Systèmes de représentation multi-échelles pour l'indexation et la restauration d'archives médiévales couleur

Julien Dombre

THÈSE

pour l'obtention du Grade de
DOCTEUR DE L'UNIVERSITÉ DE POITIERS
(Faculté des Sciences Fondamentales et Appliquées)
(Diplôme National - Arrêté du 25 avril 2002)

École Doctorale : **Sciences Pour l'Ingénieur**
Secteur de Recherche : **Traitement du signal et des images**

Présentée par :
Julien DOMBRE

Systèmes de représentation multi-échelles pour l'indexation et la restauration d'archives médiévales couleur

Directeur de Thèse :
Christine FERNANDEZ-MALOIGNE

Soutenue le 18 décembre 2003 devant la Commision d'Examen composée de :

Mme Marinette Revenu, Professeur, université de Caen, GREYC ..... Rapporteur
M. Patrick Lambert, Maître de Conférences HDR, université de Savoie, LISTIC ..... Rapporteur

M. Ludovic Macaire, Maître de Conférences, université de Lille, I^3D ..... Examinateur
M. Jean-Marc Ogier, Professeur, université de la Rochelle, L3I ..... Examinateur
Mme Christine Fernandez-Maloigne, Professeur, université de Poitiers, SIC ..... Examinateur
M. Noël Richard, Maître de Conférences, université de Poitiers, SIC ..... Examinateur

M. Éric Palazzo, Professeur, université de Poitiers, CESCM ..... Invité

REMERCIEMENTS

Je tiens tout d'abord à remercier MARINETTE REVENU et PATRICK LAMBERT d'avoir accepté d'être rapporteurs de cette thèse ainsi que LUDOVIC MACAIRE pour en avoir été un examinateur très pointilleux. La version finale de ce manuscrit a profité de leurs nombreuses et pertinentes remarques. Je tiens également à remercier JEAN-MARC OGIER d'avoir accepté d'être président du jury. Merci aussi à ÉRIC PALAZZO d'avoir pris part en tant qu'invité à ma soutenance et qui a amené le point de vue des historiens de l'art sur mes travaux.

Je voudrais remercier tout particulièrement CHRISTINE FERNANDEZ-MALOIGNE qui m'a permis de débuter cette thèse au sein du laboratoire nouvellement nommé SIC en me proposant un sujet très intéressant. Elle a su être présente dès que le besoin s'en faisait sentir que ce soit pour des problèmes administratifs ou scientifiques. Elle a su également instaurer une très bonne ambiance de travail au sein de l'équipe ACTIOM par les repas organisés régulièrement chez elle. Ne t'inquiètes pas la prochaine fois aussi tu te retrouveras dans la piscine !

Je remercie NOËL RICHARD d'avoir co-encadré ce travail et de m'avoir laissé tout ma liberté au cours de ces 3 ans. J'ai ainsi pu être entièrement maître des directions à suivre tout au cours du déroulement de ma thèse.

Je remercie aussi très vivement BENOIT pour avoir pris le temps de relire l'ensemble de mon manuscrit ainsi que pour ses nombreuses remarques qui ont fortement contribué à l'amélioration de sa qualité. Je profite également pour saluer toutes les personnes qui m'ont côtoyé au sein du même bureau et qui d'une manière ou d'une autre ont réussi à me supporter mais surtout qui ont permis de toujours travailler dans une très bonne ambiance : BENOIT (encore lui...), CARINE, DENIS, GUILLAUME et LAURENT.

Merci également à toutes les personnes de l'équipe peintures murales du CESCM, et en particulier AURÉLIA BOLOT, pour leur accueil chaleureux ainsi que pour les échanges très intéressants, nécessaires à l'avancée de cette collaboration.

Un grand merci aussi à ÉRIC et MANU qui m'ont donné la possibilité de varier un peu mes activités durant ces trois ans. Leur contact m'a permis de réfléchir plus précisément à mes perspectives d'avenir et j'espère devoir supporter *grincheux* encore longtemps...

