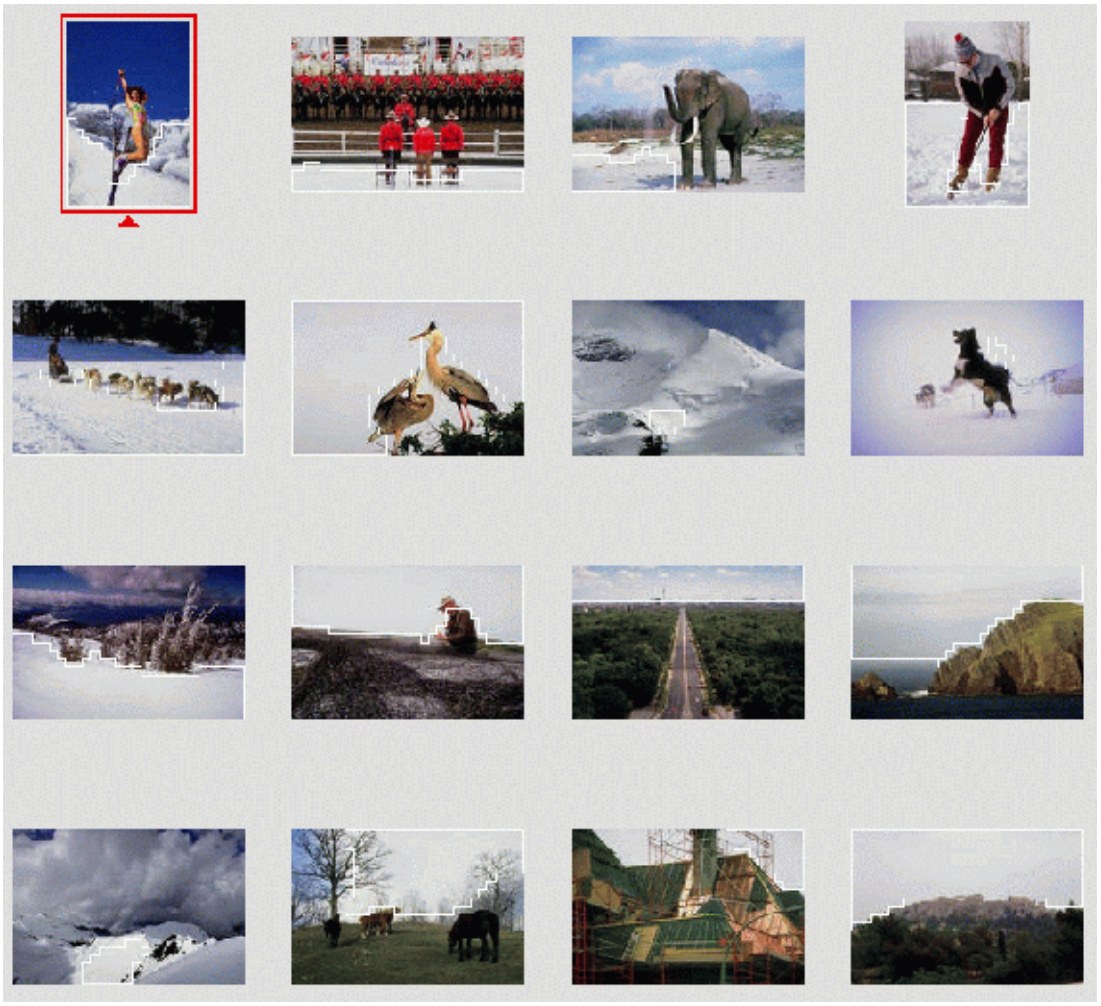


FIG. 4.14 – Résultats de recherche à partir de la région de neige en haut à gauche avec ADCS seul.

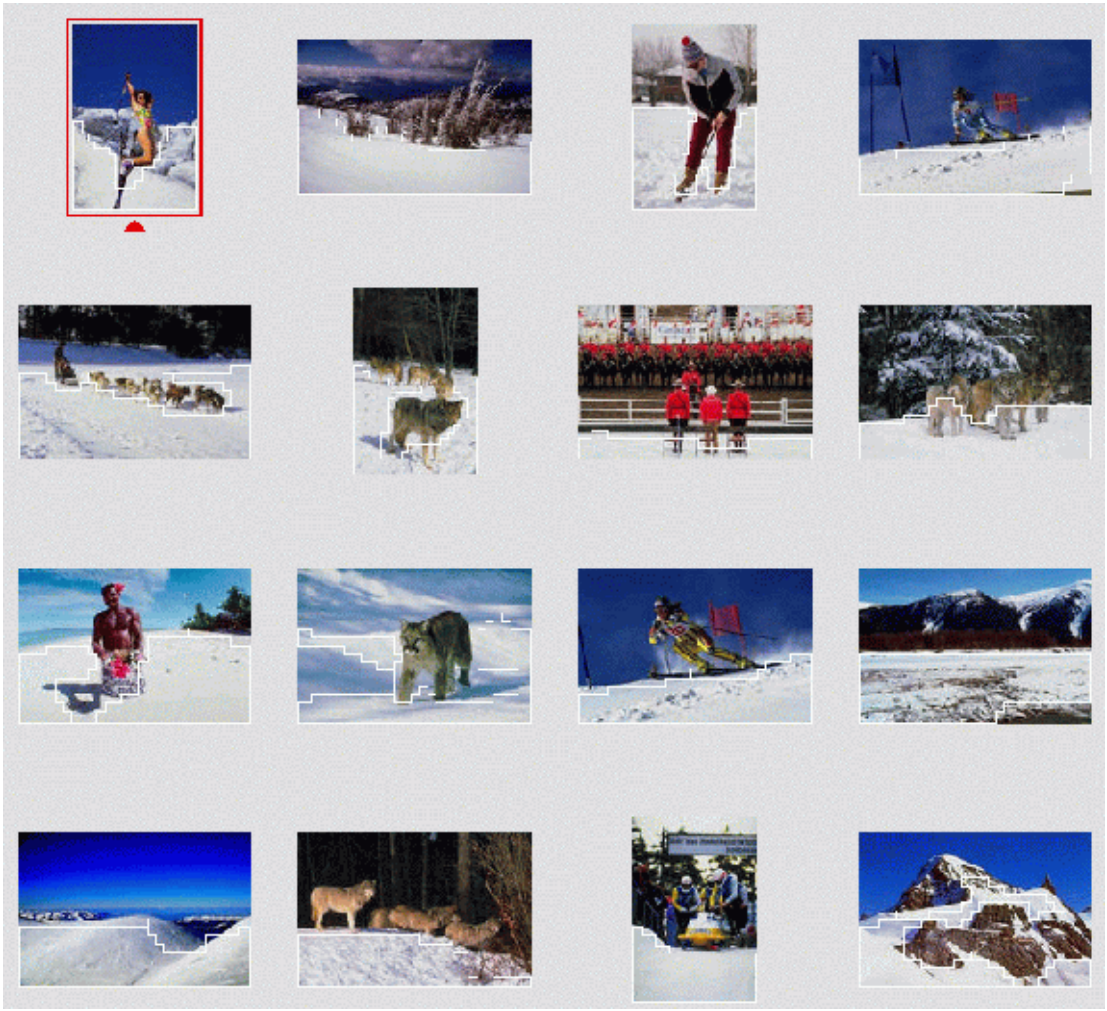


FIG. 4.15 – Résultats de recherche à partir de la région de neige en haut à gauche avec la combinaison de ADCS avec la surface et la position. Bien que les régions retournées avec le descripteur ADCS soient pertinentes en termes de photométrie, la taille et la position améliorent nettement la qualité des résultats. Les régions de neige sont un exemple de région requête pour laquelle les descripteurs géométriques sont fortement discriminants.

servé que le nombre de nuances de couleur est aussi une information exploitée dans les descripteurs ADCS. Les faux positifs d'images retournées avec l'histogramme classique étaient dus à une imprécision dans la similarité lorsque celui-ci ne permettait pas de distinguer deux régions perceptuellement différentes correspondant à des *objets* différents (voir capture d'écran 4.10). La finesse de la description ADCS couplée à la forme généralisée de la distance quadratique se traduit par un gain en fidélité de description et en précision de recherche par rapport à la représentation imprécise des couleurs dans l'histogramme classique.

La combinaison de ADCS avec des descripteurs géométriques simples a conduit à une seconde amélioration significative, à l'exception de la classe piscine pour laquelle le gain est quasiment nul dans les derniers rangs. Dans notre base vérité terrain, les régions associées à une piscine étaient soient de petites parties du fond de l'image, soit le fond entier, soit une tâche en haut ou en bas de l'image. Cette remarque est cohérente avec le fait que les descripteurs géométriques ne sont pas discriminants pour cette classe dans notre base. Cependant, pour des régions telles que *lavande*, *ciel*, *peau* ou *neige*, ils sont très discriminants. Les figures 4.14 et 4.15 illustrent deux résultats de requêtes sur des zones de neige avec ADCS seul (fig. 4.14) et avec ADCS couplé aux descripteurs géométriques (fig. 4.15).

Les index de régions sont comparés de manière exhaustive avec la région requête. Le temps moyen de recherche parmi les 56.374 régions est de 0.8s avec leur descripteur ADCS seul et de 0.5s avec ADCS et la surface et la position, avec $\alpha_A = 1.0$ et $\alpha_S = \alpha_P = 0.2$. Nous rappelons que la stratégie de rejet accélère les comparaisons dans le cas de la combinaison de plusieurs descripteurs.

4.8 Discussion : points ou régions comme description locale ?

Dans l'introduction de ce mémoire (section 1.2), nous avons présenté les diverses solutions existantes pour définir et décrire des composantes d'images. Alors que l'approche que nous venons de proposer s'attache à détecter des régions saillantes et de tailles significatives, les points d'intérêt offrent une caractérisation fine et précise de sites d'images. La complémentarité de la description par points d'intérêt [36] avec la description par régions a été mise en avant dans [8]. Dans cette partie, nous allons mettre en parallèle le principe de ces deux approches afin d'identifier leurs scénarios d'usage respectifs.

Pour la recherche par points d'intérêt, le principe est le suivant : dans chaque image de la base, les points d'intérêt identifient les voisinages de pixels présentant une très forte variation photométrique. Ils sont individuellement décrits par des

grandeurs photométriques et/ou géométriques lors de la phase d'indexation. La notion de "zone d'image" ne prend son sens qu'au moment de la requête : à la souris, l'utilisateur trace dans une image un rectangle qui définit la zone d'intérêt exemple. Les points contenus dans le rectangle constituent la clé de la requête et sont comparés aux points détectés dans les images de la base par un mécanisme de vote. Les images obtenant le meilleur score de vote sont celles contenant le plus grand sous-ensemble de points les plus similaires aux points requête. Cette approche est détaillée dans [36]. Dans cette approche, les "parties" d'images pertinentes pour une recherche sont définies *interactivement*. L'absence de définition a priori des parties d'images, contrairement à la segmentation en région, offre plus de souplesse à l'utilisateur.

Dans notre approche, rappelons que les régions caractérisent les zones importantes et photométriquement cohérentes de l'image. Elles sont détectées au moment de l'indexation. La description des régions (par ADCS) est de nature statistique dans la mesure où tous les pixels de régions sont pris en compte. Par contre, les points d'intérêt n'existent dans une image que sur les voisinages à forte variation photométrique et sont absents dans les zones uniformes et plus généralement lisses. Par exemple, dans la troisième image de la figure 4.16, le fond flou et la zone rouge lisse ne comportent pas de points. La figure 4.16 illustre les points et les régions détectés sur les mêmes images.

D'un point de vue pratique, la recherche par points présente l'avantage d'une définition interactive de zones d'intérêt, mais elle se traduit par un coût calculatoire nettement supérieur par rapport aux régions. Du point de vue de l'usage, si l'utilisateur recherche de grandes zones homogènes, l'approche région est préférable car les points détectés seront inexistants ou correspondront à du bruit. A l'inverse, pour rechercher des petites zones avec des détails caractéristiques, l'approche région est inadéquate car la segmentation peut ne pas avoir détecté ces zones. Dans le but d'une recherche précise sur des détails visuels, les points sont préférables, mais au prix d'un temps de réponse beaucoup plus long. Les points peuvent aussi être une alternative à la recherche par régions pour les cas d'images difficiles à segmenter, telle que l'image du milieu dans la figure 4.16. La figure 4.17 illustre le résultat d'une requête par points d'intérêt pour retrouver les images comportant la même partie de cellier à vin.

En résumé, il n'y a pas de meilleure solution dans l'absolu entre la description par points d'intérêt ou régions d'intérêt. Le choix doit être guidé par les critères de temps de réponse, de précision de recherche et par la nature des parties d'image recherchées. En l'absence de critères de recherche bien définis, l'approche par régions peut être préférée pour la rapidité de recherche. La complémentarité de ces deux approches permet de couvrir de très larges scénarios de requêtes partielles.

FIG. 4.16 – De gauche à droite : Images originales, régions détectées, points détectés. Alors que les points d'intérêt détectent les sites d'images correspondant à de hautes fréquences spatiales, les régions détectent des zones larges et homogènes selon la primitive locale LDQC. La scène du milieu illustre un cas difficile de segmentation et la région contenant les bouteilles n'est pas pertinente. Les points la caractérisent mieux. Les paramètres de détection de régions et de points sont ceux employés de façon standard sur la base entière.

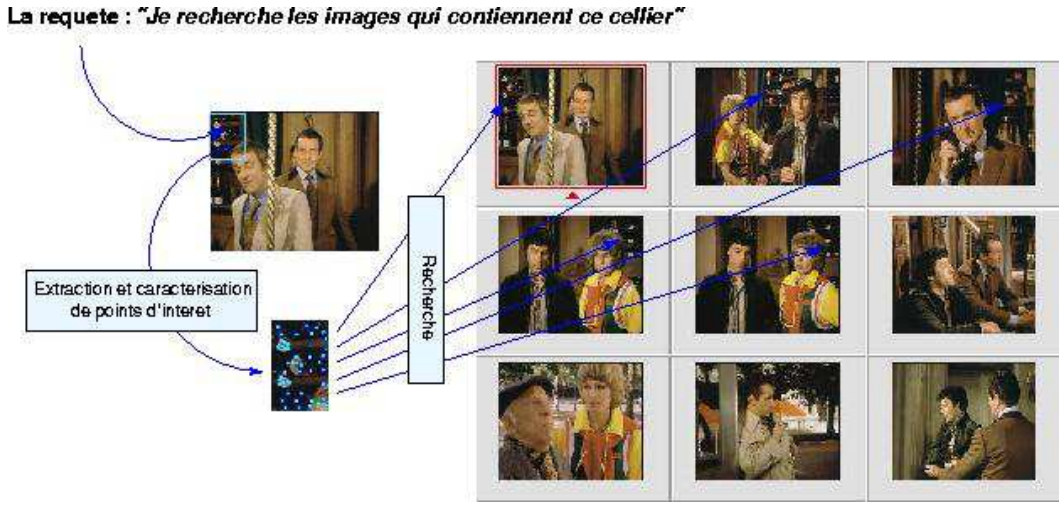


FIG. 4.17 – Résultat d'une requête par points d'intérêt pour retrouver les images comportant la même partie de cellier à vin. La requête sur cette zone n'aurait pas donné de résultat satisfaisant par une approche région qui la considèrerait comme majoritairement noire et lisse donc pas suffisamment caractéristique (image fournie par Valérie Gouet).

4.9 Perspectives

Combinaison avec d'autres primitives

Au paragraphe 4.5, nous avons introduit un ensemble de critères géométriques couplés à ADCS visant à renforcer la similarité de l'apparence visuelle des régions retrouvées. De la même manière, la combinaison d'autres primitives de texture ou de structure locale (voir section 4.2) pourrait améliorer la discrimination visuelle de la mesure de similarité. Cependant, plus le nombre de descripteurs combinés au sein d'une seule distance est grand plus le problème du choix de la combinaison optimale se pose. La combinaison optimale ne peut pas être définie a priori et, alternativement, il ne serait pas raisonnable de laisser l'utilisateur régler manuellement l'importance relative d'une dizaine de descripteurs pour effectuer sa recherche.

Nous pensons que les efforts doivent plutôt se concentrer sur la définition de "descripteurs intégrés" capables de caractériser ces différentes primitives par une seule grandeur homogène.

Transposition à d'autres descripteurs couleur

La notion de nuances de couleurs (couplée à la distance quadratique) pourrait être étendue aux variantes de l'histogramme couleur classique évoquées en

4.2 (par exemple les *geometric histograms* [89]). Ces variantes intègrent une information sur la structure couleur locale ou l'arrangement spatial des couleurs dans l'image. Bien que les régions portent en elles l'information spatiale de leur position, surface et forme, ces variantes d'histogrammes ajouteraient une information spatiale supplémentaire *à l'intérieur* des régions. La transposition de ces histogrammes sur l'ensemble des nuances de couleurs est envisagée et serait susceptible d'améliorer encore la pertinence des visuelle des régions retrouvées.

Passage à l'échelle et accélération de la recherche

Les recherches de régions s'opèrent par comparaison séquentielle avec toutes les régions de la base. La base n'est pas préstructurée. Nous avons vu qu'avec 11.479 images (donnant 56.374 régions) le temps de réponse est faible : 0.5 seconde. Dans le domaine des Bases de Données on estime que jusqu'à 2 secondes, l'utilisateur n'a pas le sentiment d'attendre le résultat de sa requête.

La manipulation de plus grandes bases nécessite l'accélération du processus de recherche. La première solution serait d'optimiser la calcul de la distance quadratique en utilisant la distance minorante rapide à calculer proposée dans [41]. La seconde solution consisterait à réduire l'espace de recherche en préstructurant la base à l'aide, par exemple, de catégories de régions. L'approche de catégorisation de régions proposée au chapitre suivant pourrait être exploitée dans ce sens.

4.10 Conclusions

Dans ce chapitre, nous avons présenté notre approche de description fine de régions issues d'une extraction grossière pour le paradigme 1 de recherche d'images par région exemple.

Afin de tenir compte de la spécificité visuelle des régions dans une base, nous avons proposé le nouveau descripteur ADCS de distribution adaptative de nuances de couleur. Tout en étant compact, il offre une description fine et adaptative de la variabilité couleur des régions. Contrairement aux représentations habituelles de la couleur qui reposent sur une palette de couleurs commune à toute la base, les nuances de couleur pertinentes pour chaque région sont propres à chacune d'elles. Nous avons aussi présenté la forme généralisée de la distance quadratique, utilisée comme mesure de similarité privilégiée pour ADCS. Nous avons montré que la finesse de description des régions conduisait à des gains importants en précision comparée à la représentation imprécise de la couleur dans les descripteurs usuels.

Nous avons aussi proposé la combinaison de ce nouveau descripteur avec des critères géométriques simples. Selon la nature des "objets" recherchés, cette com-

binaison conduit à une amélioration supplémentaire de la pertinence visuelle des régions retrouvées. Par ailleurs, nous avons envisagé le développement de nouveaux descripteurs de régions combinant notre représentation fine des couleurs à l'information de configuration spatiale des couleurs.

Chapitre 5

Paradigme 2 : recherche d'images par composition logique de catégories de régions

Qu'ils soient globaux ou partiels, la plupart des systèmes existants de recherche d'images reposent sur le même paradigme générique de *recherche par l'exemple*. Selon s'il s'agit de recherche d'images par leur apprence visuelle globale ou partielle, l'exemple peut être une image, une région (voir paradigme 1) ou plusieurs régions ou éventuellement un croquis dessiné par l'utilisateur.

Le *paradigme 2* que nous présentons dans ce chapitre diffère des approches existantes pour la recherche d'information visuelle par le contenu. Il permet à l'utilisateur d'atteindre, si elles existent dans la base, les images recherchées à partir de la représentation mentale qu'il en a (*recherche par image mentale*). Ce nouveau paradigme apporte une solution au problème de la page zéro [16]. Aucune image ou région exemple n'est nécessaire à la formulation d'une requête visuelle. Pour formuler une requête visuelle, l'utilisateur interagit avec le résumé visuel des composantes visuelles de toutes les images de la base. Généré de façon non-supervisée, ce résumé est constitué par le thesaurus photométrique des régions d'intérêt des images. A l'aide d'opérateurs de requêtes logiques, l'utilisateur peut rechercher des images par la composition de leur contenu visuel. Les requêtes logiques sur la composition des images s'apparentent à celles employées en recherche de texte.

Nous verrons que ce nouveau paradigme conduit naturellement à des perspectives riches pour la recherche d'information visuelle.

5.1 Introduction

Le nouveau "paradigme 2" diffère du paradigme 1, étudié dans le chapitre précédent, en termes d'usage et de développements scientifiques. Le paradigme 1 de recherche par région exemple a été présenté comme le paradigme canonique de la recherche d'images par régions. Il permet de proposer des solutions aux problèmes fondamentaux de détection de régions, de leur description visuelle et de mesure de similarité adéquate. En termes d'usage, le but du paradigme 1 est de retrouver parmi toutes les images de la base celles comportant des régions visuellement similaires à une région exemple. Ce paradigme suppose donc que l'utilisateur dispose au préalable d'une image comportant une région d'intérêt proche de ce qu'il recherche.