Je voudrais remercier particulièrement FRANÇOISE et SYLVIE pour leur sympathie aux pauses café et leur efficacité dans l'organisation et la résolution des problèmes administratifs.

Enfin, bien sûr, je remercie toutes les autres personnes qui m'ont permis de passer trois années inoubliables sur Poitiers, qu'ils soient au laboratoire ou non : Aline, Anne-Marie, Anne-Sophie, Benoît Papapa, Bertrand, Cédric, Christelle, Christian, Isabelle, Jean-Christophe et toutes les personnes du volley, Éric (l'autre petit), Yves, Philippe et Sylvie, Philippe (encore un petit), Pierre, Sébastien, Xavier et Gaëlle, Yannick, Yves...

TABLE DES MATIÈRES

TABLE DES FIGURES

NOTATIONS

Ci-dessous sont présentées les différentes notations qui seront utilisées tout au long de ce manuscrit.

Notations générales	
$\mathbb{R}$	Espace des réels
$\mathbb{N}$	Espace des entiers
$\mathbb{B} = \{vrai, faux\}$	Espace des booléens
Notations liées à l'image	
l	Largeur de l'image
h	Hauteur de l'image
$\Omega \in \mathbb{N}^2$	Domaine de l'image
$\Theta \in \mathbb{R}^3$	Domaine des couleurs
$I : \Omega \rightarrow \Theta$	Fonction image
$\zeta_I = \{I(x)/x \in \Omega\}$	Ensemble des couleurs de l'image I
$nbc_I = Card(\zeta_I)$	Nombre de couleurs présentes dans l'image I
$H : \zeta \rightarrow \mathbb{R}$	Fonction histogramme
$MC_I : \zeta^2 \rightarrow \mathbb{R}$	Fonction cooccurrence de l'image I
$V : \Omega^2 \rightarrow \mathbb{B}$	Fonction de voisinage
$[MC_I]$	Matrice de cooccurrence de l'image I
Notations liées aux graphes	
$G = V, E$	Un graphe G
$\|V\|$	Nombre de nœuds de G ou ordre du graphe G
$\|E\|$	Nombre d'arêtes de G
G_i	i^e nœud de G
G_{ij}	arête liant les i^e et j^e nœuds de G
$V = \{G_i/i \in [1 \ldots \|V\|]\}$	Ensemble des nœuds de G
$E = \{G_{ij}/(i,j) \in [1 \ldots \|V\|]^2\}$	Ensemble des arêtes de G

CHAPITRE 1

INTRODUCTION GÉNÉRALE

Les travaux présentés dans ce mémoire ont été réalisés au sein du laboratoire SIC (Signal Image Communications) de l'université de Poitiers. Ils ont été développés sous la direction de CHRISTINE FERNANDEZ-MALOIGNE et de NOËL RICHARD en collaboration avec le Centre d'Études Supérieures des Civilisations Médiévales (CESCM) de l'Université de Poitiers.

Depuis une dizaine d'années et la démocratisation des communications en réseau, d'importants volumes d'informations sont échangés permettant la mise en place de banques de données très variées et de plus en plus vastes. Quasiment tous les domaines applicatifs sont ainsi touchés du militaire au médical en passant par l'éducation ou les médias. Les illustrations les plus frappantes sont les moteurs de recherche tels que *Google* ou *Yahoo* qui recensent chacun quelques millions de pages Web et facilitent l'accès à leurs contenus. Mais des bases de données beaucoup plus spécifiques existent aussi. L'INA[1] propose au sein de l'Inathèque l'intégralité de la diffusion des programmes de radio et de télévision, des spots publicitaires et des vidéo-clips. De nombreux autres entreprises ou organismes d'état mettent également en place des bases de données pour organiser leur savoir et leur savoir-faire. Des bases de vidéos sont ainsi créées pour fournir du matériel aux journaux télévisés. La bibliothèque nationale de France a lancé la numérisation de ses ouvrages pour les pérenniser et les mettre à disposition du plus grand nombre. La photothèque du CESCM, avec qui cette étude a été réalisée, est en train de créer des bases d'images numériques de bâtiments anciens allant des vues architecturales jusqu'aux fresques présentes dans les édifices.

Les bases de données actuelles peuvent ainsi être très diverses : génériques ou spécialisées, contenant du texte et/ou des images et/ou des vidéos. Les problèmes principaux liés à celles-ci sont le stockage, la représentation et l'accès aux données. Quelle est la manière la plus adaptée pour entreposer les différents éléments ? Comment pouvons nous les représenter simplement et précisément ? Comment rechercher l'information qui nous intéresse ? Actuellement, la réponse généralement apportée à ces questions est l'utilisation d'une description textuelle des composants. Le stockage est alors facile et les recherches simples. Les différents moteurs de recherche sur le Web fonctionnent sur ce principe. À partir de mots-clés, ils fournissent les adresses les plus pertinentes contenant ces motifs. Pour des données textuelles, les systèmes actuels sont très performants.

Malheureusement, cette approche est fortement limitée pour décrire les images ou les vidéos. En effet, une représentation textuelle de telles données est complexe mais surtout soumise à une forte variabilité. La langue utilisée, le niveau d'expertise de la personne, les détails apportés et le temps nécessaire à la description sont autant de limitations fortes qui font que cette méthode n'est généralement pas utilisable pour de telles bases de données. Pour ces différentes raisons, les modes de recherche explorés actuellement sont fondés sur le contenu même des médias. Ainsi,

[1] Institut National de l'Audiovisuel.

a) b)

FIG. 1.1 – Fresques médiévales issues des bases du CESCM

c'est l'information contenue dans les images ou les vidéos qui est utilisée pour les caractériser automatiquement et non pas une description textuelle réalisée manuellement. Pour les images, ce type de travaux se regroupe sous le terme : indexation d'images basée sur le contenu (ou *CBIR* en anglais pour *Content Based Image Retrieval*).

L'ensemble de ce travail se situe dans cette mouvance. Le partenariat avec le CESCM a pour but de fournir aux historiens de l'art un ensemble d'outils leur permettant d'effectuer des recherches fines au sein de leurs bases d'images et en particulier celles contenant des fresques médiévales. Actuellement, ils disposent de bases d'images indexées textuellement ce qui est très limitant dans certains cas. La figure [1.1] présente quelques exemples des éléments dont ils disposent[2]. Les problèmes qui se posent aux médiévistes sont très divers :

- disposant d'un cliché inconnu, pouvons nous déterminer d'où il provient ?
- une fresque est fortement détériorée (cf. figure [1.1].a). Pouvons nous en trouver une semblable susceptible de nous aider pour la restauration ?
- pour réaliser des regroupements entre édifices et ainsi pouvoir dater une fresque ou faire ressortir un atelier d'artisans ayant travaillé sur différents lieux, pouvons nous retrouver autre part un dessin semblable d'un objet ou d'une partie d'un objet ?
- disposant d'un objet portant à confusion quant à son indexation textuelle, pouvons nous trouver des objets similaires déjà indexés ?

Les outils à mettre en place sont donc des systèmes d'aide à l'expert. Ils doivent surtout apporter un plus significatif par rapport aux données textuelles présentes dans leurs bases.

La première difficulté de cette étude est liée à la nature même des images dont nous disposons. Avec le temps, les fresques sont souvent dégradées : les couleurs sont dépigmentées, des effacements partiels peuvent être présents... Nous devons également répondre à des problèmes de recherche d'objets. Or, en indexation par le contenu, les images sont généralement caractérisées de manière globale ce qui ne permet pas de détecter efficacement des objets.