De façon plus générale en recherche d'information visuelle, le paradigme existant de la *recherche par l'exemple* repose sur l'hypothèse que l'utilisateur dispose d'une image ou une région de départ. Or la plupart du temps, l'utilisateur ne dispose que d'une représentation *mentale* des images cible. Dans ce cas, la recherche par l'exemple nécessite la recherche préalable d'une image ou région exemple et s'avère fastidieuse, voire impossible dans de grandes bases.

Le paradigme 2 que nous présentons permet de s'affranchir de cette contrainte. Cette approche diffère de l'existant aussi bien du point de vue de la formulation de la requête que du processus de recherche. Dans l'interface de requête, le **thesaurus photométrique des régions** de la base donne à l'utilisateur un aperçu des régions présentes dans la base. A partir de celui-ci, l'utilisateur peut formuler des requêtes évoluées en précisant la composition des images cible. Le système peut répondre à des requêtes aussi complexes que : "trouver les images composées de régions de ces types mais pas de ces types".

La génération du thesaurus repose sur le regroupement non-supervisée des régions de la base en **catégories de régions**. Chaque catégorie regroupe les régions visuellement similaires et permettent à l'utilisateur d'indiquer dans l'interface de requête la présence et l'absence des "types" de régions qui caractérisent les images recherchées. Ce nouveau paradigme de **recherche d'images par composition logique de catégories de régions** s'apparente à celui de la recherche de texte. Nous verrons que la spécification par l'utilisateur du contenu de l'image recherchée introduit de la "sémantique visuelle" dans la requête.

L'implantation effective de cette approche est simple et complètement non-supervisée. Nous verrons que le couplage de l'*indexation symbolique* avec la formulation des requêtes logiques constitue un outil rapide et puissant de recherche dans de grandes bases d'images. La simplicité et l'originalité de ce principe ouvre la voie à de nombreuses perspectives que nous évoquerons.

Dans la section 5.2, nous résumerons le déroulement de cette nouvelle ap-

proche. Dans la section 5.3, nous présenterons le principe de la catégorisation des régions pour la recherche de similarité. Nous définirons les catégories de régions ainsi que la notion de catégories voisines. En section 5.4, nous détaillerons le principe de *recherche d'images par composition logique de catégories de régions* ainsi que son implantation effective. L'interface graphique de recherche et le thesaurus photométrique des régions seront introduits en section 5.5 en même temps que les résultats. Des scénarios de recherche se seront présentés sur la base d'images Corel et sur une base d'images extraites d'archives video du journal télévisé de TF1. En section 5.6, les discussions porteront sur plusieurs extensions envisageables pour cette nouvelle approche.

5.2 Résumé et déroulement de l'approche

Le déroulement de l'approche est résumé par la figure 5.1 de l'étape de prés-tructuration de la base jusqu'à celle recherche des images pertinentes pour une requête logique donnée.

Les régions sont, comme pour le paradigme 1, détectées par l'algorithme de segmentation présenté au chapitre 3. Dans le cadre du paradigme 2, elles constituent les "clés" de la recherche d'images par composition. Une première étape consiste à effectuer le regroupement non-supervisé des régions similaires de la base en catégories. A chaque catégorie sont associées ses catégories voisines qui contiennent des régions proches dans l'espace de description. Au moment de la recherche les régions seront considérées comme similaires si elles a ppartiennent à la même catégorie ou à des catégories voisines. L'indexation des images se réduit à la liste des étiquettes de catégories auxquelles ses régions appartiennent. L'interface de requête présente à l'utilisateur le thesaurus photométrique des régions de la base constitué des régions représentatives de chaque catégorie. Par sélection de catégories requête, il contraint la présence et l'absence de différents types de régions qu'il juge caractéristiques de sa représentation mentale de l'image recherchée. Il formule ainsi la requête logique par composition. Grâce à l'indexation symbolique des images, le système détermine les ensembles d'images qui sont composées de chaque catégorie requête puis, par opérations ensemblistes, les images pertinentes pour la requête par composition sont retournées. Le raffinement de la recherche peut porter sur la modification du choix des types de régions et de l'étendue de leur similarité visuelle.

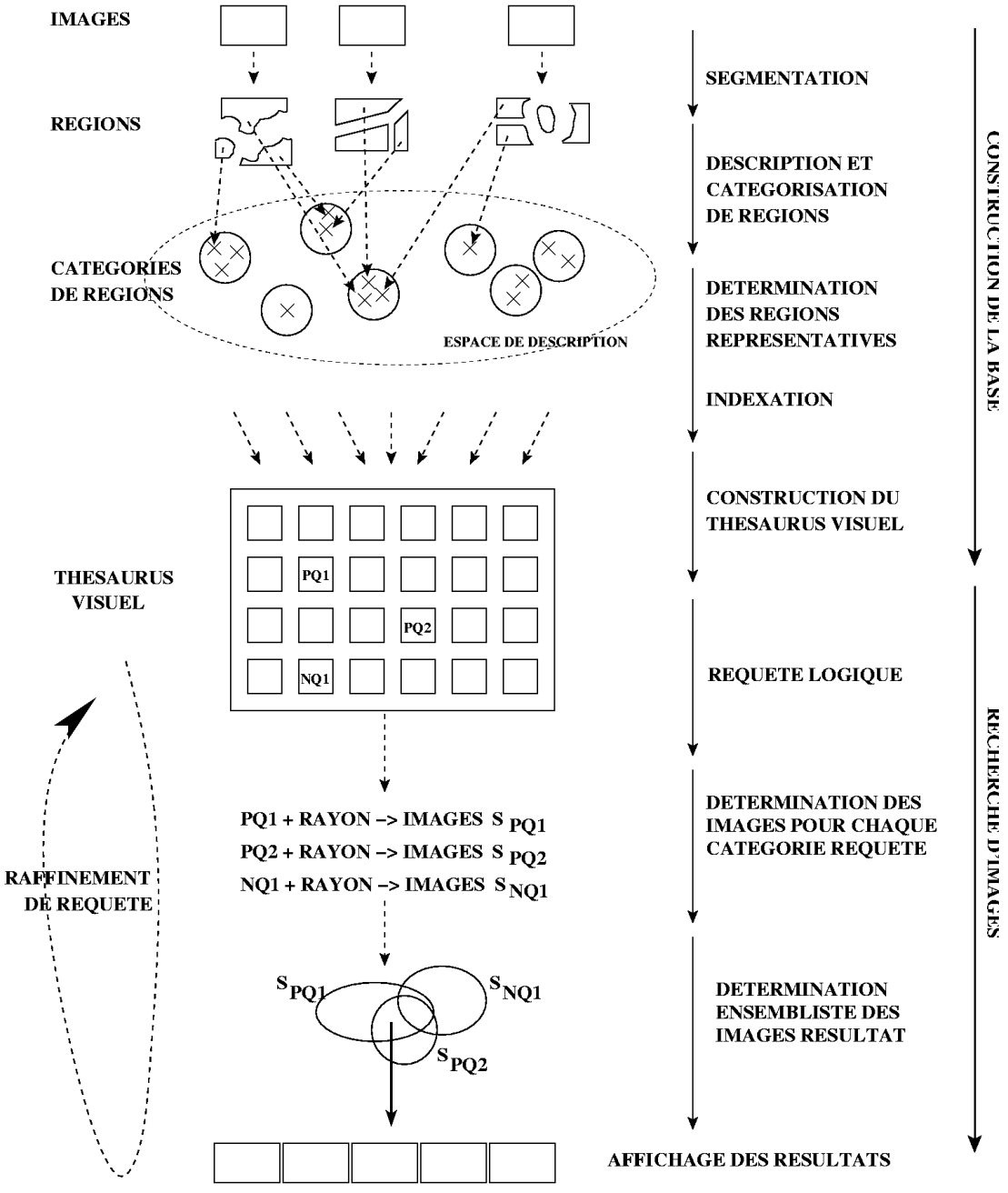


FIG. 5.1 – Etapes de structuration de la base et de requête par composition logique de catégories de régions.

5.3 Catégorisation de régions pour la recherche de similarité

Définition 5. Etant donnée une base d'images segmentées en régions $\{R_1, ..., R_N\}$, les **catégories de régions** $C_1, ..., C_P$ de cette base sont ainsi définies :

- $C_i \neq \emptyset$
- $\cup_{i=1}^{P} C_i = \cup_{j=1}^{N} R_j$
- $C_i \cap C_h = \emptyset, \forall i \neq h$
- $\forall R_j, R_k \in C_i$, R_j et R_k sont d'apparences visuelles très proches

La catégorisation d'images est définie comme le regroupement non-supervisé de leurs descripteurs visuels.

Les catégories définissent les "types" de régions à partir desquels l'utilisateur pourra formuler sa requête de composition.

5.3.1 Représentation des régions dans l'espace de description

L'apparence visuelle des régions est décrite ici par leur couleur moyenne et les groupes seront alors formés de régions de couleurs moyennes proches. Il est important de noter que tout autre descripteur que la couleur moyenne peut être utilisé.

Afin de comprendre les prérequis de la catégorisation des régions, nous avons représenté, dans l'espace Luv, les 50.220 points correspondant aux couleurs moyennes des 50.220 régions extraites des 50.220 images de la base Corel. De cette représentation tridimensionnelle (dont la figure 5.2 présente une vue), nous observons tout d'abord qu'une partie majoritaire de l'espace Luv ne contient aucune donnée. Ceci est dû à trois facteurs : 1. certaines valeurs de l'espace Luv n'ont pas de correspondant dans l'espace de départ RGB, 2. les couleurs saturées, localisées aux sommets, sont minoritaires dans une base d'images naturelles 3. le moyennage des couleurs d'une région est une fonction convexe.

Par ailleurs, nous avons pu observer un regroupement relativement compact des points, avec un maximum de densité autour de l'axe L de luminance. Sur l'ensemble des données, on ne peut pas observer de groupement naturel des points. Cette remarque semble cohérente avec la nature hétérogène des images de la base, donc des régions extraites. De plus, l'effet de compacité globale, autrement dit de *continuum* des points est destiné à s'accentuer sur des bases encore plus grandes. Notons que le problème de regroupement rencontré ici est différent de celui de segmentation étudié au chapitre 3 dont le but est d'exhiber les groupements

naturels de primitives locales qui forment les régions saillantes de chaque image. Nous allons voir ici que l'absence de regroupement naturel n'est pas un problème.

Dans notre approche de recherche par catégories de régions, la base de la similarité entre les régions est l'appartenance aux catégories générées. En l'absence d'hypothèse de groupes naturels de régions dans l'espace de description, définir la similarité entre les régions comme la simple appartenance à une même catégorie n'est pas suffisant. En effet, pour toute catégorisation générée, non réduite à une seule catégorie, il existera toujours des régions pouvant être visuellement similaires mais qui appartiennent à des catégories différentes. C'est pour cette raison que nous introduirons la notion de "catégories voisines" qui permettra de définir la similarité entre les régions comme l'appartenance à une même catégorie *ou bien* à une catégorie voisine. La question du choix des catégories voisines sera abordé en section 5.4. Dans un premier temps, nous allons nous intéresser à la définition des catégories et des catégories voisines.

5.3.2 Catégorisation de régions et catégories voisines

La catégorisation des régions en catégories est obtenu par l'algorithme CA (voir 2.4) avec une granularité fine de classification. Les primitives regroupées sont les couleurs moyennes des régions dans l'espace Luv. L'homogénéité des catégories est obtenue avec un nombre de classes relativement élevé et en autorisant des classes faiblement peuplées. et une population minimale faible. Dans CA, la granularité fine est contrôlée à l'aide d'une population minimale faible $\epsilon = 0.005$, un nombre de classes initiales de 100 (qui constitue une surestimation de nombre de catégories désirées) et une importance relativement faible accordée à l'agglomération de classes avec $\eta = 4$ et $\tau = 7$. L'initialisation des 100 prototypes est effectuée par une sélection aléatoire parmi les données. Cette granularité correspond à un compromis entre l'homogénéité des régions au sein des différentes classes et un nombre raisonnable de catégories pour constituer le thesaurus de régions dans l'interface de requête. A chaque catégorie C_q correspond un prototype p_q donné par la classification. Ils correspondent aux centroïdes de classes.

Dans le cas de la base Corel comportant $9,995$ images dont sont extraites 50.220 régions, la classification des 50.220 triplets de couleur est effectuée dans l'espace Luv.

91 catégories sont obtenues dont les populations varient de 112 à 2048 régions. La figure 5.2 illustrent les 50.220 points dans l'espace couleur ainsi que les 91 prototypes obtenus pour chaque catégorie.

Bien que les données s'étalent selon un certain continuum et qu'elles n'occupent qu'une fraction de l'espace Luv, CA a pu détecter les prototypes avec espacement régulier et dans les différents lieux de l'espace présentant une densité notable de points. Inversement, aucun prototype n'est détecté dans les lieux de

l'espace ne comportant aucune ou très peu de données. Nous verrons au moment des résultats que les catégories les plus peuplées correspondent aux couleurs moyennes faiblement saturées, car c'est en ces lieux que les densités sont les plus fortes.

Les observations détaillées en section 5.3.1 nous conduisent à ne pas supposer de regroupement naturel des régions dans l'espace de description. Dans un souci de généralisation à divers types de bases, nous n'exploiterons aucune information a priori sur la nature de la base pour former les catégories de régions.

La stratégie adoptée consiste alors à combiner la *finesse de la granularité* de la classification avec la prise en compte des catégories voisines et la fonctionnalité de "range-query" sur les catégories. Nous allons voir que cette stratégie est cruciale pour la recherche des régions similaires dans l'espace de description.

Imposer une granularité fine de classification conduit à la formation de classes présentant une forte homogénéité. Etant donné que la similarité entre les régions sera définie d'abord comme l'appartenance à une même catégorie, une granularité fine de classification permet de retrouver les régions très proches visuellement. Afin de retrouver les régions selon des degrés *variables* de similarité, le système prendra en compte les catégories proches dans l'espace de description des régions, appelées "catégories voisines". La prise en compte de ces "catégories voisines" dans notre système permet de formuler des "range-query" sur les types de régions.

Définition 6. La **catégorie voisine** d'une catégorie C_q de prototype p_q est définie comme la catégorie C_j dont le prototype p_j satisfait $d(p_q, p_j) \leq \gamma$, pour un seuil donné γ, appelé **rayon de recherche**. La distance $d(.,.)$ est celle associée à l'espace de description des régions (donc L^2 ici pour l'espace Luv).

Définition 7. Pour toute catégorie C_q, nous définissons l'**ensemble des catégories voisines** de C_q, noté $V^\gamma(C_q)$, l'ensemble suivant :

$$V^\gamma(C_q) = \{(V_1, d_1), (V_2, d_2), ..., (V_K, d_K) \mid d_K \leq \gamma\}$$

où les V_j sont les catégories voisines de C_i triées par distance croissante d_j entre les prototypes, où $d_j = d(C_i, V_j)$. Par convention, nous définissons V_1 comme étant C elle-même (i.e. $d_1 = 0$).

Nous introduisons par ailleurs les ensembles suivants :

Ensemble de catégories $CI(I)$ **:** catégories auxquelles appartiennent les régions qui composent l'image I.

Ensemble d'images $IC(C)$ **:** images composées d'au moins une région appartenant à la catégorie C. C'est la table inversée de $CI(I)$.

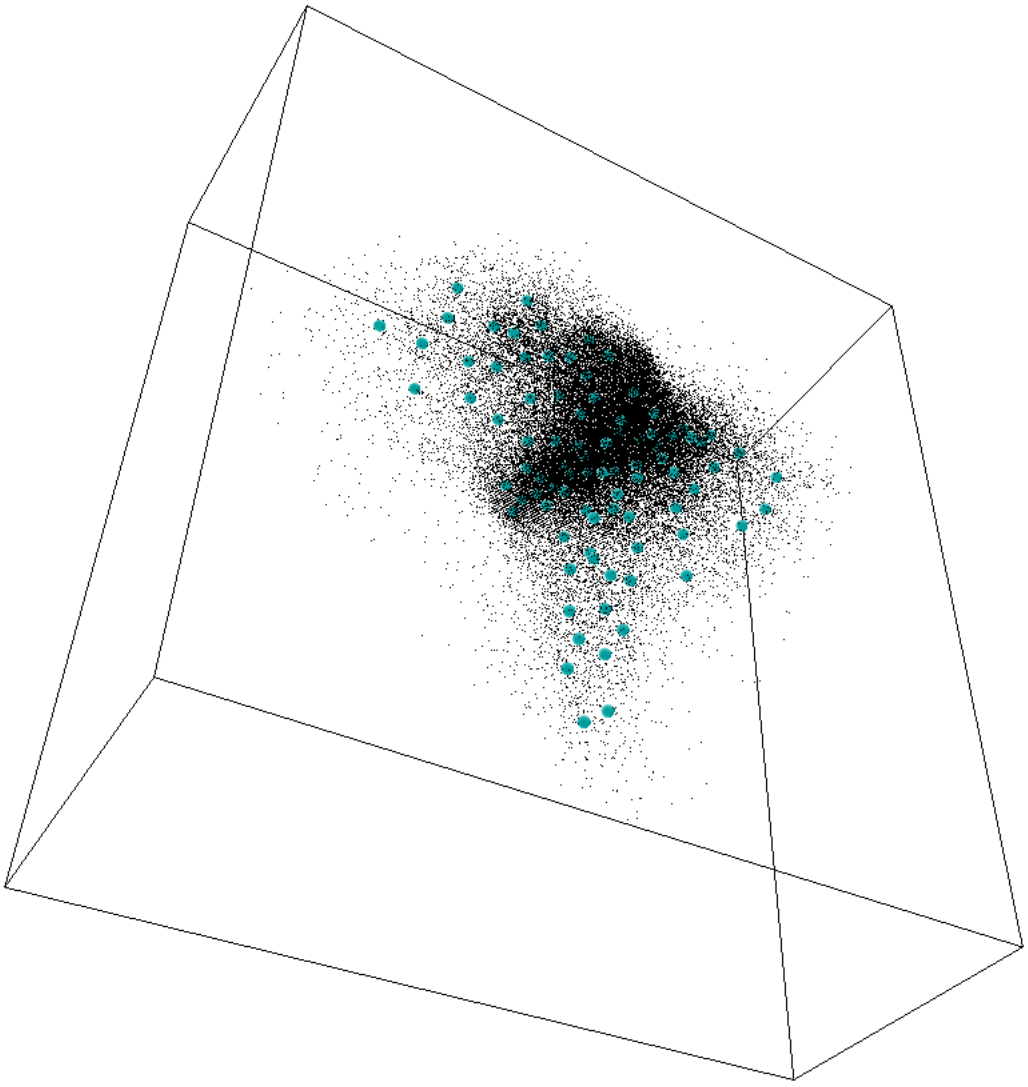


FIG. 5.2 – **Prototypes des catégories obtenus par CA :** Les points identifient les couleurs moyennes de 50.220 régions de la base Corel et les sphères représentent les prototypes des 91 catégories Le parallélépipède représente l'espace LUV entier. Malgré la relative compacité des données et l'absence de groupes naturels, l'algorithme CA parvient de façon non-supervisée à trouver les prototypes repartis équitablement dans les densités de points. Les plus importantes densités sont observées le long de l'axe L de luminance. On remarque qu'une faible partie de l'espace est occupée par l'ensemble des points (voir explications dans le texte).

Pour toute catégorie C et pour toute image I, $V^{\gamma}(C_q)$, $CI(I)$ et $IC(C)$ définissent trois tables d'association (illustrées par la figure 5.3). Celles-ci vont constituer la base du principe d'indexation.

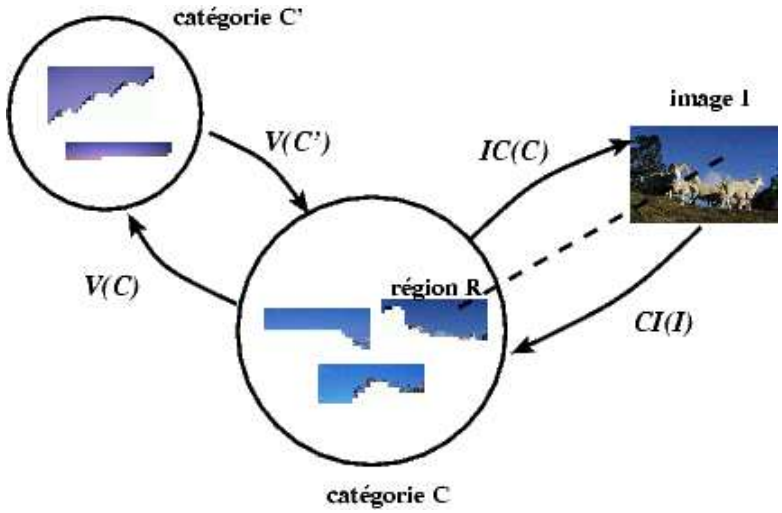


FIG. 5.3 – Les trois tables d'association $V^{\gamma}(C), CI(I), IC(C)$ nécessaires à l'indexation fournissent les informations suivantes : $V(C)$ est la catégorie voisine de C, $CI(I)$ est une catégorie ayant une région de I et $IC(C)$ est une image contenant une région (la région R) de la catégorie C.