Sans aller jusqu'à une description sémantique des images, nous devons pouvoir caractériser les images le plus précisément possible pour répondre aux problèmes qui nous sont posés. Un état

[2] voir aussi annexe B.8.

de l'art des diverses structures classiquement employées en traitement d'images nous a amené à proposer l'utilisation d'une structure de représentation multi-échelles des images : le graphe pyramidal. En effet, les objets à détecter pouvant être de taille variable, il paraît intéressant de travailler à plusieurs niveaux de détails. Le graphe pyramidal met ainsi en évidence les objets au sein de la représentation rendant possible leurs caractérisations et par conséquent leurs recherches. De plus, l'utilisation de différentes échelles de segmentation permet la détection des petits objets mais aussi de décrire la composition des différents éléments de la scène. Cette structure nécessitant diverses partitions plus ou moins fines de l'image, nous avons développé un algorithme de segmentation multi-échelles fondé sur un critère de rupture de modèle textural. Chaque région du graphe pyramidal est ensuite décrite par sa couleur, sa texture et sa forme. Pour ces trois statistiques, nous avons étudié les différentes techniques existantes pour trouver celles qui s'adaptaient le mieux à notre structure. Pour la couleur, nous avons également proposé de nouvelles techniques pouvant être utilisées pour comparer efficacement des histogrammes à supports colorimétriques différents.

La structure étant complètement définie, elle caractérise de manière fine le contenu des images en fournissant une description des différents objets et de leurs compositions. Nous avons alors proposés divers algorithmes pour comparer les graphes pyramidaux permettant d'effectuer des recherches d'images de manière globale. De plus, cette nouvelle représentation des images rend très efficace les requêtes partielles utiles pour retrouver un objet au sein de bases quelconques et en particulier celle contenant les fresques médiévales du CESCM.

Pour présenter ce travail, l'organisation de ce manuscrit est la suivante.

Le premier chapitre se propose de mettre en place le contexte général d'une chaîne d'analyse d'images numériques couleur en faisant le parallèle avec le système visuel humain.

Le deuxième chapitre s'intéresse alors plus spécifiquement à l'indexation d'images. Il pose les différents problèmes rencontrés dans ce domaine. Il expose également les différentes méthodes déjà mises en place pour répondre à certains d'entre eux et montre la nécessité de trouver une représentation des images adaptée à l'application visée.

Le troisième chapitre traite des différents systèmes de représentation d'images classiques. Après avoir montré la limitation des structures planes, nous présentons l'intérêt de la notion de multi-résolutions pour la description d'images. Cette présentation aboutit alors à la définition d'un nouveau système de représentation multi-échelles d'images, le graphe pyramidal. Celui-ci est inspiré des pyramides irrégulières et permet une description fine des objets contenus dans les images.

Cette nouvelle structure nécessite de disposer d'un ensemble de partitions plus ou moins fines des images. Le quatrième chapitre présente par conséquent le problème de la segmentation multi-échelles ainsi que le détail de l'algorithme que nous avons développé et les résultats obtenus.

Au sein du graphe pyramidal, il est nécessaire de décrire chaque région indépendamment comme peut l'être une image pour un système d'indexation classique. Le cinquième chapitre présente donc l'étude comparative que nous avons menée pour déterminer les statistiques de couleur, de forme et de texture les plus adaptées à notre structure.

Maintenant que la description des images que nous avons définie est complète, le sixième chapitre expose les différentes méthodes développées pour déterminer la similarité entre images ou pour effectuer une recherche d'objet. Divers résultats sont également exposés ainsi que des comparaisons avec des systèmes d'indexation d'images existants.

Le dernier chapitre s'intéresse particulièrement au problème des fresques médiévales et présente deux applications différentes mises en place dans ce cadre. La première est utile pour déterminer l'origine géographique d'une image inconnue et la seconde se focalise sur la recherche d'objets et est illustrée en prenant comme exemple les apôtres.

Nous conclurons sur les intérêts du système développé et les perspectives de ces travaux de recherche.

CHAPITRE 2

L'IMAGERIE COULEUR - DE LA PERCEPTION À L'ANALYSE

Sommaire

2.1 Introduction

Les sciences dites *pour l'ingénieur* tentent généralement de modéliser des phénomènes réels pour les analyser, les reproduire et/ou les contrôler. L'imagerie couleur, discipline liée au traitement du signal et à l'informatique, fait partie intégrante de ce schéma. Ainsi, elle essaye de comprendre et de reproduire de manière artificielle les comportements de l'être humain dans tous les domaines de la vision. Cette tâche est extrêmement complexe car la proportion de neurones du cerveau humain entièrement dévoués à cette tâche est estimée à 80-90%[YOUNG91]. Par conséquent, la vision peut être considérée comme le sens principal de l'être humain.

Toute action de la vision humaine peut sommairement être divisée en trois phases : la perception, la transmission et l'analyse. L'homme perçoit le monde qui l'entoure grâce à ses yeux auxquels parvient un flux lumineux composé de photons. Le signal reçu comporte ainsi différentes longueurs d'ondes. Une fois cette information acquise, elle est transformée puis transmise au cerveau via le nerf optique. Enfin, l'analyse est effectuée par le cerveau ; grâce à elle, l'homme étudie son environnement et prend alors une décision et/ou réalise une action (cf. figure [2.1]).

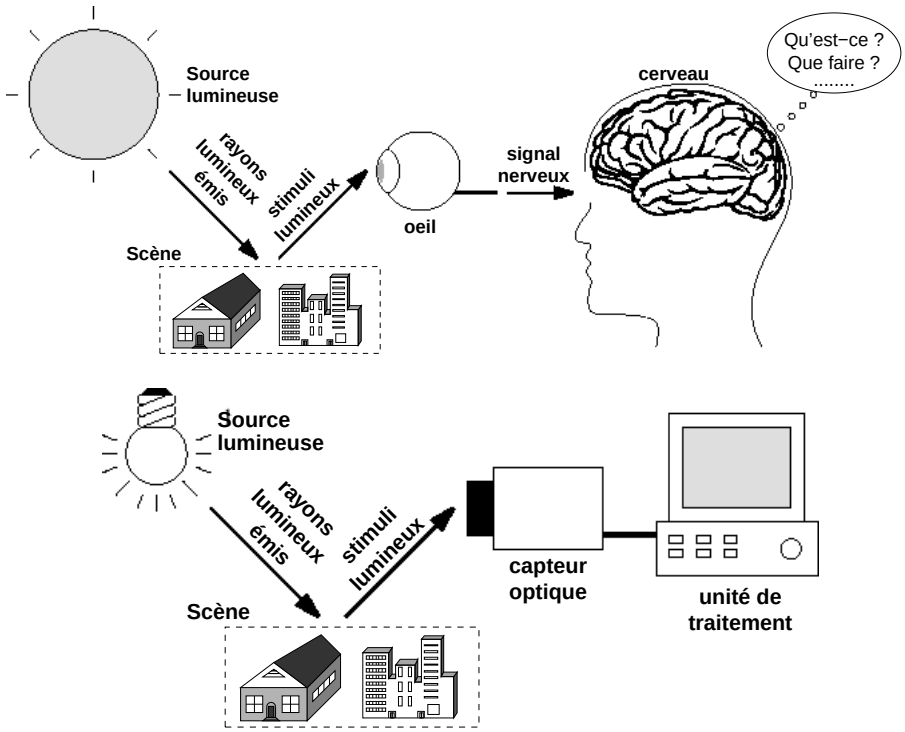


FIG. 2.1 – Schémas de principe comparés de la vision humaine et artificielle
*En haut : perception humaine, en bas : vision artificielle (inspirés de [*VANDEN00*].)*

De même, l'imagerie couleur se décompose en trois étapes. Un capteur optique acquiert tout d'abord des images tout comme le fait l'œil. Ces images sont ensuite stockées sous forme numérique pour être finalement analysées et/ou traitées en fonction de l'application visée (cf. figure [2.1]).