Pour constituer le thesaurus photométrique de régions dans l'interface de requête, chaque catégorie sera illustrée par sa région représentative.

Définition 8. La **région représentative** d'une catégorie C est la région dont la couleur moyenne est la plus proche du prototype de C dans l'espace de description des régions[1].

Lors de la recherche, l'utilisateur sélectionnera un ensemble de catégories-requête qui correspondent aux types de régions devant apparaître (ou ne pas apparaître) dans les images retrouvées. Pour chaque catégorie requête C, seront définies comme similaires les régions appartenant à $V^{\gamma}(C)$. Plus le rayon de recherche γ sera fixé à une faible valeur au moment de la requête, moins de catégories voisines seront prises en compte et plus les régions considérées pour la recherche seront similaires. La combinaison de catégories homogènes de régions avec l'intégration de catégories voisines est un point-clé dans la définition du procédé de recherche par "range-query". Il permet d'effectuer des recherches efficaces

[1]Notons que les prototypes sont calculés dans CA par une moyenne des données et ne correspondent donc pas nécessairement à une région effective de la base.

dans l'espace des descripteurs même en l'absence d'hypothèse de regroupement naturel des régions dans l'espace de description.

5.4 Recherche d'images-cible par composition de catégories

Dans le paradigme 2, la structuration des régions de la base en catégories est directement liée à la formulation de la requête par l'utilisateur. Dans l'interface graphique, chaque catégorie est illustrée par sa région representative (voir section précédente) afin que l'utilisateur sélectionne les catégories visuellement pertinentes pour caractériser l'image mentale recherchée. Dans le but de répondre à des requêtes telles que "trouver des images composées de régions de ces types et pas de ces types", l'utilisateur pourra sélectionner un ensemble de catégories requête positives et un ensemble de catégories requête négatives.

Etant donné l'ensemble des catégories requête sélectionnées, nous allons présenter le principe de recherche par composition logique de catégories de régions afin de déterminer les images pertinentes pour la requête de l'utilisateur.

5.4.1 Principe

Dans cette section, nous allons développer l'approche formelle de requête par composition en nous intéressant à l'expression d'une composition logique simple puis complexe. Nous déterminerons graduellement l'ensemble des images satisfaisant la composition requête grâce à l'introduction d'opérateurs logiques.

Nous introduisons les notations suivantes :

Catégories Requête Positives ou "CRP" : ensemble des catégories sélectionnées par l'utilisateur devant apparaître dans les images retournées. Elles sont notées $\{C_{pq_1}, C_{pq_2}, ..., C_{pq_M}\}$.

Catégories Requête Negatives ou "CRN" : ensemble des catégories sélectionnées par l'utilisateur *ne devant pas* apparaître dans les images retournées. Elles sont notées $\{C_{npq_1}, C_{npq_2}, ..., C_{npq_R}\}$.

Ensemble des images retournées S_{res} **:** images de la base ayant une composition de catégories de régions pertinente pour la requête.

Les sélections de "CRP" et de "CRN" constituent la requête de l'utilisateur et S_{res} est l'ensemble d'images à déterminer.

Recherche de présence d'un type de région

La requête la plus simple consiste à retrouver les images composées d'au moins une région appartenant exclusivement à une catégorie CRP. Si C_q désigne cette CRP, l'ensemble des images pertinentes s'écrit simplement comme suit :

$$IC(C_q)$$

Les régions retrouvées sont visuellement très similaires car appartiennent à la même catégorie.

Afin d'étendre la similarité des régions par range-query, nous souhaitons ensuite prendre en compte aussi les catégories voisines de C_q. En ajustant le rayon de recherche γ au moment de la requête, l'utilisateur peut adapter le degré de similarité selon la nature des régions recherchées. Lorsque nous chercherons les images composées de régions provenant de la CRP C_q, nous chercherons aussi les images composées de régions de C_q `OU` de régions provenant des catégories voisines de C_q. La disjonction `OU` se traduit par une *union* d'ensembles d'images :

$$\bigcup_{C \in V^{\gamma}(C_q)} IC(C) \tag{5.1}$$

Ceci constitue le fondement de la recherche de catégories de régions par range-query. L'influence du rayon de recherche γ sur la définition des catégories voisines est illustrée dans la figure 5.4.

Recherche de présence de plusieurs types de régions

A présent nous allons étendre la requête à plus d'une CRP. Nous supposons que l'utilisateur a sélectionné M CRP notées : $C_{pq_1}, C_{pq_2}, ..., C_{pq_M}$. Nous recherchons les images ayant une région dans C_{pq_1} ou ses voisines `ET` une région dans C_{pq_2} ou ses voisines `ET` .... `ET` une région dans C_{pq_M} ou ses voisines. La conjonction `ET` se traduit par l'*intersection* des ensembles d'images obtenus dans l'expression (5.1) :

$$S_{PQ} = \bigcap_{i=1}^{M} \Big[\bigcup_{C \in V^{\gamma}(C_{pq_i})} IC(C) \Big] \tag{5.2}$$

L'ensemble S_{res} d'images retournées s'écrit : $S_{res} = S_{PQ}$. Voir l'illustration 5.5.

Recherche de présence et d'absence de plusieurs types de régions

En plus de l'ensemble de catégories requête positives $C_{pq_1}, C_{pq_2}, ..., C_{pq_M}$, nous supposons que l'utilisateur a en plus sélectionné un ensemble de catégories requête

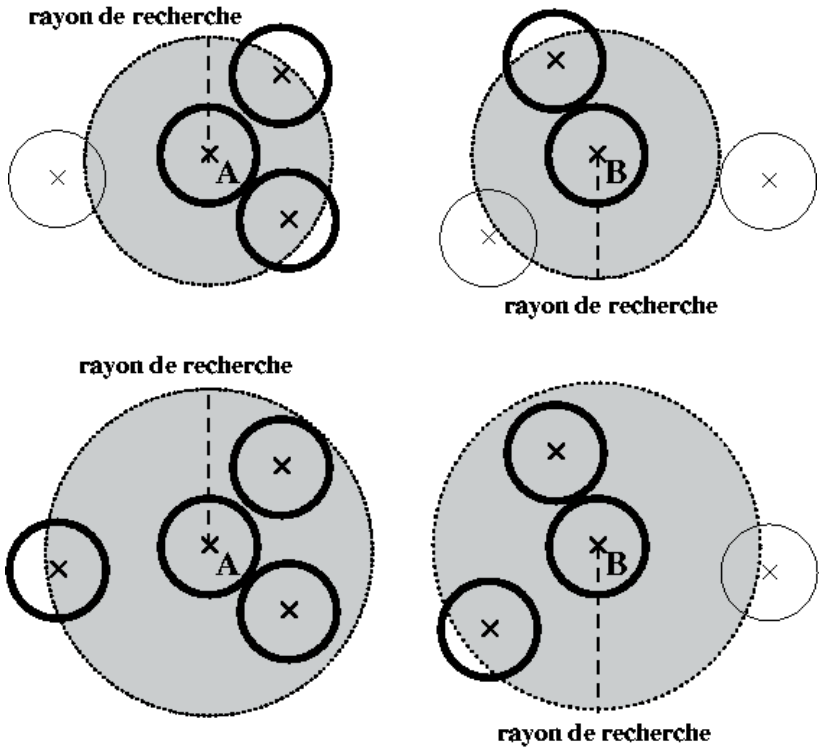


FIG. 5.4 – Range query par sélection des catégories voisines. A et B désignent deux catégories requête. Le choix du rayon de recherche détermine les ensembles des catégories voisines autour de A et B (respectivement $V^{\gamma}(A)$ et $V^{\gamma}(B)$). Un rayon faible (figure du haut) et un rayon élevé (figure du bas) vont couvrir plus ou moins de catégories voisines pour définir les types de régions recherchées. Les disques grisés de rayon de recherche γ indiquent les catégories voisines sélectionnées. Les catégories voisines sont tracées avec des contours plus épais que les autres catégories. Les prototypes sont repérés par des croix.

négatives $C_{nq_1}, C_{nq_2}, ..., C_{nq_R}$. Elles désignent les types de régions ne devant pas apparaître dans les images recherchées. Nous allons prendre aussi en compte les catégories voisines des catégories requête négatives.

L'ensemble S_{NQ} d'images "indésirables" contenant les CRN et leurs voisines s'écrit alors :

$$S_{NQ} = \bigcap_{i=1}^{R} \Big[\bigcup_{C \in V^{\gamma}(C_{nq_i})} IC(C) \Big] \tag{5.3}$$

La formulation de la requête consiste en une liste d'indices de CRP $\{pq_1, pq_2, ..., pq_M\}$ et de CRN $\{nq_1, nq_2, ..., nq_R\}$. Nous nous intéressons à l'ensemble S_{res} des images retournées qui sont composées de régions appartenant aux différentes CRP `ET PAS` de région appartenant aux CRN. La négation est traduite par la *soustraction ensembliste* de S_{PQ} (formule (5.1)) avec S_{NQ} :

$$\begin{aligned} S_{res} &= S_{PQ} \setminus S_{NQ} \\ &= \bigcap_{i=1}^{M} \Big[\bigcup_{C \in V^{\gamma}(C_{pq_i})} IC(C) \Big] \setminus \bigcap_{i=1}^{R} \Big[\bigcup_{C \in V^{\gamma}(C_{nq_i})} IC(C) \Big] \end{aligned} \tag{5.4}$$

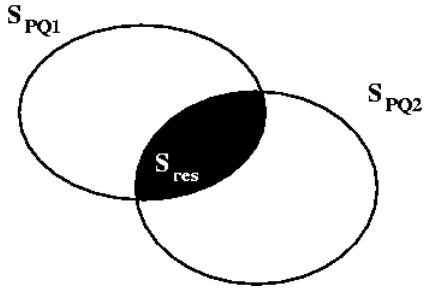


FIG. 5.5 – Composition de Catégories Requête Positives : l'ensemble des images résultats pour la recherche de présence de 3 types de régions est traduite par l'*intersection* de 3 ensembles d'images $S_{PQ_1}, S_{PQ_2}, S_{PQ_3}$, associés aux trois catégories requête positives. Il s'écrit : $S_{res} = S_{PQ_1} \bigcap S_{PQ_2} \bigcap S_{PQ_3}$.

Ce dernier mode de recherche regroupe les trois fonctionnalités de recherche suivants : range-query, requête par la négative, requête par types multiples (plusieurs CRP et plusieurs CRN). Ces trois modes ne sont mis en oeuvre que pour les requêtes les plus complexes. Cependant, l'expression (5.4) se réduit dans le cas de requêtes plus simples, par exemple si l'utilisateur ne spécifie qu'une seule CRP ou aucune CRN ou bien un rayon de recherche nul.

Le tableau 5.1 résume les fonctionnalités de requête de la méthode de recherche par composition avec les opérateurs logiques correspondants.

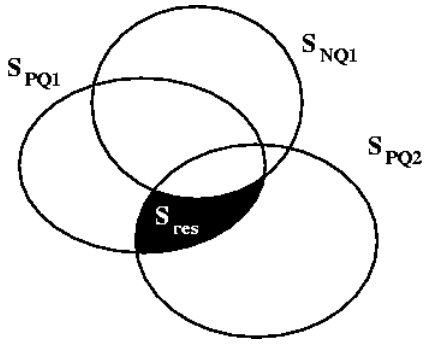


FIG. 5.6 – Composition par Catégories Requête Positives et *Négatives* : par rapport à la figure 5.5, nous avons ajouté un ensemble d'images S_{NQ_1} ayant des régions dans une catégorie requête négative (ou ses voisines). L'ensemble des images résultats avec la contrainte supplémentaire d'absence de régions se traduit par la *soustraction* de S_{NQ_1}. avec les ensembles d'images $S_{PQ_1}, S_{PQ_2}, S_{PQ_3}$.
Il s'écrit : $S_{res} = S_{PQ_1} \bigcap S_{PQ_2} \bigcap S_{PQ_3} \setminus S_{NQ_1}$.

mode	opérateur	usage
"range query"	`OU` ($\vee$)	prise en compte des catégories voisines pour adapter l'étendue de la recherche de régions
requête multiple	`ET` ($\wedge$)	contrainte de présence de plusieurs types de régions
requête par la négative	`ET NON` ($\wedge \neg$)	contrainte d'absence de plusieurs types de régions

TAB. 5.1 – Résumé des trois fonctionnalités de requête possibles avec leurs opérateurs logiques et leur usages associés.

Ce mode de requête permet de retomber sur des mécanismes existants en recherche d'information et en bases de données. L'usage de la table d'association catégorie/image $IC(C)$ est à rapprocher des tables inversées (ou "inverted tables") qui permet un accès direct aux données pertinentes (images, ici). Par ailleurs, les conjonctions et négations dans la formulation des requêtes (voir table 5.1) sont à rapprocher du modèle booléen en recherche d'information.

Nous avons en plus introduit dans les requêtes les disjonctions avec les catégories voisines. Ceci a été motivé par la nature *continue* de nos données (descripteurs visuels) qui diffèrent des objets manipulés généralement en recherche d'information et en bases de données qui sont de types "meta-données".

5.4.2 Implantation algorithmique

La recherche par composition logique de catégories de régions, dont nous venons de présenter le principe, répond à requêtes telles que : "trouver des images composées de régions dans ces CRP et pas de région de ces CRN".

La traduction de ces requêtes logiques par des calculs d'unions et d'intersections sur des ensembles d'images va se prêter naturellement à une mise en oeuvre efficace de ce principe.

A partir d'une liste de CRP et de CRN et du rayon de recherche γ sélectionnés par l'utilisateur, le but est d'évaluer l'expression (5.4) de S_{res} qui définit l'ensemble des images pertinentes pour la requête. L'approche naïve consisterait à déterminer si chaque image contient des régions appartenant aux CRP (et leurs voisines) et aucune appartenant aux CRN (et leurs voisines). Pour une base comportant des dizaines de milliers de régions (50.220 pour la base Corel présentée) et des requêtes constituées typiquement de trois CRP et une CRN en tenant compte de leurs voisines respectives, le nombre de comparaisons que nécessiterait une requête serait prohibitif avec cette approche. Notons qu'une comparaison consiste ici à déterminer si une étiquette de catégorie (donc un entier) est associée à une image donnée. Nous allons voir qu'en utilisant les propriétés d'intersections ensemblistes et en choisissant une stratégie d'indexation adéquate, ce nombre de comparaisons pourra être réduit simplement. La stratégie présentée permet d'accéder directement aux sous-ensembles pertinents d'images de la base sans perte de précision.

Afin de satisfaire la requête, c'est-à-dire déterminer S_{res}, nous allons initialiser S_{res} avec un sur-ensemble des images pertinentes puis éliminer celles qui ne satisfont pas toutes les contraintes de la requête. Etant donné que S_{res} s'exprime comme des intersections et des soustractions d'ensembles d'images, l'idée est de l'initialiser avec l'un des ensembles d'images et ensuite de supprimer les images

qui n'appartiennent pas aux autres ensembles. Cette initialisation évite l'accès à chaque image de la base et permet de démarrer la recherche directement à partir d'un ensemble potentiellement pertinent d'images qui constitue un sur-ensemble de S_{res}. Développons l'expression (5.4) de l'ensemble S_{res} que nous cherchons à déterminer :

$$\begin{aligned} S_{res} =& S_{PQ} \setminus S_{NQ} \\ =& \bigcap_{i=1}^{M} \Big[\bigcup_{V^\gamma(C_{pq_i})} IC(C) \Big] \setminus S_{NQ} \\ =& \bigcup_{V^\gamma(C_{pq_1})} IC(C) \cap \bigcap_{i=2}^{M} \Big[\bigcup_{V^\gamma(C_{pq_i})} IC(C) \Big] \setminus S_{NQ} \\ =& S_{PQ_1} \cap (S_{PQ_2} \cap \ldots \cap S_{PQ_M}) \setminus S_{NQ} \end{aligned} \tag{5.5}$$

où $S_{PQ_i} = \bigcup_{V^\gamma(C_{pq_i})} IC(C)$.

Dans la dernière ligne de (5.5), nous venons d'isoler l'ensemble S_{PQ_1}. Il correspond à l'ensemble des images ayant une région dans la CRP 1 ou dans l'une de ses voisines et constitue donc un *sur-ensemble* de S_{res}. L'ensemble S_{PQ_1}, quel qu'il soit, est donc d'un candidat à l'initialisation de S_{res}. Pour déterminer S_{res}, nous supprimerons parmi les images de S_{PQ_1} celles qui n'appartiennent pas à S_{PQ_2} et celles qui appartiennent à S_{NQ}.

L'algorithme d'évaluation de S_{res} se déroule ainsi :

1. $S_{res} = S_{PQ_1}$. Initialisation de S_{res} avec l'ensemble S_{PQ_1} (usage des tables IC et V^γ).
2. $S_{res} = S_{res} \cap S_{PQ_2}$. Suppression dans S_{res} des images qui n'ont de régions ni dans C_{pq_2}, ..., ni dans C_{pq_M}, ni dans aucune de leurs voisines (usage des tables CI et V^γ). A ce stade, S_{res} correspond à la partie "positive" de la requête, i.e. $S_{res} = S_{PQ}$ (formule (5.2)).
3. $S_{res} = S_{res} \setminus S_{NQ}$. Suppression dans S_{res} des images qui ont une région dans C_{nq_1}, ..., ou dans C_{nq_R}, ou dans l'une de leurs voisines (usage des tables CI et V^γ). A ce stade, on a $S_{res} = S_{PQ} \setminus S_{NQ}$ (formule (5.4))

Progressivement, S_{res} est réduit de S_{PQ_1} à $S_{PQ} \setminus S_{NQ}$. Grâce à cette stratégie, nous verrons en section 5.5.6 que l'on évite l'accès inutile à une importante fraction de l'ensemble des images de la base par rapport à l'approche naïve.

Nous introduisons une optimisation supplémentaire de l'algorithme de recherche par une considération sur le tri des S_{PQ_i} S_{res} est initialisé comme l'ensemble S_{PQ_1} des images associées à la catégorie C_{pq_1} et à ses voisines. Au fil

de l'algorithme, il est réduit par intersections successives avec les ensembles $S_{PQ_2}, ..., S_{PQ_M}$. Ces différentes étapes de traitement impliquent des tests d'appartenance sur chaque image de l'ensemble S_{res}. Nous proposons de choisir un ordre d'étiquetage des $\{S_{PQ_i}\}$, donc des $\{C_{pq_i}\}$, qui réduise les coûts de calculs. Les M CRP sont interchangeables et, sans perte de généralité, on peut considérer que C_{pq_1} est la catégorie de cardinalité minimale parmi C_{pq_1}, ..., C_{pq_M}. Considérons que les CRP sont étiquetées par ordre de cardinalité croissante au moment de leur construction :

$$\mid C_{pq_1} \mid \leq \mid C_{pq_2} \mid \ ... \ \leq \mid C_{pq_M} \mid$$

En faisant l'hypothèse que les voisines d'une catégorie peu peuplée sont peu peuplées[2] et en remarquant que la fonction ensembliste $C \mapsto IC(C)$ est croissante, on en déduit :

$$\mid \bigcup_{V^\gamma(C_{pq_1})} IC(C) \mid \leq \mid \bigcup_{V^\gamma(C_{pq_2})} IC(C) \mid ... \leq \mid \bigcup_{V^\gamma(C_{pq_M})} IC(C) \mid$$

c'est-à-dire :

$$\mid S_{PQ_1} \mid \leq \mid S_{PQ_2} \mid ... \ \leq \mid S_{PQ_M} \mid \tag{5.6}$$

S_{res} étant initialisé par S_{PQ_1}, puis réduit par intersections successives avec les $S_{PQ_2}, ..., S_{PQ_M}$, le tri des C_{pq_1}, donc des $\{S_{PQ_i}\}$ (inégalités (5.6)), par cardinalités croissantes permet de diminuer le nombre de tests d'appartenance des images de S_{res}.

La stratégie d'initialisation de S_{res} par S_{PQ_1} et le tri des catégories par cardinalités par cardinalités croissantes sont deux facteurs qui permettent de réduire le temps de recherche par rapport à une approche naïve, avec des résultats identiques.

5.4.3 Indexation symbolique

Dans cette section, nous mettons en avant la nature symbolique de la méthode d'indexation sous-jacente sur laquelle repose le principe de recherche qui vient d'être décrit.

Les descripteurs de régions (couleur moyenne ici) ont servi à former les catégories mais ne sont plus exploités au moment de la recherche contrairement aux approches existantes. En effet, nous venons de voir que la détermination de S_{res} exploite uniquement les associations entre les images, les catégories de régions ainsi que les catégories voisines. Ces associations sont représentées par trois tables d'association : IC, CI et V^γ qui ont été définies en section 5.3.2.

[2]Ceci semble se vérifier sur nos données dans l'espace de description.

Dans l'algorithme de recherche (section précédente), la table d'association IC a un rôle de "point d'entrée" dans la base au moment de la requête. Elle permet d'initialiser l'ensemble des résultats instantanément comme un sous-ensemble de la base. La seconde table d'association CI permet ensuite de supprimer dans ce sous-ensemble les images qui ne vérifient pas l'ensemble des conjonctions de la requête logique. Finalement, la troisième table V^γ permet d'étendre la recherche aux catégories proches.

Ces trois tables d'association IC, CI et V^γ sont construites à l'issue de la génération des catégories. Au moment de la recherche, elles évitent le parcours séquentiel des images et des catégories. Elles renseignent sur les associations entre images, catégories et catégories voisines, mais pas sur les régions directement (voir illustration 5.3). Chaque image est indexée avec la liste des étiquettes des catégories auxquelles ses régions appartiennent.

En résumé, pour un rayon de recherche donné, un ensemble de catégories requête positives et negatives donnés, la recherche par composition de catégories de régions dans la base d'images se ramène à des accès aux tables d'association V, CI, IC. Au moment de la recherche, l'accès aux entités régions n'est plus nécessaire : nous n'exploitons avec les tables que les associations entre catégories et images (IC et CI) et entre catégories entre elles (V^γ).

Même pour des requêtes logiques évoluées, le processus de recherche est très rapide, d'une part grâce à l'accès direct aux groupes de régions similaires et, d'autre part car il n'implique que des opérations élémentaires sur des entiers, contrairement aux approches classiques qui nécessitent le calcul de distances entre des descripteurs multidimensionnels.

5.5 Interface Utilisateur et résultats

Dans cette partie, nous présentons d'abord l'interface utilisateur dans un cadre général puis nous verrons son fonctionnement en détail avec la présentation des résultats.

Notre approche a été testée sur deux scénarios d'application correspondant à des bases de natures différentes. Le premier correspond à un scénario d'agence photo de recherche dans la photothèque généraliste Corel. Le second est celui de recherche d'images dans des fonds d'archives video de journaux télévisés de la chaîne TF1.

5.5.1 Interaction utilisateur

L'interface de recherche par composition logique de catégories de régions permet la sélection d'un ensemble des catégories requête positives et négatives. Pour

choisir les catégories, l'utilisateur va pouvoir les visualiser afin de sélectionner celles qui correspondent aux types de régions dont sont composées son *image mentale*. L'élément central de l'interface de requête est le thesaurus photométrique de régions qui fournit un résumé de l'ensemble des types de régions de la base.

Une fois la requête formulée, l'interface de résultats affichera l'ensemble S_{res} des images pertinentes pour la requête par composition. Dans l'interface de résultats, s'il le souhaite, l'utilisateur pourra facilement modifier sa requête afin de raffiner la recherche.

Nous pourrons constater que le mode d'interaction présenté pour la recherche d'images est très différent des approches existantes du point de vue de la formulation de la requête, du processus de recherche et de la rapidité de traitement.

Interface de requête

Dans l'interface de requête, chaque catégorie est identifiée visuellement par sa région représentative (section 5.3.2). Etant définie à partir du prototype, elle constitue en quelque sorte la "région moyenne" de la catégorie. Nous définissons le **thesaurus photométrique de régions** comme l'ensemble des régions représentatives des différentes catégories de la base. Dans l'interface de requête, ce thesaurus permet à l'utilisateur de visualiser et de sélectionner les catégories pour sa requête. La figure 5.9 illustre le thesaurus des 91 catégories obtenues pour la base des 50.220 régions de la base Corel. Dans ce scénario, l'utilisateur n'a pas à naviguer dans la base pour rechercher une image ou une région exemple. Chaque catégorie de régions dans l'interface définit un type de régions qui peut potentiellement composer les images recherchées.

Sous chaque vignette de la région représentative d'une catégorie, l'utilisateur coche la case verte pour spécifier la présence des types de régions associées ou sur la rouge pour en spécifier l'absence. L'ensemble des catégories dont la case verte a été cochée définit la liste des CRP $\{pq_1, pq_2, ..., pq_M\}$ et les cases rouges cochées définissent la liste $\{nq_1, nq_2, ..., nq_R\}$ des CRN, telles qu'elles ont été introduites en section 5.4.1. La valeur du rayon de recherche γ est sélectionnée dans le menu déroulant en bas de la fenêtre.

Le contenu intégral de chaque catégorie peut être consulté en cliquant sur la vignette de sa région représentative. La figure 5.7 illustre, par exemple, les régions contenues dans deux catégories de la base Corel. Cette visualisation permet à l'utilisateur de vérifier si cette catégorie correspond effectivement aux types de régions qu'il recherche.

Sélectionné dans l'interface de requête, le rayon de recherche γ définit l'étendue de range query. Il permet d'intégrer dans la recherche un ensemble variable de catégories voisines "autour" de chaque catégorie requête. Pour la description par

couleur moyenne LUV des régions, le rayon sélectionné définit la valeur de distance euclidienne maximum entre les prototypes des catégories. Sachant que la distance minimum entre deux prototypes a été fixée à 2 à l'issue de la catégorisation des régions, une valeur par défaut $\gamma = 5$ a été jugée convenable pour intégrer par défaut des catégories voisines "proches". Cette valeur peut être modifiée par l'utilisateur pour ajuster la similarité des régions retrouvées.

La formulation de la requête est plus rapide et plus intuitive que dans les paradigmes de recherche par l'exemple, car on accède directement aux groupes de régions similaires présentes dans la base par l'intermédiaire du thesaurus.

Interface des résultats

A partir de la liste de CRP, de CRN et du rayon de recherche γ, le système exprime et affiche la formulation logique de la requête telle qu'elle est traitée. Nous renvoyons le lecteur au tableau 5.1 pour les fonctionnalités de recherche correspondant à chaque opérateur logique. Un exemple de formulation logique est illustrée dans l'interface des résultats (figure 5.11) avec les régions représentatives de chaque catégorie requête et de leurs catégories voisines respectives. Elles sont séparées par les opérateurs logiques suivants : OR entre les catégories voisines, AND entre les catégories requête, et $ANDNOT$ pour les catégories requête négative.

Dans l'interface (figure 5.12) sont affichées les images retrouvées par cette requête logique, correspondant à la formule (5.4). Notons qu'il n'y a pas d'ordre dans les images retrouvées.

Au vu des images retrouvées, l'utilisateur peut choisir de raffiner sa requête à partir de l'interface de résultats. Il peut augmenter ou réduire le rayon de recherche, supprimer ou ajouter des catégories requête afin de raffiner la caractérisation visuelle qu'il recherche. Notons qu'il est possible qu'aucune image ne soit retournée si la requête est très complexe donc très contrainte et plus généralement si la composition définie par l'utilisateur est trop spécifique par rapport à la base d'images. Dans ce cas, les contraintes sur la requête doivent être relâchées en spécifiant un rayon de recherche plus large ou en supprimant des catégories requête.

Des exemples de requêtes sur les bases Corel et TF1 seront présentées.

5.5.2 Résultats

Les tests ont été effectués sur un PC à 498 MHz avec 192 Mo de mémoire vive. La partie interface est en HTML et communique via la librairie CGI avec le programme écrit en C++. Sur la base Corel comportant 9.995 images et 50.220

régions, le processus de recherche prend au maximum 0.03 seconde sur un PC à 498MHz.

La pertinence des résulats de la recherche par composition de catégories de régions ne peut être évaluée que sur la satisfaction de l'utilisateur car pour une même recherche les scénarios de requête peuvent être multiples. L'évaluation dépend des facteurs suivants : de la connaissance que l'utilisateur a de la base, de son jugement personnel dans les images retrouvées, mais aussi du rayon de recherche, de l'ensemble de catégories requête qu'il a sélectionnés et de l'utilisation ou non du raffinement de requête.

La pertinence des régions mises en correspondance dans les images retournées repose sur les processus d'extraction et de regroupement de régions. Dans les images retournées, les régions correspondant aux catégories requête positives sont visuellement saillantes. Les faux positifs de régions parmi les régions mises en correspondance sont peu nombreux. Ils correspondent à des cas de segmentation difficile sur des images composites de scènes naturelles. Dans ce cas, une région mise en correspondance peut ne pas être significative même si sa couleur moyenne correspond à une catégorie requête.

Concernant la mise en correspondance de composition dans les images retournées, la simplicité des stratégies d'indexation et de recherche (comparaison des étiquettes de catégories dans les index d'images) assure la satisfaction attendue, car il s'agit d'une mise en correspondance *exacte* pour une formulation de requête donnée. Dans les images retournées, les régions satisfont bien les contraintes de présence des régions issues des CRP et d'absence des régions provenant des CRN. Le raffinement de la requête est un moyen supplémentaire de satisfaction de l'utilisateur (voir section 5.5.5).

Avec les deux scénarios présentés, nous allons pouvoir approfondir l'étude de la pertinence des résultats.

5.5.3 Application à une photothèque

Le premier scénario de recherche est testé sur 9.995 images de la photothèque Corel[3]. Elle comporte des scènes de nature très différente : paysages, portraits, objets, dessins, peintures, architecture, jardins, animaux, fleurs,...

50.220 régions sont extraites de la base par la méthode de segmentation présentée au chapitre 3. Le regroupement des 50.220 couleurs moyennes des régions prend 150 secondes avec l'algorithme CA. 91 catégories sont automatiquement générées. La disposition des points dans l'espace Luv ainsi que celles des 91 prototypes ont été montrées précédemment dans la figure 5.2. Les figures 5.7 et 5.8 illustrent le contenu de trois des 91 catégories obtenues.

[3]http ://www.corel.com

La granularité fine de classification de CA produit des catégories de régions homogènes en couleurs moyennes. Au sein de chaque catégorie, la variation photométrique est due aux différentes textures de régions qui présentent une couleur moyenne proche.

Le thesaurus photométrique de régions, constitué pour cette base des 91 régions représentatives, apparaît dans l'interface de requête (fig. 5.9). L'affichage des régions représentatives dans l'interface suit l'ordre des populations croissantes de catégories. La catégorie 0, la moins peuplée, possède 112 régions et la 90, la plus peuplée en comporte 2048. On remarque ainsi que les catégories les moins peuplées (premiers numéros de catégorie) sont celles correspondant principalement à des couleurs moyennes saturées et plus on avance dans les catégories peuplées moins elles sont saturées (noir, blanc, gris). Cette remarque est cohérente avec l'observation précédente qu'une grande majorité des couleurs moyennes dans l'espace Luv se situait dans le voisinage de l'axe des intensités, correspondant donc à des saturations faibles. La conséquence est que les catégories correspondant à des couleurs moyennes relativement saturées sont plus discriminantes pour la recherche.

Dans le thesaurus de régions, il est normal que certaines régions représentatives paraissent très proches visuellement entre elles. Ceci est dû à la contrainte de granularité fine de classification nécessaire à la formation de catégories *homogènes* de régions. Pour une paire de régions représentatives très proches visuellement, la sélection de l'une ou l'autre pour formuler la requête modifie peu les images retrouvées même pour un rayon de recherche faible. Cependant, par souci de clarté, nous envisagerons dans le futur de regrouper dans l'interface de requête les régions représentatives proches .

Dans un contexte d'agence photo, l'utilisateur peut souhaiter par exemple rechercher des paysages urbains. Ce type de scènes peut être caractérisé visuellement par la présence de zone de ciel et de bâtiment, et par l'absence de verdure. La figure 5.10 illustre la traduction de cette recherche par la composition visuelle suivante : "région grise et région bleue et pas de région verte". Pour un rayon de recherche donné, le système détermine les voisines de chaque catégorie requête et réduit la requête en une composition logique (voir figure 5.11). La figure 5.12 montre l'ensemble des images retournées pour cette requête. Dans ces images, les régions grises correspondent effectivement majoritairement à des immeubles, mais aussi à des monuments, et à quelques rochers, les régions bleues à des zones de ciel. La deuxième partie de l'interface de résultats (figure 5.13) montre l'ensemble des images rejetées à cause de la présence d'une région verte, en plus d'une région grise et d'une région bleue. Ces images s'avèrent correspondre presque toutes à des paysages de nature et ont été éliminées à juste titre.

Au sein des catégories sélectionnées pour une requête donnée, de nombreuses régions ne sont généralement pas pertinentes d'un point de vue sémantique. A

l'observation du contenu des images retournées, il est intéressant de remarquer que ces régions non-pertinentes sémantiquement ne sont généralement pas retenues alors qu'elles sont cohérentes photométriquement. La contrainte de composition Dans le scénario de la recherche de paysages urbains, ces régions non-pertinentes sémantiquement sont les régions bleues qui ne sont pas du ciel, des grises qui n'identifient pas des bâtiments et des vertes ne provenant pas de zones de nature. La requête logique impose une contrainte de composition de différents types de régions au sein d'une même image. Cette contrainte semble plus probablement vérifiée par des régions sémantiquement pertinentes pour la recherche de l'utilisateur. Dans l'exemple de recherche de paysages urbains, la catégorie grise sélectionnée contient en fait une faible proportion de bâtiments mais qui a été retrouvée par la contrainte de présence dans une même image avec une région bleue et sans région verte (voir le diagramme d'illustration 5.20). La requête logique par composition semble introduire naturellement une forme de *sémantique visuelle* dans la recherche d'images.

FIG. 5.7 – Exemples de 2 catégories de région dans la base Corel : la catégorie 23 (en haut) regroupe les régions ayant une couleur moyenne orange proche et la catégorie 48 (en bas) correspond à une couleur moyenne verte. Chaque catégorie est homogène par rapport au descripteur utilisé lors de la catégorisation, à savoir la couleur moyenne ici.

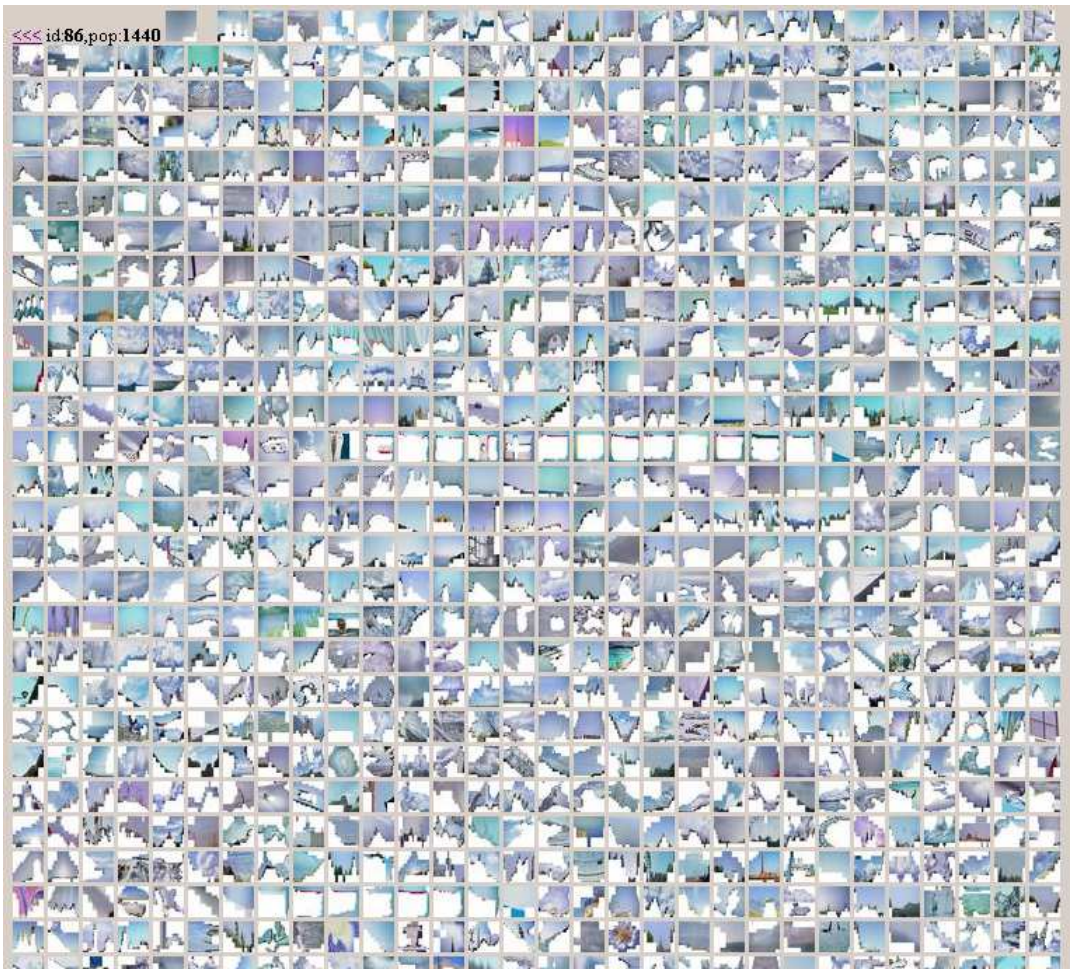


FIG. 5.8 – Un autre exemple de catégorie de région de la base Corel : la catégorie 86 contient les régions correspondant à une couleur moyenne bleu ciel.

FIG. 5.9 – Interface de requête : les régions représentatives des 91 catégories constituent le "thesaurus photométrique de régions" de la base Corel. Chaque catégorie peut être sélectionnée pour former la requête. Aucune image exemple ni région exemple n'est nécessaire ici. Le contenu de chaque catégorie peut être consulté en cliquant sur sa région représentative.

FIG. 5.10 – Exemple de recherche de paysages urbains : le rayon de recherche est fixé à sa valeur par défaut, les catégories 39 (bleu) et 88 (gris) sont les CRP la 48 (vert) la CRN.

FIG. 5.11 – Expression complète de la composition logique de catégories de régions : elle est formulée par le système à partir de la requête utilisateur (figure 5.10).

FIG. 5.12 – Résultats : les images retournées satisfont la requête par composition. Notons que la contrainte de présence de régions vertes et des régions bleues et l'absence de région verte est principalement satisfaite par des paysages urbains et aussi par des images des monuments et de ruines. Les images ne sont pas ordonnées.

FIG. 5.13 – Les images rejetées lors de la recherche de paysages urbains à cause de la présence d'une région verte (en plus d'une région bleue et d'une région grise). Elles s'avèrent représenter des paysages naturels.

5.5.4 Application au journal télévisé

Le deuxième scénario s'applique à la recherche dans les bases de video de journaux télévisés. La base est constituée des images clés extraites d'un journal télévisé de TF1 : 910 images ont été extraites de 3 minutes de videos. De cette base, 6362 régions ont été extraites par segmentation et 65 catégories ont été obtenues par catégorisation. Par rapport à la base Corel, le thesaurus photométrique de régions de la base TF1 présente moins de catégories dans l'ensemble (la base est plus petite) et en particulier moins de catégories correspondant à des couleurs saturées. Cependant le thesaurus comporte un nombre remarquable de catégories de couleurs bleues et noires caractéristiques de la charte graphique du journal télévisé de TF1. Deux catégories sont illustrées à la figure 5.14 qui correspondent à des couleurs moyennes bleues et couleur chair.

L'observation du contenu des différentes catégories montre qu'elles contiennent des sous-groupes de parties sémantiquement identifiables telles que :

- dans la catégorie verte, de la végétation
- dans la catégorie noire, des moitiés de costumes (provenant du présentateur du journal et de personnes interviewées) et des parties sombres du décor de plateau
- dans la catégorie "chair", des visages
- dans les catégories correspondant à différentes nuances de bleu saturé, différentes parties d'incrustation

Les trois requêtes que nous allons présenter sur la base TF1 correspondent à des problèmes pratiques posés par les documentalistes de cette chaîne : la détection des scènes de plateau dans les videos, les images comportant des incrustations et finalement des images ayant pour fond une façade de bâtiment institutionnel.

Scènes de plateau

La recherche de scènes de plateau est effectuée par la sélection comme CRP des catégories 57, 25 et 55 (voir figure 5.14). La catégorie 55 a été sélectionnée pour cette requête car elle contient des visages et les 25 et 57 car elles représentent des parties caractéristiques du décor du plateau télévisé de TF1. La formulation logique de la requête ainsi que les résultats sont présentés dans la figure 5.16. Les images retournées correspondent toutes à des images de plateau.

Nous avons aussi effectué des requêtes à partir de chacune de ces trois catégories individuellement. Les images retournées incluaient les scènes de plateau recherchées mais aussi de nombreuses images non pertinentes. Cela signifie que chaque catégorie ne suffit pas individuellement à caractériser une image de plateau, mais ce n'est qu'avec la conjonction des trois catégories que nous avons obtenu exclusivement les images de plateau. Et, inversement, dans

chaque image retournée, les régions mises en correspondance correspondent sémantiquement aux zones d'intérêt (élément de décor et visage) attendues par l'utilisateur. Cette remarque rejoint celle de la recherche de paysages urbains dans la base Corel. Ces exemples illustrent la sémantique visuelle obtenus par la contrainte de composition.

Incrustations
Pour la recherche d'incrustations dans les images, nous avons sélectionné comme CRP la catégorie 34 qui illustre un bleu clair saturé propre aux incrustations graphiques de TF1. Les images retournées correspondent effectivement à des incrustations (voir figure 5.17). Si la requête porte en particulier sur les incrustations de portraits, elle peut être raffinée en imposant l'absence de rouge qui caractérise les images non-pertinentes (les diagrammes en camembert). La figure 5.18 illustre la nouvelle requête avec pour CRN supplémentaire la catégorie rouge 41. Les images retournées sont alors réduites aux incrustations de portraits seulement. Il s'agit d'un exemple de raffinement de requête.