Ce chapitre présente les analogies entre la vision humaine et l'imagerie couleur aux trois niveaux : perception/acquisition, transmission/représentation et analyse/traitement.

2.2 Perception - l'œil en tant qu'objectif

Cette partie se propose de présenter les analogies entre l'œil et les systèmes artificiels mis en place pour l'imiter : les appareils photographiques ou les caméras. La composition de ces récepteurs est équivalente : une surface photosensible et un système optique réalisant la mise au point. Nous présentons tout d'abord le fonctionnement de l'œil[DARRAS95] puis celui des imageurs artificiels qui s'en sont fortement inspirés.

2.2.1 L'œil

Globalement, l'œil est une sphère pilotable suivant trois axes de rotation. Les éléments principaux le composant sont, dans le sens de cheminement de l'information image (cf. figure [2.2]) :

- **la sclère** n'intervient pas dans le processus de vision ; c'est l'enveloppe externe de l'œil qui se transforme sur l'avant en une cornée quasi-ronde et transparente. Elle est normalement blanche et transparente ;
- **la cornée** est la couche externe convexe de l'œil. Elle est composée de quatre couches successives. Sa courbure dépend des individus et varie aussi avec l'âge ;
- **l'humeur aqueuse** est un liquide salin et alcalin sous pression qui maintient la rigidité du globe oculaire ;
- **l'iris** est un diaphragme vertical percé en son centre par la pupille. Il régule la quantité de lumière atteignant la rétine en ajustant la taille et la forme de la pupille ;
- **le cristallin** est une lentille transparente placée derrière l'iris. Il complète le rôle de la cornée et assure la mise au point de l'image optique ; ce phénomène constitue l'accommodation ;
- **le corps vitré** constitue 80% du volume de l'œil et est composé d'un liquide albumineux parfaitement transparent sous forme de gelée. Son indice de réfraction varie suivant la longueur d'onde ; ce phénomène est la cause de l'aberration chromatique. Il diminue aussi l'acuité visuelle ;
- **la rétine** constitue une interface entre le flux lumineux et le système nerveux dont elle fait partie intégrante. Elle est constituée de deux types de photorécepteurs : les cônes et les bâtonnets qui vont transformer l'information photonique en stimuli électriques ;
- **la fovéa** est la zone centrale de la rétine qui se trouve au niveau de l'axe optique de l'œil. Elle est essentiellement constituée de cônes ;
- **le nerf optique** permet le transport de l'information visuelle électrique issue de la rétine vers le cerveau. Au niveau de la connexion rétine/nerf optique, une zone appelée **papille** ne contient aucun photorécepteur.

La liste précédente met en évidence deux structures principales : le système optique réalisant la mise au point de l'image sur la rétine et l'interface photosensible transformant les signaux lumineux reçus en un signal neuronal. La zone qui nous intéresse plus particulièrement est la zone effectuant la transformation d'un signal lumineux multispectral en un signal compréhensible par le cerveau, c'est-à-dire la rétine.

2.2.1.1 La rétine

La rétine tapisse le fond de l'œil. C'est un tissu d'environ 4 cm de diamètre et 250 μm d'épaisseur qui reçoit le flux lumineux focalisé par l'ensemble du système optique. Elle est directement connectée au nerf optique auquel elle transmet une traduction des signaux lumineux en impulsions neuronales. Pour cela, la lumière doit pénétrer jusqu'au fond de la rétine où sont situées les cellules photosensibles. Ces photorécepteurs sont de deux types :

- **les cônes** pointus sont responsables de la vision diurne (vision photopique) et chromatique. Ils sont en moyenne 6.5 millions divisés en trois catégories sensibles au bleu, vert et rouge respectivement dénommés S (pour *Short wavelengths*), M (*Medium*) et L (*Long*). Pour un cône S, 20 M et 40 L sont généralement présents. Leur maxima de réponse spectrale sont situés à 440 nm (S), 530 nm (M) et 560 nm (L) (cf. figure [2.3]) ;
- **les bâtonnets** sont achromatiques et adaptés à la vision nocturne (vision scotopique). Ils sont utiles pour la perception des formes et des mouvements. Ils sont beaucoup plus nombreux