Façades officielles
Un autre type de requête a fait l'objet de l'intérêt des documentalistes de TF1 : les façades d'institutions de bâtiments officiels. Ce type de requête portant sur le *fond* d'une image leur permet de retrouver des sujets dans lesquels une personnalité politique a été interviewée. Dans la figure 5.19, la sélection d'une catégorie de couleur gris-ocre et une de vert permet de retrouver une série d'images dont le fond est une façade de bâtiment du type recherché et contenant un arbre. Notons que ce type classique de recherche transversale permet de retrouver un sujet principal de façon détournée.

La constitution d'une base plus importante d'images extraites de journaux télévisés de TF1 est en cours de réalisation. Elle permettra d'étudier dans quelle mesure se généralise l'association entre les catégories et des éléments distinctifs de ces requêtes spécifiques (plateau, visage, charte graphique).

Notons que les trois requêtes illustrées exploitent une certaine connaissance de la spécificité visuelle des images traitées : celles des décors de plateau, des incrustations, des façades de bâtiments. Les utilisateurs dans un tel scénario étant des documentalistes des archives video de la chaîne de télévision, l'hypothèse de cette connaissance est raisonnable. La préstructuration de la base en régions étant totalement non-supervisée, notre approche peut être transposée à tout autre domaine, et la connaissance de certaines spécificités du domaine peut être intégrée dans la formulation des requêtes.

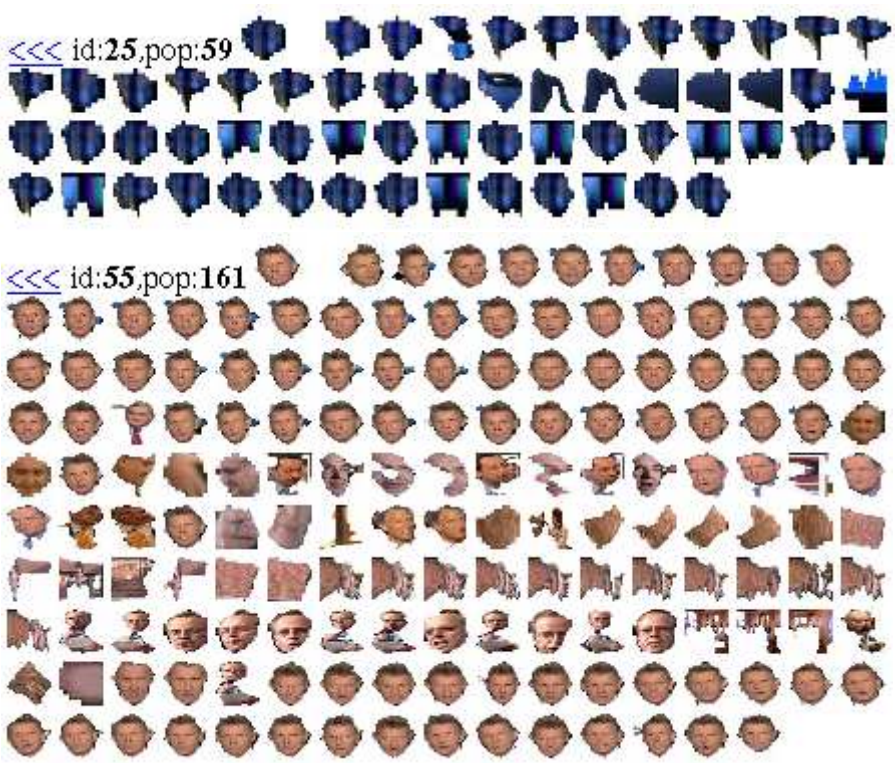


FIG. 5.14 – Exemple de deux catégories de régions de la base TF1 : la catégorie 25 (haut) correspond à une couleur moyenne bleu foncé et contient principalement des décors des scènes de plateau. La catégorie 55 (bas) correspond à une couleur chair et comporte principalement des visages.

FIG. 5.15 – Interface de requête : les 65 catégories constituent le “thesaurus photométrique de régions” de la base TF1.

FIG. 5.16 – Expression de la requête logique et résultats correspondants pour retrouver les images de plateau. Trois catégories ont été sélectionnées : deux catégories correspondant à deux bleus foncé contiennent des éléments typiques de décor. La troisième catégorie correspondant à la couleur chair contient principalement des visages. Il est intéressant de noter que les régions de cette catégorie qui ne sont pas pertinentes (i.e. qui ne sont pas de visages) sont automatiquement rejetées.

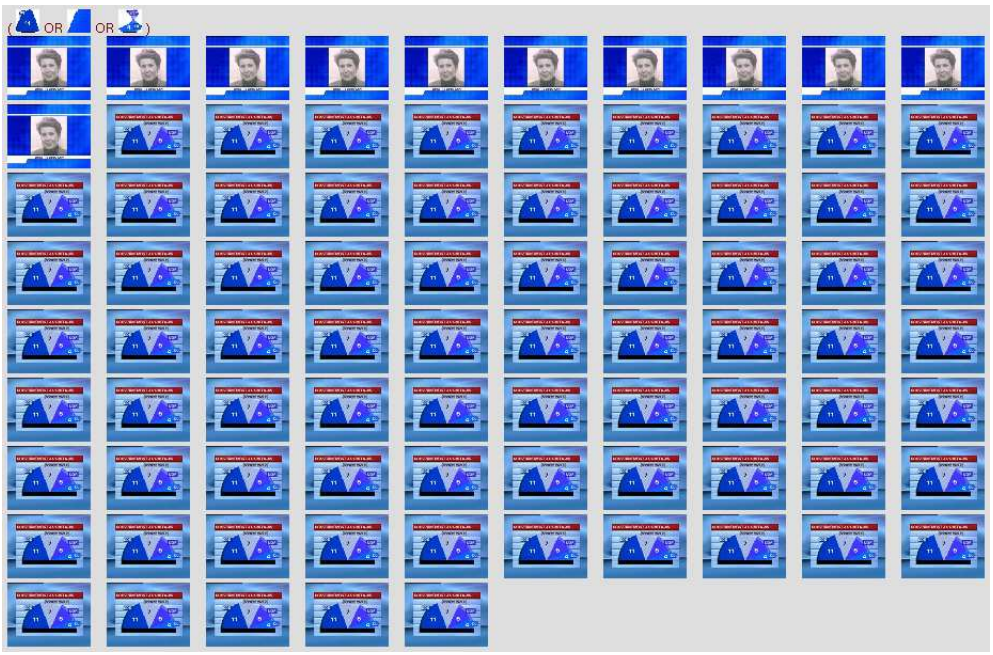


FIG. 5.17 – Expression de la requête logique et résultats correspondants pour retrouver les incrustations graphiques. La requête est réduite à une catégorie de bleu vif et ses deux voisines. Ce bleu est spécifique aux incrustations.

FIG. 5.18 – Raffinement de la requête pour rejeter les diagrammes en ajoutant une catégorie rouge comme CRN.

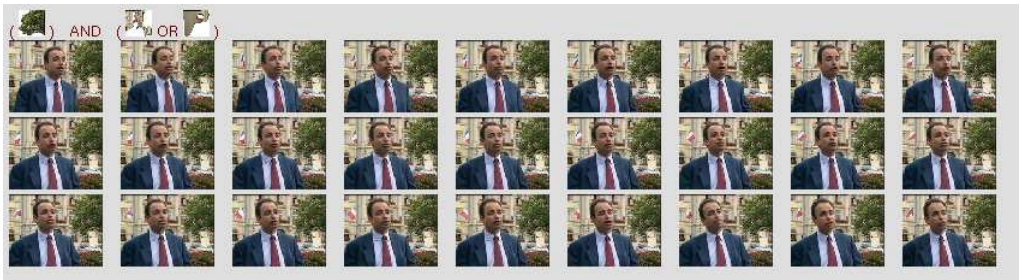


FIG. 5.19 – Pour retrouver cette scène spécifique d'interview, c'est le fond, constitué d'une façade de bâtiment institutionnel et d'un arbre, qui est recherché par la composition de deux catégories.

5.5.5 Raffinement de requête - interaction sur les résultats

Le système permet à l'utilisateur d'interagir avec les résultats retournés pour modifier la formulation de sa requête et obtenir un nouvel ensemble de résultats qui corresponde mieux à ses attentes. Il peut, d'une part, utiliser la fonctionnalité de "range-query" pour adapter l'homogénéité des types de régions qu'il recherche et, d'autre part, modifier la requête logique. Le premier cas correspond à un raffinement au niveau *région* et le second au niveau *composition de scène*.

Pour la fonctionnalité de "range-query", la valeur de rayon de recherche γ proposée par défaut permet d'intégrer dans la requête les catégories voisines proches de chaque catégorie requête CRP et CRN. Elle définit ainsi l'étendue de la similarité visuelle pour chaque type de régions. Il est important que cette information soit ajustable *dynamiquement*, car elle dépend de la nature *sémantique* des zones auxquelles pense l'utilisateur. Pour une recherche de scènes d'extérieur, par exemple, il pourra spécifier une zone de bleu pour le ciel. S'il pense à une scène ensoleillée, il pourra limiter la requête à un bleu saturé grâce à un rayon faible. A l'inverse si la recherche porte sur des scènes d'extérieur plus générales, le rayon devra être augmenté afin de tenir compte d'une plus grande variété de cieux. Le nombre d'images retournées est croissant avec la valeur du rayon. A l'extrême, un rayon trop élevé risque de définir comme similaires des régions qui n'ont plus de lien perceptuel entre elles. Le choix d'un rayon convenable peut être établi interactivement sur les résultats. Selon les régions mises en correspondance dans la composition des images retrouvées, l'utilisateur juge de la similarité visuelle des régions retrouvées. S'il estime qu'une plus grande tolérance sur la similarité est nécessaire, il augmentera le rayon de recherche et il le diminuera s'il souhaite plus contraindre la similarité aux catégories requête. Le cas échéant, la modification du rayon et l'envoi de la requête modifiée se font directement dans l'interface des résultats.

La seconde possibilité de raffinement se situe au niveau de la composition des scènes. A l'observation des premiers résultats, l'utilisateur peut se rendre compte si les types de régions requête qu'il a indiqués sont effectivement caractéristiques ou non pour les scènes qu'il recherche. Les images retournées peuvent aussi lui suggérer d'autres types de régions caractéristiques. L'association entre les images retournées pour la requête formulée et l'image mentale de l'utilisateur peut lui suggérer d'ajouter ou de supprimer certains types de régions. La requête peut alors être directement modifiée dans l'interface des résultats.

En fonction des images retournées par la première requête, l'utilisateur peut donc raffiner à la fois l'étendue de la similarité des types de régions et la composition logique. Ces deux modes d'interaction sur les résultats permettent d'améliorer la sémantique des images retournées grâce à un "dialogue" entre le système et l'utilisateur.

5.5.6 Coût de la recherche

Sur la base Corel comportant 9.995 images et 50.220 régions, le processus de recherche prend au maximum 0.03 seconde pour les requêtes les plus complexes sur le PC à 498MHz. Des requêtes sont dites "complexes" lorsque le rayon de recherche choisi est élevé (plus de catégories voisines sont prises en compte) et lorsque le nombre de CRP et CRN sélectionnées est élevé. Ce temps est particulièrement faible étant donnée la combinatoire très élevée d'une requête logique sur plusieurs types de régions dans une base de cette taille.

Il est important de remarquer que le temps de recherche restera faible *quelque soit la dimension des descripteurs* de régions envisagés (couleur moyenne, texture, distributions de couleur). De façon plus générale dans notre approche, le choix du descripteur et de sa distance associée ne concerne que le processus de regroupement des régions qui est effectué "off-line". Il n'influe donc pas sur le temps de recherche. En effet, la seule information exploitée au moment de la recherche sont les distances précalculées (pour la détermination des catégories voisines) et l'association entre les catégories et les images. Seules des opérations sur des entiers sont impliquées.

Les facteurs suivants sont à l'origine de la rapidité du processus de recherche :

- la recherche repose sur le seul accès à 3 tables d'association (contrairement aux approches classiques, il n'y a pas de descripteur numérique, ni de distance à calculer pour la recherche)
- les régions ne sont pas recherchées individuellement mais seulement par catégories. Par exemple, dans le cas du scénario Corel nous n'utilisons que les 91 catégories de régions au lieu des 50.220 régions
- il n'accède pas à l'intégralité des images de la base

La représentation de la structure de la base a été conçue dans le but de fournir les accès directs aux associations (entre images et catégories) utiles au processus de recherche. Toute recherche séquentielle sur l'ensemble de la base est ainsi évitée. En moyenne sur les requêtes par composition, la fraction d'images à laquelle on accède de situe autour de 12% pour chaque requête. Elle est calculée comme le rapport entre le nombre d'images pour l'initialisation de S_{res} et le nombre total d'images de la base. Avec les notations de l'expression (5.5), ce rapport est défini comme : $\mid S_{PQ_1} \mid / \mid S_{DB} \mid$, où S_{DB} est l'ensemble des images de la base.

5.6 Discussions

Analogie avec la Recherche de Texte

Les approches existantes de recherche d'images par le contenu représentent les images (ou les régions d'images) comme des points multidimensionnels dans

l'espace de description visuelle de façon individuelle. Dans notre approche, le regroupement des régions similaires en catégories conduit à une indexation *symbolique* plutôt que numérique du contenu des images. Les requêtes exploitent les occurrences des mêmes types de régions dans les images. Outre l'avantage considérable de gain en rapidité de recherche, la représentation symbolique combinée à l'usage d'opérateurs logiques de requête (`ET`, `OU`, `ET NON`) rend cette approche semblable à celle de la recherche de texte, selon la correspondance terminologique suivante :

- image → document
- région → terme
- catégorie de régions → concept
- catégories voisines → concepts similaires/synonymes
- ensemble de catégories de régions → thesaurus
- requête par composition logique → requête de type `Google`[4]

Différentes techniques éprouvées dans la recherche de texte [96] pourraient ainsi être appliquées à notre approche. Wang et Du [115] ont par exemple proposé l'extension modèle *tf.idf* à la recherche d'images par régions afin de favoriser les régions peu fréquentes.

De la composition vers la sémantique

Nous avons observé dans les exemples des scénarios de recherche dans la photothèque et les images de journal télévisé que la composition de types de régions dans les images faisait émerger une forme de *sémantique visuelle*. Une catégorie requête que sélectionne l'utilisateur afin d'identifier un type d'objet contient un ensemble de régions visuellement similaires, mais pas toutes pertinentes sémantiquement pour la requête. Nous avons pu observer que la contrainte de présence et d'absence de plusieurs types de régions avait tendance à éliminer naturellement les régions sémantiquement non-pertinentes dans leurs catégories respectives. La recherche par composition permet de désambiguïser la sémantique visuelle des différents types de régions requête. Ce principe est illustré par la figure 5.20.

Ce type de désambiguisation est couramment utilisé pour la raffiner la recherche de texte. Dans un moteur de recherche sur internet, pour chercher des documents relatifs aux "noeuds" au sens de mesure de vitesse nautique, une requête par le mot-clé "noeuds" retournera probablement beaucoup de documents comportant le mot "noeud", mais dans d'autres contextes : noeuds d'une corde, noeuds d'un graphe, ... Si maintenant, au lieu de simplement rechercher le mot "noeud", nous recherchons "noeud ET vitesse", nous augmentons nos chances de retrouver des documents relatifs à la vitesse nautique et de rejeter les documents

[4]http ://www.google.com

non-pertinents pour la recherche.

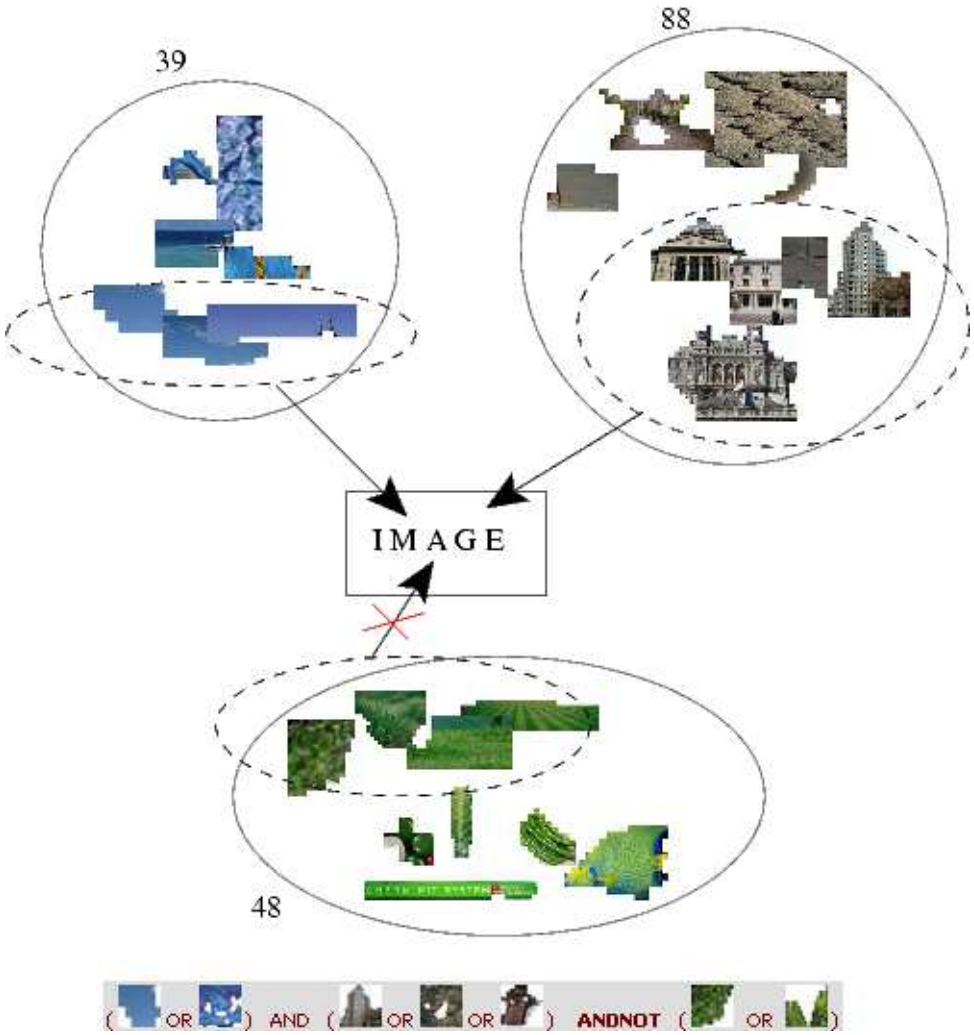


FIG. 5.20 – Illustration de la sémantique visuelle induite par la requête par composition logique de catégories de régions : bien que les régions photométriquement similaires au sein d'une même catégorie puissent avoir une sémantique différente, la contrainte d'absence et de présence de régions à l'intérieur d'une *même image* tend à éliminer les régions sémantiquement non-pertinentes même si visuellement similaires. Dans le schéma, les catégories sont illustrées par des traits pleins. Les pointillés délimitent les sous-ensembles de régions retenus pour la composition.

Positionnement de notre approche

L'expression "thesaurus visuel" a été précédemment introduite dans la littérature (voir [86], [62] et [112]). Cependant la conception et la finalité de leurs méthodes diffèrent des nôtres.

Picard a proposé de constituer un "thesaurus visuel" [86] à partir du système FourEyes [75]. L'utilisateur est assisté dans les associations et les regroupements entre entités visuelles intra- ou inter-images par un bouclage supervisé avec le système. Le thesaurus constitué est l'ensemble de ces regroupements et associations pour un type d'application donnée.

Dans [112], Town et Sinclair se sont intéressés à la classification sémantique de régions d'images parmi un ensemble prédéfini de 11 classes sémantiques (brique, nuage, fourrure...). A l'issue d'une phase d'apprentissage, un réseau de neurones associe au descripteur visuel d'une région un score d'appartenance à l'une des 11 classes. Dans l'interface de requête, un thesaurus visuel de régions sur les classes apprises peut être utilisé pour formuler des requêtes sur plusieurs régions.

Dans un but de navigation et de recherche de motifs de texture dans une base d'images aériennes, Ma et Manjunath [62] constituent un thesaurus de blocs de texture. La base des blocs et de régions extraites des images est partitionnée hiérarchiquement en classes selon deux niveaux. Le regroupement comporte une phase supervisée afin de prendre en compte la similarité perceptuelle propre au domaine d'application. Les motivations de cette approche sont la navigation et la recherche de motifs similaires à un motif donné.

Bien que différentes les unes des autres, ces trois approches conçoivent un "thesaurus visuel" dans un cadre *supervisé* d'apprentissage de primitives visuelles. L'apprentissage est employé dans la phase de construction des thesauri afin de modéliser la sémantique [112] ou la perception visuelle humaine sur les classes apprises [62][86]. A l'inverse, notre approche est totalement non-supervisée et donc transposable a priori à tout type de domaine d'application et de descripteurs mis en oeuvre. Au niveau de la construction de notre thesaurus photométrique de régions, nous ne cherchons pas à former des catégories sémantiques ou perceptuellement idéales, mais des catégories de régions cohérentes visuellement. C'est plutôt dans l'interaction avec l'utilisateur, concentrée sur l'étape de recherche, que la sémantique visuelle émerge au niveau des requêtes logiques par composition (voir section 5.6).

5.7 Perspectives

Simple et originale, l'approche proposée dans ce chapitre suggère de nombreuses perspectives dans le domaine de la recherche d'information visuelle.

Concernant le paradigme 2 de recherche d'images par composition logique de catégories de régions, il peut bénéficier des avancements de divers domaines de recherche tels que la recherche de texte, la description visuelle, la classification non-supervisée, l'indexation spatiale, l'optimisation de requêtes, la navigation et la visualisation dans les bases de données. Nous évoquerons aussi plus généralement des applications en recherche d'information visuelle par le contenu qui dépassent le cadre de la recherche d'images par régions.

5.7.1 Description de régions

L'approche a été présentée avec la couleur moyenne comme description visuelle de régions. Bien que très simple, ce descripteur a montré sa capacité à retrouver des images par composition de catégories de régions. L'intégration de descripteurs géométriques et plus riches photométriquement (voir ceux employés dans le paradigme 1) pourrait être envisagée pour raffiner la similarité visuelle des régions retournées.

Choix du descripteur

Le choix du descripteur employé (couleur, géométrie, texture, forme) peut être guidé par les besoins du domaine d'application. A priori, tout type de descripteur visuel associé à une métrique (qui est nécessaire à la formation des catégories de régions similaires) peut convenir. Les catégories seront alors homogènes selon le descripteur employé et, pour une même base, deux choix de descripteurs différents produiront deux catégorisations de régions différentes. Selon le type de scènes et d'"objets" recherchés, différents descripteurs peuvent s'avérer nécessaires (comme la position en plus de la couleur pour distinguer par exemple une zone de ciel d'une zone de piscine).

La question se pose de la façon dont ils peuvent être combinés dans notre approche. Deux types d'intégration multi-descripteurs sont envisageables. La première consiste à procéder à autant de catégorisations de régions qu'il y a de combinaisons possibles de descripteurs. L'utilisateur doit pouvoir basculer d'une catégorisation à l'autre selon sa requête. La seconde serait de produire une catégorisation *hiérarchique* des régions dans laquelle chaque niveau correspond à un descripteur. La sélection d'un type de régions selon plusieurs critères visuels s'effectuerait par descente dans l'arbre.

Catégorisation de régions

Dans les scénarios de recherche présentés, nous avons observé une importante compacité des données dans l'espace de description des couleurs moyennes. Dans

ces conditions, il est difficile d'espérer identifier des regroupements naturels de données. Malgré ces observations, l'approche de nous avons présentée permet de former des catégories cohérentes de régions similaires.

Cependant, rien ne nous permet de généraliser cette observation (de compacité des données dans l'espace de description) à d'autres descripteurs et d'autres types de bases. Il serait donc utile d'approfondir l'étude de la répartition des données dans un cas plus général de descripteurs et de domaines d'application en se posant les questions suivantes : existe-t-il un regroupement naturel des données ? si oui, en combien de groupes ? des quelles populations ? de quelles densités ? de quelles formes ? La méthode de catégorisation devrait permettre de modéliser au mieux ces groupements naturels afin de produire des catégories visuellement adaptées.

5.7.2 Recherche dans de très grandes bases

De par sa structure intrinsèque, notre système convient naturellement à la recherche dans de grandes bases d'images. Les similarités entre les régions sont précalculées. Le temps de recherche est totalement indépendant du descripteur adopté. Nous nous interrogeons ici sur le comportement de notre système sur de très grandes bases et proposons des ébauches de solutions.

La plus grosse base sur laquelle nos tests ont porté comportait 9.995 images et 50.220 régions. Les temps de recherche obtenus sur un PC à 498MHz sont inférieurs ou égaux à 0.03 seconde pour les requêtes logiques les plus complexes. Les facteurs de rapidité de la recherche ont été détaillés en section 5.5.6. Sachant que les PC les plus rapides actuellement sont cadencés à 3GHz et que l'on considère usuellement que le temps raisonnable d'attente maximum pour l'attente des résultats est de 2 secondes, la marge de calcul supplémentaire permet d'envisager des implantations sur des bases d'images bien plus importantes qui dépassent le million d'images.

Outre l'aspect de temps de calcul, l'augmentation du nombre d'entrées de la base nécessitera de se pencher en particulier sur la granularité de la catégorisation des régions et sur le nombre d'images résultat, comme nous allons le voir dans le paragraphe suivant.

Granularité de la catégorisation

Le regroupement des régions en catégories a un usage double dans notre approche. Le premier est celui de préstructuration et d'indexation de la base (les catégories définissent la similarité entre les régions) et le second est celui de visualisation (les régions représentatives des catégories constituent le thesaurus exploité par l'utilisateur). A chaque usage correspond une contrainte sur la granularité de

la catégorisation. Pour l'aspect de préstructuration et d'indexation, les catégories doivent contenir des régions homogènes (donc besoin d'une granularité fine). Cette considération favorise un nombre *élevé* de catégories. Quant à la visualisation, le nombre de régions représentatives, donc de catégories, doit rester relativement *faible* pour que l'utilisateur puisse rapidement sélectionner les catégories pertinentes dans l'interface de requête (donc besoin d'une granularité faible). Ainsi le choix de granularité de la catégorisation doit alors être adapté à la taille de la base afin de prendre en compte des contraintes antagonistes. Dans nos tests sur la base Corel, la représentation des 50.220 régions par 91 catégories nous a semblé satisfaisante aussi bien pour l'aspect de préstructuration et d'indexation que de visualisation.

Avec l'augmentation du nombre d'images et de régions dans la base, la catégorisation peut produire soit trop de catégories (problème de la "lisibilité" de l'interface), soit des catégories comportant trop de régions (risque de regrouper des régions peu similaires donc d'obtenir des résultats peu pertinents). Si un choix de granularité de catégorisation ne peut satisfaire aux deux types de contraintes pour une très grande base, une solution naturelle serait de proposer une catégorisation *hiérarchique* des régions. Cette hiérarchie permettrait à la fois d'avoir des catégories homogènes, tout en présentant à l'utilisateur un nombre raisonnable de régions représentatives à chaque niveau de la hiérarchie. L'utilisateur pourrait naviguer dans les différents niveaux de la hiérarchie afin de trouver la catégorie correspondant au type de régions qu'il recherche.

Les niveaux de la hiérarchie pourraient correspondre à des catégorisations de plus en plus fines selon le même descripteur de région, ou bien à des catégorisations selon différents descripteurs (voir la section 5.7.1).

Structuration des résultats

A requête logique équivalente, une grande base fournit plus d'images résultat qu'une petite base. Il peut facilement arriver qu'une requête retourne des centaines d'images (par exemple si la requête ne porte que sur une seule catégorie), voire quelques milliers sur une très grande base. Dans ce cas, la présentation désordonnée d'un très grand nombre d'images ne permet plus à l'utilisateur d'appréhender les résultats efficacement. Il serait alors nécessaire d'étudier la structuration de ces résultats soit en les triant à partir d'une image exemple sélectionnée par l'utilisateur parmi les images résultat, soit sous forme de regroupements homogènes selon différents critères (autres descripteurs par exemple).

5.7.3 Interaction système-utilisateur

Sélection des catégories voisines

Au moment de la formulation de la requête, l'ensemble des catégories voisines des différentes catégories requête est donné par le rayon de recherche γ sélectionné par l'utilisateur. Plus la valeur sélectionnée de γ est élevée, plus de catégories voisines seront prises en compte et inversement si γ est faible. Cette interaction est cruciale puisque c'est d'elle que dépend la fonctionnalité de range-query : l'utilisateur peut adapter l'étendue de la similarité des régions selon la nature des régions recherchées. L'implantation proposée de la sélection de ce rayon de recherche est rudimentaire et mériterait d'être améliorée sur deux points. Premièrement, le rayon devrait pouvoir être sélectionné individuellement pour chaque catégorie requête car chaque type de région au sein d'une même requête peut nécessiter une étendue de similarité différente. Deuxièmement, la sélection numérique (voir figure 5.10) de γ n'est pas intuitive pour l'utilisateur ; elle devrait plutôt être *perceptuelle*.

Amélioration du thesaurus

Dans l'interface de requête, le thesaurus photométrique de régions présente les régions représentatives par ordre de population croissante de catégorie. Pour l'utilisateur, il serait plus naturel de proposer un arrangement spatial de ces représentants selon leur similarité visuelle. Deux régions représentatives proches dans l'espace bidimensionnel du thesaurus devraient pouvoir être proches perceptuellement. Une telle organisation permettrait d'appréhender plus aisément l'ensemble des catégories de la base. Une solution adéquate consisterait à utiliser les cartes auto-organisatrices de Kohonen [52] (ou SOM pour *Self-Organising Maps*) qui permettent la conservation de la topologie d'un ensemble multidimensionnel en deux dimensions. La transformation s'appliquerait aux descripteurs régions ayant servi à former les catégories et permettrait un arrangement bidimensionnel des représentants selon leur similarité. Un exemple d'application de ces cartes à des descripteurs visuels a été proposé dans PicSom [54] pour la visualisation d'une base d'images dans espace bidimensionnel.

Une seconde piste à étudier concerne le choix des régions représentatives de catégories. Les régions représentatives affichées dans l'interface de requête guident l'utilisateur dans le choix des catégories-requête. Leur pertinence visuelle participe donc à l'efficacité de la recherche.

Les catégories étant obtenues par classification des descripteurs, nous avons naturellement défini la région représentative d'une catégorie comme étant la ré-

gion la plus proche du prototype de la catégorie obtenu avec l'algorithme CA. Ce choix est le plus simple et le plus naturel, mais il serait intéressant d'étudier d'autres manières de représenter chaque catégorie dans l'interface afin que l'utilisateur en ait le meilleur aperçu possible. Les meilleures régions candidates devraient être significatives visuellement, c'est-à-dire favoriser celles qui sont issues d'une segmentation "facile" et qui sont, de plus, de forme suffisamment compacte pour être visibles.

Par ailleurs, il serait intéressant d'étudier si l'utilisation de *plusieurs* régions peut permettre de représenter une catégorie de façon plus pertinente pour l'utilisateur.

5.7.4 Perspectives en recherche d'information visuelle

Le paradigme 2 permet à l'utilisateur de désigner explicitement dans le thesaurus les composantes visuelles typiques des images qu'il recherche. La notion d'image représentée par ses composantes visuelles qui sont caractérisées par leurs descripteurs peut être étendue à la notion de "document" (entendu au sens large) représenté par un ensemble d'attributs numériques. Dans des domaines d'usage variés, le paradigme 2 permettrait la recherche de documents à partir d'un thesaurus représentant leurs attributs (ou composantes) typiques. Le thesaurus ne serait plus nécessairement exclusivement visuel, mais dépendrait de la nature des attributs numériques. Ce paradigme constituerait une alternative intéressante aux modes existants de recherche par mots-clés ou par l'exemple.

Concernant l'image, la catégorisation pourrait être effectuée selon différents attributs simultanément (descripteurs visuels globaux ou partiels ou metadonnées) et être exploitée pour des recherches multi-critères. Dans le contexte de la recherche de videos, nous pourrions envisager la catégorisation des scènes détectées [1] et proposer un thésaurus de scènes typiques sur une base entière. Ce paradigme offre des perspectives dans le cadre plus général de la recherche de documents multimedia composites.

De telles extensions nécessiteraient de réfléchir à la manière de combiner efficacement des catégorisations générées selon des attributs *de natures différentes*.

Par ailleurs nous pourrions étudier l'intégration de modèles de représentation d'image plus abstraits reposant sur les prédicats, les concepts associés aux régions et leurs relations (voir les modèles EMIR [70, 71], DISIMA [80]).

5.8 Conclusions

Nous venons de présenter le paradigme 2 de recherche par composition logique de catégories de régions qui permet de retrouver des images à partir de requêtes

telles que : “trouver les images composées de régions de ce type et de ce type, mais pas de ce type, ni de ce type”.

Radicalement différent du paradigme habituel de recherche par l'exemple, ce nouveau paradigme ne nécessite aucune image ou région exemple pour formuler une requête, l'*image mentale* recherchée suffit. Le thesaurus visuel fournit un aperçu direct de l'ensemble des régions qui composent les images de la base. Il permet de formuler des requêtes complexes sur la composition des images. L'interaction de l'utilisateur avec le système s'opère au niveau de ce thesaurus et, éventuellement, dans l'interface des résultats pour raffiner la recherche. Il devient possible de rechercher des images de la même façon que l'on recherche des documents de texte à partir d'une expression logique de mots-clés. L'utilisateur peut adapter la similarité visuelle des différents types de régions selon la nature de la requête.

Nous avons observé que l'expression de l'utilisateur dans la composition des scènes recherchées fait émerger une forme de *sémantique visuelle* au niveau de la composition logique des types de régions, même avec un descripteur élémentaire de région.

L'indexation symbolique des images et la recherche non-séquentielle sur les entrées de la base permettent des temps de recherche très rapides. Ce temps de recherche est indépendant de la nature du descripteur visuel employé. Au moment de la recherche, on ne représente plus les régions individuellement, mais seulement les catégories de régions similaires.

Nous avons proposé différentes pistes de travaux futurs concernant l'approche elle-même de recherche d'images par régions, mais aussi de l'étendre à d'autres problèmes de recherche d'information visuelle.

Chapitre 6

Perspectives

Concernant les perspectives propres aux différentes contributions (segmentation, paradigme 1 et paradigme 2), nous invitons le lecteur à se référer aux chapitres correspondants. Nous présentons dans ce chapitre des perspectives plus larges concernant la recherche d'images par régions d'intérêt.

Association de mots-clés aux régions

Dans l'introduction de cette thèse, nous avons commencé par évoquer l'annotation d'images par mots-clés comme première approche historiquement pour la recherche d'images. Plus précise que l'annotation d'images, l'annotation des régions détectées permettrait de décrire les images par leur composition en termes d'*objets sémantiques*. Des requêtes sophistiquées pourraient alors être formulées sur le contenu aussi bien en termes de composition visuelle que sémantique.

Par ailleurs, toutes les régions annotées par des mots-clés pourraient être regroupées en catégories sémantiques et intégrées au paradigme 2. Il serait alors possible de formuler des requêtes par composition logique à la fois de composantes photométriques et de composantes sémantiques.

L'inconvénient majeur de l'annotation demeure la pénibilité de la tâche d'annotation sur une base entière. Afin d'assister l'utilisateur, l'annotation des régions pourrait semi-automatisée en propageant les mots-clés de régions annotées vers les régions d'aspect photométrique similaire.

Relations spatiales

Dans la représentation partielle des images en régions, nous nous sommes intéressés à la description photométrique et géométrique ainsi qu'à la spécification de la présence et de l'absence de certains types de régions. L'information de relations spatiales entre les régions au sein d'une image devrait être intégrée à

notre système [25]. De nombreux travaux dans cette direction ont été développés pour les Systèmes d'Information Géographiques. La spécification des relations spatiales dans une requête, si le besoin se présente, permettrait d'améliorer encore la pertinence des images retournées par rapport aux attentes de l'utilisateur. Cet axe fait actuellement l'objet d'une étude au sein de l'équipe IMEDIA.

Bouclage de pertinence sur les régions

Le mécanisme de bouclage de pertinence appliqué avec succès à la recherche d'images globales a été très peu étudié dans le cas des régions. Au fil des itérations de requêtes, le mécanisme pourrait concerner aussi bien la description individuelle de chaque région que la composition en elle-même. Le but serait d'estimer au mieux la similarité visuelle associée à chaque région mais d'estimer aussi leurs relations spatiales. Des travaux sont en cours dans l'équipe dans cette direction.

Chapitre 7

Résumé des contributions et Conclusions

Résumé des contributions

Nous résumons les contributions apportées dans nos travaux de recherche d'images par régions d'intérêt. Nous avons :

1. introduit la distance quadratique dans sa forme généralisée comme mesure de similarité sur des distributions adaptatives de couleur
2. proposé l'usage d'un algorithme de classification évolué par rapport à l'état de l'art pour la quantification couleur d'images et de régions
3. justifié notre choix des régions d'intérêt par rapport à d'autres représentations partielles d'images et présenté sa complémentarité avec les points d'intérêt
4. présenté une nouvelle approche pour le paradigme 1, motivée par les spécificités du problème de recherche par régions-exemple : la détection grossière et description fine de régions
5. développé une nouvelle méthode de segmentation pour la détection grossière de régions
6. proposé un nouveau descripteur (ADCS) de région. Nous montré que, tout en étant plus compacte, une représentation plus fine de la variabilité couleur des régions conduisait à une amélioration de la précision de recherche
7. introduit le nouveau paradigme 2 de recherche d'images par composition logique de catégories de régions
8. exhibé l'émergence d'une sémantique visuelle des régions dans le paradigme 2

9. proposé des perspectives riches pour le paradigme 2 pour la recherche d'images par régions mais aussi pour des problèmes plus généraux en recherche d'information visuelle
10. implémenté et évalué les paradigmes 1 et 2 dans la plate-forme logicielle Ikona
11. établi le lien entre le paradigme 2 et la recherche de texte

Conclusions

Dans le contexte de la recherche d'images par le contenu, nous jugeons nécessaire de tenir compte du fait qu'une image est généralement perçue par l'utilisateur comme une entité visuelle composite plutôt qu'atomique.

Les différentes contributions de nos travaux ont porté sur de nouvelles approches pour permettre à l'utilisateur d'atteindre plus efficacement les images qu'il recherche en désignant explicitement leurs composantes visuelles. En se positionnant par rapport aux représentations locales d'images existantes, nous avons justifié notre choix pour les régions d'intérêt qui correspond au meilleur compromis en termes de rapidité de recherche et de pertinence visuelle.

Pour le paradigme 1 de recherche par région-exemple, nous avons présenté une nouvelle approche de *détection grossière et description fine de régions*. Elle a été motivée par la spécificité des bases de descripteurs de régions. La nouvelle méthode de segmentation présentée détecte les régions visuellement saillantes, dites "grossières", qui sont intuitives et et susceptibles de constituer des clés de requête visuelle pertinentes pour l'utilisateur. Nous avons proposé un nouveau descripteur de variabilité couleur pour les régions. Produisant une description plus fidèle et plus compacte, nous avons montré qu'une indexation plus fine de la couleur améliorait la similarité des régions retournées. De cette approche résulte un système de recherche d'images par région-exemple plus intuitif et plus précis.

Notre dernière contribution a été l'introduction du "paradigme 2" de recherche d'images par composition logique de catégories de régions. Très différent des paradigmes existants en recherche d'information visuelle, il ne nécessite aucune image ou région exemple. Seule la représentation mentale de l'image cible de l'utilisateur suffit. En interagissant avec le thesaurus photométrique de régions, l'utilisateur peut retrouver des images présentant une composition visuelle spécifique. Nous avons mis en évidence l'émergence d'une forme de "sémantique visuelle" dans la spécification par l'utilisateur de la composition logique des types de régions recherchée. Des perspectives riches ont été proposées pour l'exploitation de cette approche dans le cadre plus général de la recherche d'information visuelle.

Annexe A

Codage d'une image segmentée

Dans le chapitre 3, nous avons présenté la méthode de segmentation permettant d'extraire les régions d'intérêt d'une image en vue de leur utilisation pour les paradigmes 1 et 2 de recherche d'images.

L'information ainsi extraite est principalement de deux natures : les masques de régions et le graphe d'adjacence associé. Nous présentons ici la syntaxe adoptée pour la représentation de ces informations.

Les masques de régions sont représentés par un tableau bidimensionnel correspondant à un sous-échantillonnage spatial de l'image, dans lequel chaque valeur indique l'identifiant de la région de chaque pixel. Le graphe d'adjacence de régions (ou "RAG") contient les informations géométriques élémentaires de chaque région extraites ainsi que les informations relationnelles sur leur adjacence.

Ces informations ont eu les usages suivants dans nos travaux :

- les masques pour l'extraction de descripteur sur chaque région
- les informations géométriques (position, surface, compacité) comme descripteurs géométriques
- les masques et le chaînage des contours pour la sélection et l'affichage des régions dans l'interface de requête

Nous illustrons ici le codage de cette représentation sur l'image exemple "plage" (figure A.1) dont l'image des contours de régions obtenues par segmentation est présentée en figure A.2.

```
171123300000000000000000[...]666666666666[...]5555555555
// Image
//
IMAGE_NAME      001209_024_IDR_.jpg
MASK_WIDTH_HEIGHT       171 123
//
// Regions
//             (7 initial CC's)
```

FIG. A.1 – Image "plage"

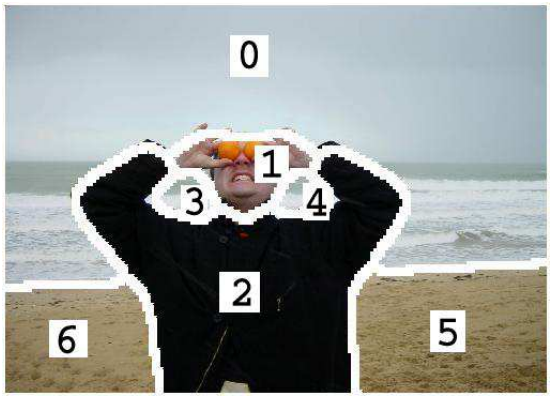


FIG. A.2 – Image des contours des régions extraites et leurs identifiants.

```
NB_OF_REGIONS   7
// Region 0
        SURFACE_RATIO           51
        COMPACTNESS             41
        BARYCENTER_X_Y          87 35
        NB_OF_CONTOURS  1
//              Contour 0
                PERIMETER       665
                START_X_Y       0 0
                CONTOUR_FREEMAN                 33333333333333[...]
        NB_OF_ADJ_REGIONS       4
                ADJ_REGIONS     1 2 5 6
// Region 1
        SURFACE_RATIO           3
        COMPACTNESS             29
        BARYCENTER_X_Y          75 51
        NB_OF_CONTOURS  1
//              Contour 0
                PERIMETER       147
                START_X_Y       62 41
                CONTOUR_FREEMAN                 35535535355353[...]
        NB_OF_ADJ_REGIONS       4
                ADJ_REGIONS     0 2 3 4
// Region 2
        SURFACE_RATIO           23
        COMPACTNESS             40
        BARYCENTER_X_Y          76 85
        NB_OF_CONTOURS  1
//              Contour 0
                PERIMETER       443
                START_X_Y       44 45
                CONTOUR_FREEMAN                 3555355353[...]
        NB_OF_ADJ_REGIONS       6
                ADJ_REGIONS     0 1 3 4 5 6
// Region 3
        SURFACE_RATIO           0
        COMPACTNESS             21
        BARYCENTER_X_Y          57 60
        NB_OF_CONTOURS  1
//              Contour 0
                PERIMETER       65
                START_X_Y       56 53
                CONTOUR_FREEMAN                 355535353[...]
```

```
        NB_OF_ADJ_REGIONS       2
                ADJ_REGIONS     1 2
// Region 4
        SURFACE_RATIO           0
        COMPACTNESS             21
        BARYCENTER_X_Y          93 60
        NB_OF_CONTOURS  1
//              Contour 0
                PERIMETER       59
                START_X_Y       91 53
                CONTOUR_FREEMAN                 35355335[...]
        NB_OF_ADJ_REGIONS       2
                ADJ_REGIONS     1 2
// Region 5
        SURFACE_RATIO           11
        COMPACTNESS             18
        BARYCENTER_X_Y          140 102
        NB_OF_CONTOURS  1
//              Contour 0
                PERIMETER       209
                START_X_Y       166 81
                CONTOUR_FREEMAN                 35555555[...]
        NB_OF_ADJ_REGIONS       2
                ADJ_REGIONS     0 2
// Region 6
        SURFACE_RATIO           7
        COMPACTNESS             17
        BARYCENTER_X_Y          23 105
        NB_OF_CONTOURS  1
//              Contour 0
                PERIMETER       169
                START_X_Y       37 86
                CONTOUR_FREEMAN                 35535555[...]
        NB_OF_ADJ_REGIONS       2
                ADJ_REGIONS     0 2
```

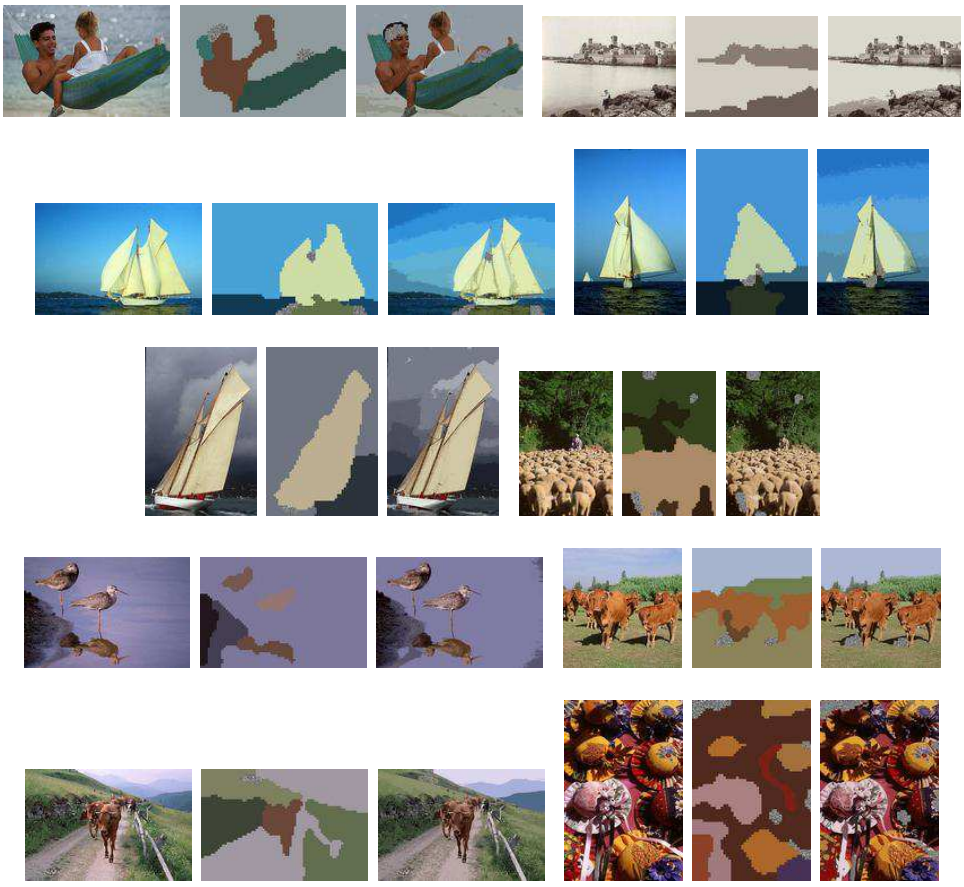
La première ligne définit les masques de régions. Elle donne, pour chaque pixel de l'image considérée à une résolution inférieure, l'identifiant de la région associée. Ici, les 7 régions détectées sont identifiées par les chiffres 0 à 6. Les sept premiers chiffres de la ligne (ici "1711233") constitue l'en-tête nécessaire à la lecture du fichier.

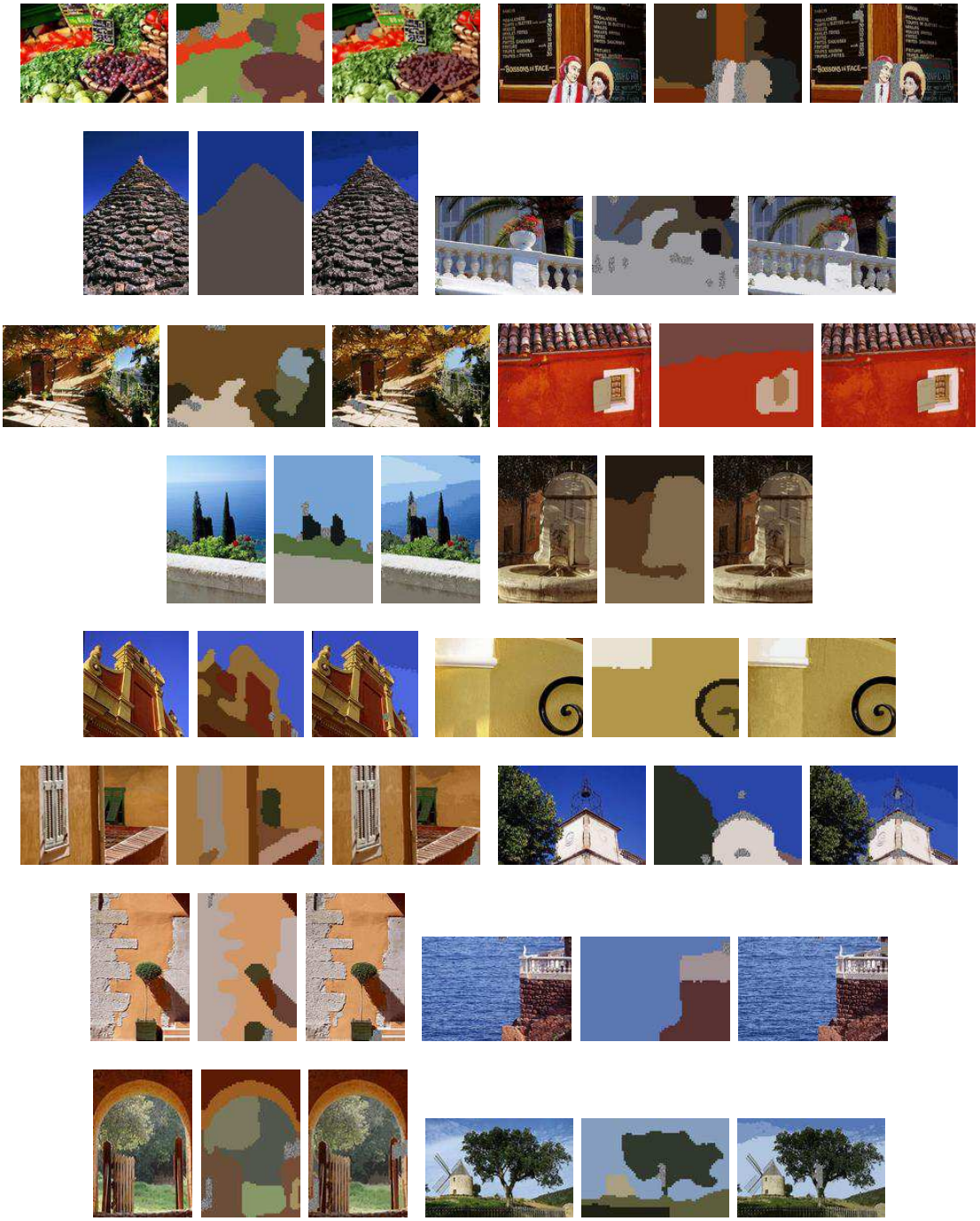
Le reste de la structure renseigne sur les informations géométriques des sept régions : surface, compacité, position, contours et adjacence. Chaque contour de chaque région

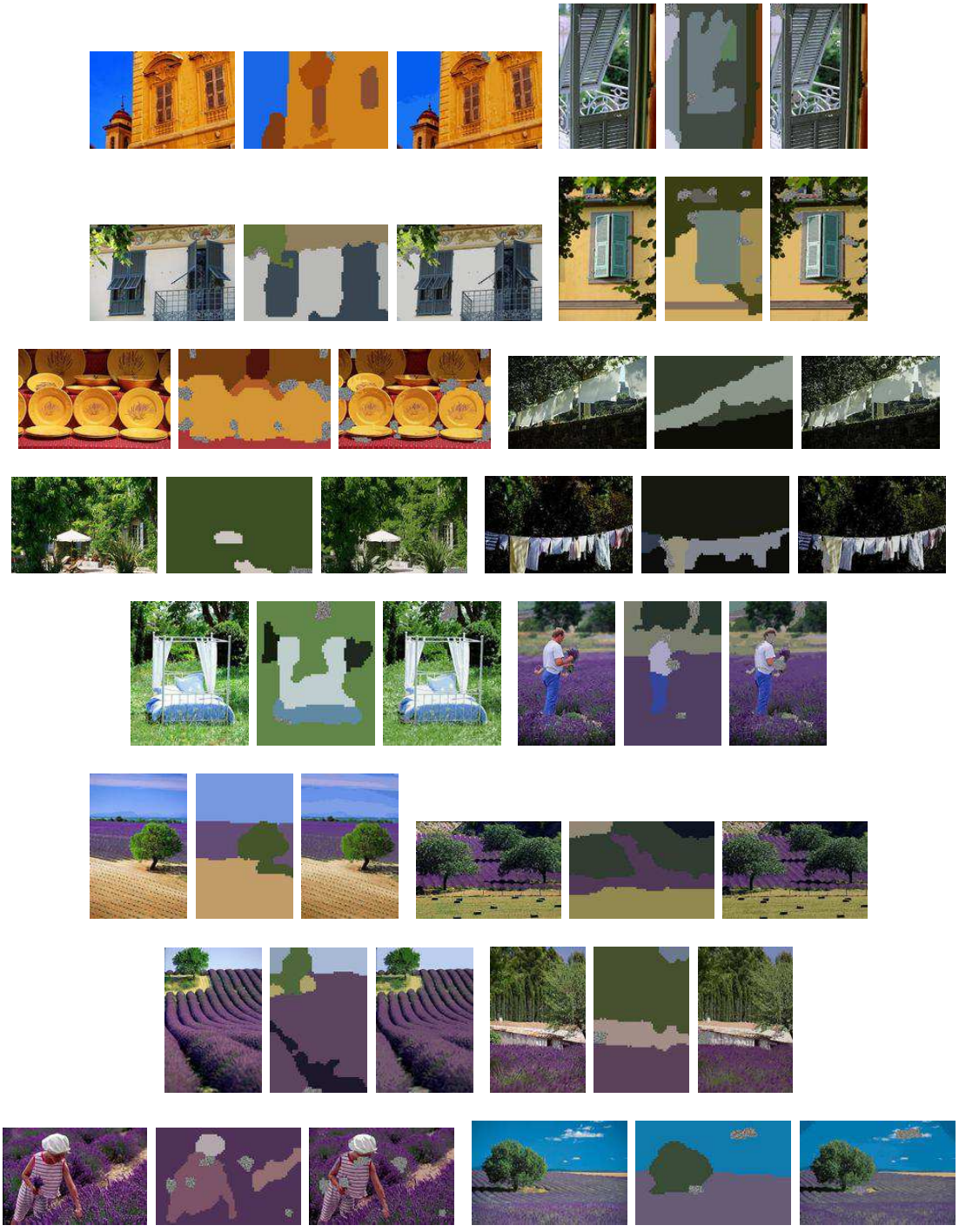
est décrit par son périmètre, les coordonnées de son point de départ et les coordonnées des points suivants par le codage de Freeman [31]. L'adjacence est décrite par la liste des identifiants des régions adjacentes.

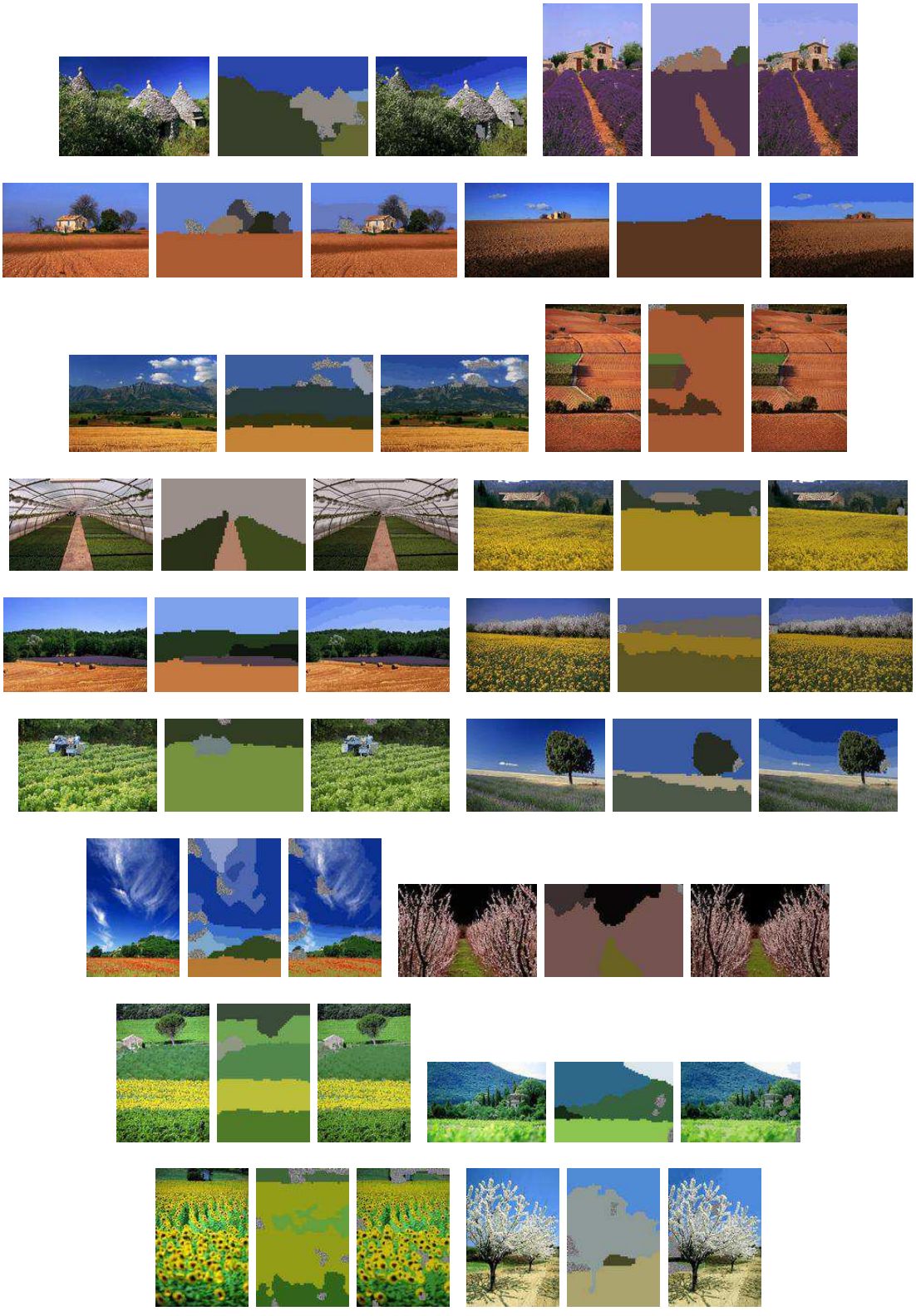
Annexe B

Résultats de segmentation grossière et description fine









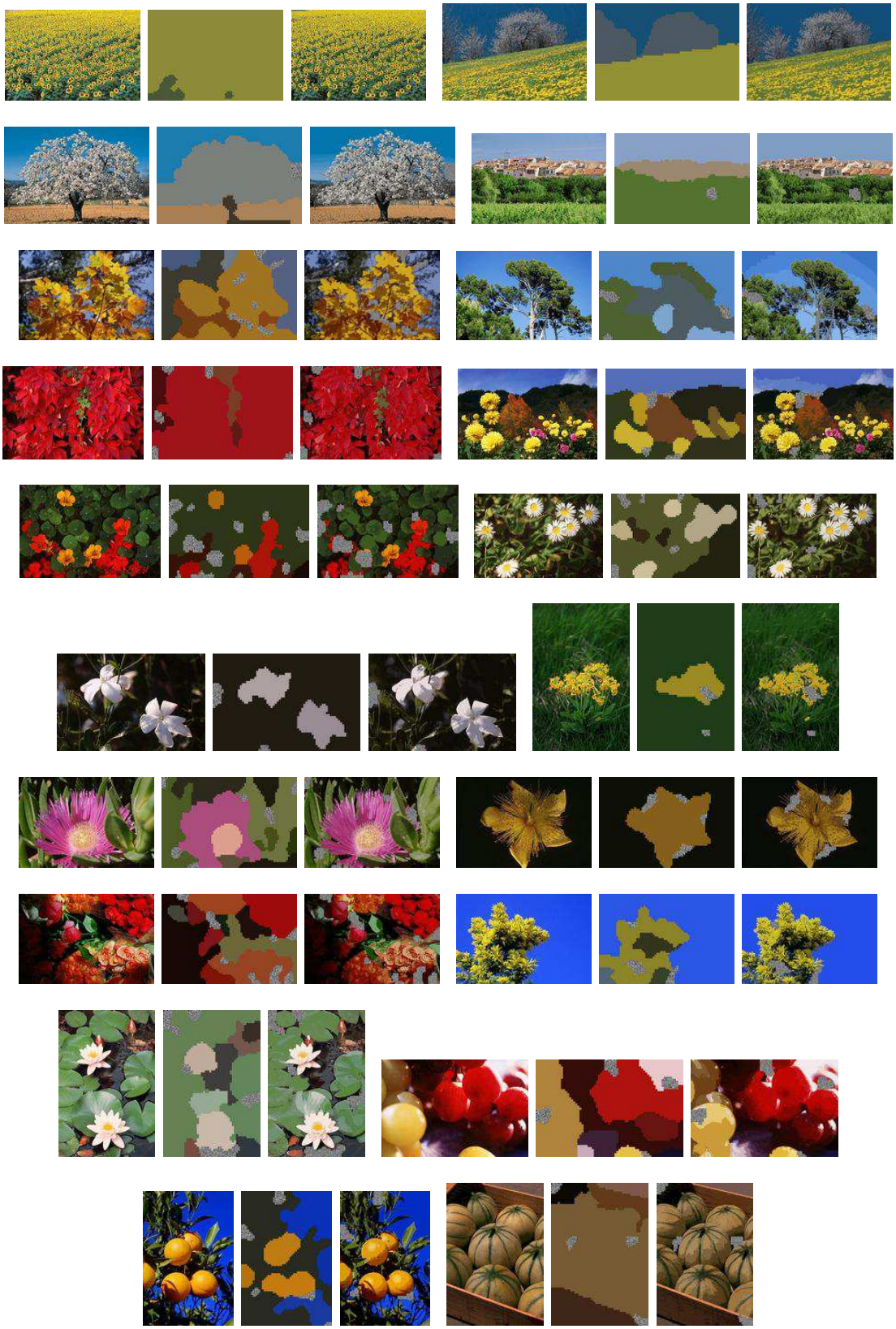


FIG. B.1 – Illustration de la description fine. Chaque paire d'images comporte l'image originale et l'image des régions avec leur nuances couleur utilisées pour leur indexation. Les petites régions supprimées sont grisées. La forte similarité visuelle entre chaque image originale et l'image des nuances de couleur illustre la précision du descripteur.

Bibliographie

[1] P. Aigrain, P. Joly, et V. Longueville. *Medium Knowledge-Based Macro-Segmentation of Video into Sequences.* Intelligent Multimedia Information Retrieval, A.P.M. Press, 1997.

[2] S. Ardizzoni, I. Bartolini, et M. Patella. Windsurf : Region-based image retrieval using wavelets. *DEXA Workshop*, 1999.

[3] Photo Marketing Association. The path from pixels to prints, the challenge of bringing digital imaging to the mass market, 2003.

[4] S. Belongie, C. Carson, H. Greenspan, et J. Malik. Color- and texture-based image segmentation using em and its application to content-based image retrieval. *Proceedings International Conference on Computer Vision (ICCV'98)*, 1998.

[5] J. C. Bezdek. *Pattern Recognition with Fuzzy Objective Functions.* Plenum, New York NY, 1981.

[6] N. Boujemaa. On competitive unsupervized clustering. *International Conference on Pattern Recognition (ICPR'00)*, 2000.

[7] N. Boujemaa, J. Fauqueur, M. Ferecatu, F. Fleuret, V. Gouet, B. Le Saux, et H. Sahbi. Ikona : Interactive generic and specific image retrieval. *International workshop on Multimedia Content-Based Indexing and Retrieval (MMCBIR'2001), Rocquencourt, France*, pages 25–28, 2001.

[8] N. Boujemaa, J. Fauqueur, et V. Gouet. *What's beyond query by example ? to appear in* Trends and Advances in Content-Based Image and Video Retrieval*, L. Shapiro, H.P. Kriegel, R. Veltkamp (ed.).* LNCS, Springer Verlag, 2004.

[9] N. Boujemaa et G. Stamon. Fuzzy modeling in early vision - application to medical image segmentation, 1994.

[10] R. Brunelli et O. Mich. Compass : an image retrieval system for distributed databases. *IEEE International Conference on Multimedia and Expo (ICME'00), New York City*, 2000.

[11] N. Boujemaa C. Vertan. Upgrading color distributions for image retrieval : can we do better ? *Proc. of International Conference on Visual Information System (VIS'00)*, pages 178–188, 2-4 Nov. 2000.

[12] N. Boujemaa C. Vertan. Using fuzzy histograms and distances for color image retrieval. *Challenge of Image Retrieval, Brighton*, 2000.

[13] C. Carson et al. Region-based image querying. *CVPR'97 Workshop on Content-based, Access to Image and Video libraries (CBAIVL)*, 1997.

[14] C. Carson et al. Blobworld : A system for region-based image indexing and retrieval. *Proc. of International Conference on Visual Information System, LNCS vol. 1614*, pages 509–517, 1999.

[15] C. Carson, S. Belongie, H. Greenspan, et J. Malik. Blobworld : Image segmentation using expectation-maximization and its application to image querying. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) 24(8) : 1026-1038*, 2002.

[16] M. La Cascia, S. Sethi, et S. Sclaroff. Combining textual and visual cues for content-based image retrieval on the world wide web. *IEEE Workshop on Content-based Access of Image and Video Libraries (CBAIVL'98)*, june 1998.

[17] Y. Chen et James Z. Wang. A region-based fuzzy feature matching approach to content-based image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 24(9) :1252–1267, 2002.

[18] Y. Cheng. Mean shift, mode seeking, and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 17(8) :790–799, 1995.

[19] D. Comaniciu et P. Meer. Robust analysis of feature spaces : Color image segmentation. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'97), Puerto Rico*, pages 750–755, 1997.

[20] I. Daubechies. *Ten lectures on wavelets*. Philadelphia : SIAM, 1992.

[21] R. N. Dave. Characterization and detection of noise in clustering. *Pattern Recognition Letters*, 12, 1991.

[22] A. DelBimbo. *Visual Information Retrieval*. Morgan Kauffman, San Francisco, CA, 1999.

[23] A.P. Dempster, N.M. Laird, et D.B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of Royal Statistical Society*, 39 :1–38, 1977.

[24] Y. Deng et B. S. Manjunath. An efficient low-dimensional color indexing scheme for region based image retrieval. *Proc. IEEE Intl. Conference on Acoustics, Speech and Signal Processing (ICASSP'99), Phoenix, Arizona*, March 1999.

[25] M. J. Egenhofer et R. D. Franzosa. Point-set topological spatial relations. *International Journal for Geographical Information Systems*, 1991.

[26] J. Fauqueur et N. Boujemaa. Recherche d images par régions d'intérêt : Segmentation grossière rapide et description couleur fine. *Techniques et Sciences Informatiques (TSI), numero special Indexation de Bases d Images fixes ou animees*, 2004.

[27] J. Fauqueur et N. Boujemaa. Region-based retrieval : Coarse segmentation with fine signature. *IEEE International Conference on Image Processing (ICIP)*, pages 609–612, 2002.

[28] J. Fauqueur et N. Boujemaa. Logical query composition from local visual feature thesaurus. *International Workshop on Content-Based Multimedia Indexing (CBMI'03), Rennes, France*, 2003.

[29] J. Fauqueur et N. Boujemaa. New image retrieval paradigm : logical composition of region categories. *IEEE International Conference on Image Processing (ICIP)*, pages 601–604, 2003.

[30] J. Fauqueur et N. Boujemaa. Region-based image retrieval : Fast coarse segmentation and fine color description. *Journal of Visual Languages and Computing (JVLC), special issue on Visual Information Systems*, 15(1) :69–95, 2004.

[31] H. Freeman. Computer processing of line drawing images. *Surveys*, 6(1) :57–97, March 1974.

[32] H. Frigui et R. Krishnapuram. Clustering by competitive agglomeration. *Pattern Recognition*, 30(7) :1109–1119, 1997.

[33] K.S. Fu et J.K. Mui. A survey on image segmentation. *Pattern Recognition*, 13 : 3–16, 1981.

[34] Theo Gevers et Arnold Smeulders. The pictoseek WWW image search system. In *ICMCS, Vol. 1*, pages 264–269, 1999.

[35] S. Gibson et R. Harvey. Morphological color quantization. *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR'01)*, 2001.

[36] V. Gouet et N. Boujemaa. Object-based queries using color points of interest. *IEEE Workshop on Content-Based Access of Image and Video Libraries (CBAIVL'01)*, 2001.

[37] R. M. Gray et D. L. Neuhoff. Quantization. *IEEE Transaction in Information Theory*, 44(6), october 1998.

[38] H. Grecu et P. Lambert. Image retrieval by partial queries. *Proc. International Conference on Image Processing (ICIP'01)*, 2001.

[39] A. Gupta et al. The virage image search engine : an open framework for image management. *SPIE Storage and Retrieval for Image and Video Databases*, 2670, 1996.

[40] E. E. Gustafson et W. C. Kessel. Fuzzy clustering with a fuzzy covariance matrix. *IEEE CDC, San Diego, California*, 1979.

[41] J. Hafner, H. Sawhnay, et al. Efficient color histogram indexing for quadratic form distance functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 17(7) :729–736, July 1995.

[42] R. M. Haralick. Statistical and structural approaches to texture. *Proceedings of IEEE*, 67(5) :786–804, 1979.

[43] R.M. Haralick et L.G. Shapiro. Survey, image segmentation techniques. *Computer Vision Graphics and Image Processing*, 29 :100–132, 1985.

[44] K. Hirata, E. Kasutani, et Y. Hara. On image segmentation for object-based image retrieval. *IAPR International Conference on Pattern Recognition (ICPR'02)*, 2002.

[45] A. Hiroike, Y. Musha, A. Sugimoto, et Y. Mori. Visualization of information spaces to retrieve and browse image data. *International Conference on Visual Information System (VIS'99)*, 1999.

[46] H. Houissa. Segmentation d'images par classification adaptative pour les requêtes partielles. *Mémoire de DEA, Université Paris-sud XI*, 2003.

[47] J. Huang, S. R. Kumar, M. Mitra, W. J. Zhu, et R. Zabih. Image indexing using color correlograms. *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR'97)*, 1997.

[48] T. Huang, S. Mehrotra, et K. Ramchandran. Multimedia analysis and retrieval system (mars) project. *Proceedings of the 33rd Annual Clinic on Library Application of Data Processing - Digital Image Access and Retrieval*, 1996.

[49] A. K. Jain et M. N. Murty. Data clustering : A review. *ACM Computing Surveys*, 1999.

[50] T.Y. Kim et J.H. Han. Partial image matching by measures from connected color regions. *IEEE International Conference on Multimedia and Expo (ICME'00)*, 2000.

[51] B. Ko, H. S. Lee, et H. Byun. Region-based image retrieval system using efficient feature description. *The fifteenth International Conference on Pattern Recognition (ICPR00)*, 2000.

[52] T. Kohonen, editor. *Self-Organizing Maps*. Springer series in Information Science, Berlin, Heidelberg, New-York, third Edition, 2001.

[53] I. Kompatsiaris, E. Triantafillou, et M. G. Strintzis. Region-based color image indexing and retrieval. *IEEE International Conference on Image Processing (ICIP'01)*, 2001.

[54] J. Laaksonen, E. Oja, M. Koskela, et S. Brandt. Analyzing low-level visual features using content-based image retrieval. *International Conference on Neural Information Processing (ICONIP 2000). Taejon, Korea*, 2000.

[55] W. K. Leow et R. Li. Adaptive binning and dissimilarity measure for image retrieval and classification. *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR'01)*, 2001.

[56] J. Li, J. Z. Wang, et G. Wiederhold. Irm : Integrated region matching for image retrieval. *Proc. ACM Multimedia Conference*, pages 147–156, 2000.

[57] J. Lin. Divergence measures based on the shannon entropy. *IEEE Transactions on Information Theory*, 37 :145–151, 1991.

[58] Y. Linde, A. Buzo, et R. M. Gray. An algorithm for vector quantizer design. *IEEE Transactions on Communications*, COM-28 :84–95, 1980.

[59] F. Liu et R. W. Picard. Periodicity, directionality, and randomness : Wold features for image modeling and retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 18(7) :722–733, July 1996.

[60] S. Lloyd. Least squares quantization in pcm's. *Bell telephone laboratory paper*, 1957.

[61] W. Ma et B. S. Manjunath. Edgeflow : A technique for boundary detection and segmentation. *IEEE Transactions on Image Processing*, 2000.

[62] W. Y. Ma et B. S. Manjunath. A texture thesaurus for browsing large aerial photographs. *Journal of the American Society of Information Science*, 49(7) : 633–648, 1998.

[63] W. Y. Ma et B. S. Manjunath. Netra : A toolbox for navigating large image databases. *Multimedia Systems*, 7(3) :184–198, 1999.

[64] W. Y. Ma et B.S. Manjunath. Edgeflow : A framework of boundary detection and image segmentation. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'97)*, pages 744–749, 1997.

[65] J. MacQueen. Some methods for classification and analysis of multivariate observations. *Proc. of the Fifth Berkeley Symp. on Math. Stat. and Prob.*, 1 :281–296, 1967.

[66] J. Malki, N. Boujemaa, C. Nastar, et A. Winter. Region queries without segmentation for image retrieval by content. In *Proc. of International Conference on Visual Information System (VIS'99)*, pages 115–122, 1999.

[67] B. S. Manjunath et W. Y. Ma. Texture features for browsing and retrieval of image data. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Special Issue on Digital Libraries*, 18(8) :837–842, 1996.

[68] B.S. Manjunath, P. Salembier, et T. Sikora. *Introduction to MPEG-7 : Multimedia Content Description Interface*. Wiley, ISBN : 0-471-48678-7, 2002.

[69] J. Mao et A. K. Jain. Texture classification and segmentation using multiresolution simultaneous autoregressive models. *Pattern Recognition*, 25(2), 1992.

[70] M. Mechkour. Un modèle étendu de représentation et de correspondance d'images pour la recherche d'informations. *Thèse de doctorat, Université Joseph Fourier, Grenoble*, 1995.

[71] M. Mechkour, C. Berrut, et Y. Chiaramella. Using a conceptual graph framework for image retrieval. *International conference on MultiMedia Modeling (MMM'95), Singapore*, 1995.

[72] C. Meilhac et C. Nastar. Feedback and category search in image databases. *IEEE International Conference on Multimedia Computing and Systems*, 1999.

[73] D. S. Messing, P. Van Beek, et J. H. Errico. The mpeg-7 colour structure descriptor : image description using colour and local spatial information. *IEEE International Conference on Image Processing (ICIP'01)*, 2001.

[74] D. S. Messing, P. Van Beek, et J. H. Errico. Using color and local spatial information to describe images. *MPEG-7 Technical Report TR13-07, January*, 2001.

[75] T. P. Minka et R. W. Picard. Interactive learning using a society of models. *MIT Technical Report TR349*, 1996.

[76] B. Moghaddam, H. Biermann, et D. Margaritis. Defining image content with multiple regions of interest. *IEEE Workshop on Content-Based Access of Image and Video Libraries (CBAIVL'99)*, 1999.

[77] P. Montesinos, V. Gouet, et R. Deriche. Differential invariants for color images. In *Proceedings of* 14^{th} *International Conference on Pattern Recognition (ICPR'98)*, Brisbane, Australia, 1998.

[78] N. Nes et M. L. Kersten. Region-based indexing in an image database. In *Proceedings International Conference on Imaging Science, Systems, and Technology*, pages 207–215, Las Vegas, ND, USA, July 1997.

[79] W. Niblack, R. Barber, W. Equitz, M. Flickner, et al. The qbic project : querying images by content using color, texture, and shape. *Proc. SPIE (Storage and Retrieval for Image and Video Databases)*, 1908 :173–187, 1993.

[80] V. Oria, M. Tamer Ozsu, L. Liu, X. Li, J. Z. Li, Y. Niu, et P. J. Iglinski. Modeling images for content-based queries : The disima approach. *International Conference on Visual Information System (VIS'97)*, 1997.

[81] N.R. Pal et S.K. Pal. A review on image segmentation techniques. *Pattern Recognition*, 26(9) :1277–1294, 1993.

[82] G. Pass et R. Zabih. Histogram refinement for content-based image retrieval. *IEEE Workshop on Applications of Computer Vision*, pages 96–102, 1996.

[83] G. Pass et R. Zabih. Comparing images using joint histograms. *Multimedia Systems*, 7(3) :234–240, 1999.

[84] Theodosios Pavlidis. *Structural Pattern Recognition*. Springer-Verlag, Berlin, 1977.

[85] A. Pentland, R. Picard, et S. Sclaroff. Photobook : Content-based manipulation of image databases. *SPIE Storage and Retrieval for Image and Video Databases*, II(2185), Feb. 1994.

[86] R. W. Picard. Toward a visual thesaurus. *MIT Technical Report TR358*, 1995.

[87] J. Puzicha, J. Buhmann, Y. Rubner, et C. Tomasi. Empirical evaluation of dissimilarity measures for color and texture. *IEEE International Conference on Computer Vision (ICCV)*, 1999.

[88] A. Rao, R. Srihari, et Z. Zhang. Spatial color histograms for content-based image retrieval. *IEEE International Conference on Tools with Artificial Intelligence*, pages 183–186, 1999.

[89] A. Rao, R. Srihari, et Z. Zhang. Geometric histogram : A distribution of geometric configurations of color subsets. *SPIE : Internet Imaging, 3964*, January 2000.

[90] R. Rickman et J. Stonham. Content-based image retrieval using color tuple histograms. *SPIE proceedings*, 2670 :2–7, 1996.

[91] J. Rissanen. Modeling by shortest data description. *Automatica*, 1978.

[92] P. K. Robertson. Visualizing color gamuts : A user interface for theeffective use of perceptual color spaces in data displays. *IEEE Computer Graphics and Applications*, pages 50–64, sept. 1988.

[93] J. J. Rocchio. *Relevance feedback in information retrieval*. Prentice Hall, Englewood Cliffs, New Jersey, USA, 1971.

[94] Y. Rubner. Perceptual metrics for image database navigation. *PhD Thesis, Stanford University*, 1999.

[95] Y. Rubner, C. Tomasi, et L. Guibas. The earth mover distance as a metric for image retrieval. *Stanford University Technical Report*, 1998.

[96] G. Salton. *Automatic Text Processing : The Transformation, Analysis, and Retrieval of Information by Computer*. Addison-Wesley, 1989.

[97] S. Sangwine et R. Horne. *The colour image processing handbook*. Chapman and Hall, 1999.

[98] B. Le Saux. Navigation dans les bases d'images. *PhD dissertation, INRIA Rocquencourt*, 2003.

[99] B. Le Saux et N. Boujemaa. Unsupervised robust clustering for image database categorization. *IAPR International Conference on Pattern Recognition (ICPR'02)*, 2002.

[100] C. Schmid et R. Mohr. Local grayvalue invariants for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 19(5) :530–535, 1997.

[101] S. Sclaroff, L. Taycher, et M. La Cascia. Imagerover : A content-based image browser for the world wide web. *IEEE Workshop on Content-based Access of Image and Video Libraries (CBAIVL'97)*, june 1997.

[102] A. Smeulders, M. Worring, et S. Santini. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 22(12), 2000.

[103] J. R. Smith. Integrated spatial and feature image systems : Retrieval, analysis and compression. *PhD Dissertation, Columbia University*, 1997.

[104] J. R. Smith et S. F. Chang. Tools and techniques for color image retrieval. *IST/SPIE Proceedings*, pages 426–437, 1996.

[105] J. R. Smith et S. F. Chang. Visualseek : A fully automated content-based image query system. *ACM Multimedia Conference, Boston, MA, USA*, pages 87–98, 1997.

[106] V. Sridhar, M. A. Nascimento, et X. Li. Region-based image retrieval using multiple-features. *Proc. of International Conference on Visual Information System (VIS'02), Hsin-Chu, Taiwan*, 2002.

[107] R. O. Stehling, M. A. Nascimento, et A. X. Falcao. An adaptive and efficient clustering-based approach for content-based image retrieval in image databases. *International Data Engineering and Application Symposium*, 2001.

[108] R. O. Stehling, M. A. Nascimento, et A. X. Falcao. Microm : A metric distance to compare segmented images. *Proc. of International Conference on Visual Information System (VIS'02), Hsin-Chu, Taiwan*, 2002.

[109] M. Stricker et M. Orengo. Similarity of color images. *Storage and Retrieval for Image and Video Databases III, SPIE Proceedings Series*, 2420, 1995.

[110] M. Swain et D. Ballard. Color indexing. *International Journal of Computer Vision (IJCV)*, 7(1) :11–32, 1991.

[111] P.M. Tardif et A. Zaccarin. Multiscale autoregressive image representation for texture segmentation. *Image Processing*, 3 026 :327–337, 1997.

[112] C. Town et D. Sinclair. Content based image retrieval using semantic visual categories. *ATT Technical Report*, 2001.

[113] C. Vertan et N. Boujemaa. Embedding fuzzy logic in content based image retrieval. *19th International Meeting of the North American Fuzzy Information Processing Society NAFIPS*, 2000.

[114] J. Wang. Simplicity : A region-based image retrieval system for picture libraries and biomedical image databases. *Proc. ACM Multimedia Conference, Los Angeles*, 2000.

[115] J. Z. Wang et Y. Du. Rf*ipf : A weighting scheme for multimedia information retrieval. *IEEE International Conference on Image Analysis and Processing (ICIAP)*, 2001.

[116] J. Z. Wang, Jia Li, et Gio Wiederhold. Simplicity : Semantics-sensitive integrated matching for picture libraries. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 2001.

[117] L.L. Wang et S.N. Yang. Color image retrieval based on hidden markov models. *IEEE Transactions on Image Processing*, 6(2) :332–339, 1997.

[118] G. Wyszecki et W. S. Stiles. *Color Science : concepts and methods, quantitative data formulae.* John Wiley, New York, 1982.

[119] T. Huang Y. Rui et S. Mehrotra. Content-based image retrieval with relevance feedback in mars. *IEEE International Conference on Image Processing (ICIP'97)*, 1997.

[120] Z. Zhang, R. Deriche, O. Faugeras, et Q.-T. Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *Artificial Intelligence Journal*, 78 :87–119, 1995.