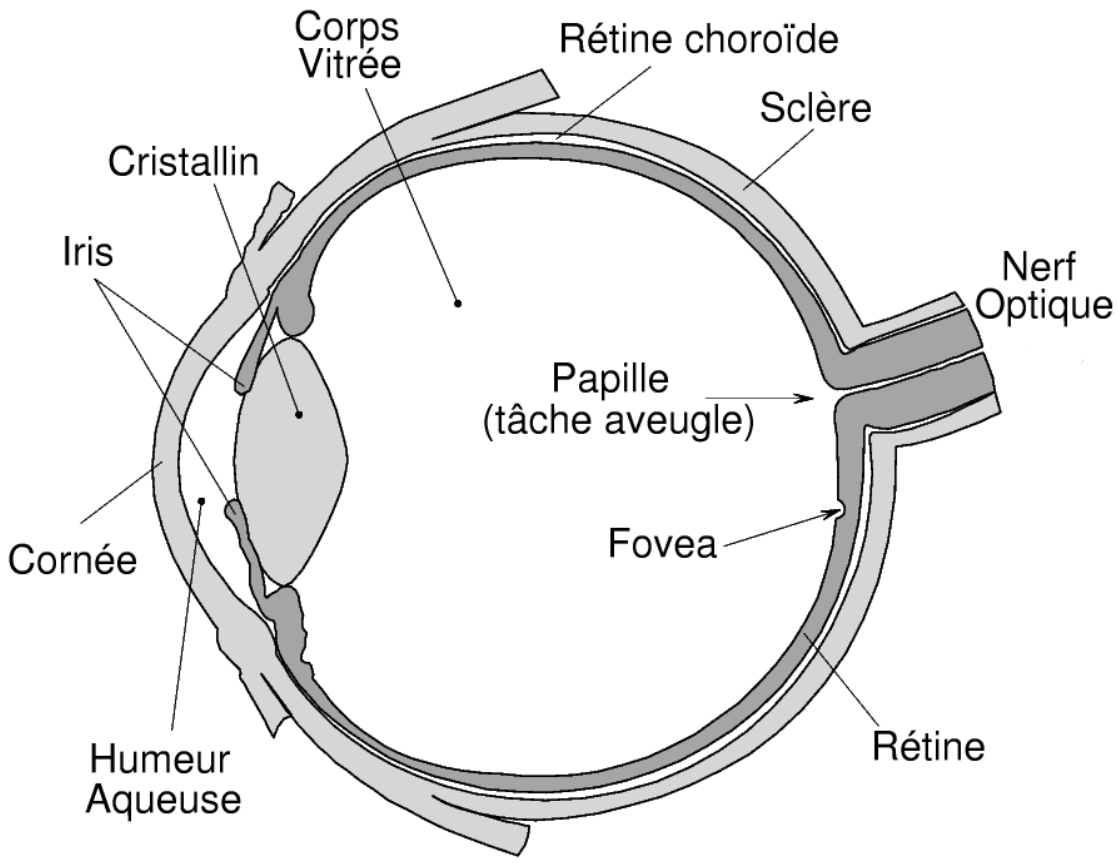


FIG. 2.2 – Coupe schématique de l'œil

que les cônes (120 millions en moyenne). Étant constitués de rhodopsine, ils sont sensibles à une large bande du spectre (cf. figure [2.3]).

Les cônes sont concentrés au niveau de la fovéa qui est une région d'un diamètre angulaire d'environ 1° proche du centre de la rétine alors que les bâtonnets se retrouvent en grande majorité à la périphérie. Il est à noter qu'il existe aussi une tâche aveugle, ou papille, sans aucun photorécepteur au niveau de la connexion avec le nerf optique (cf. figure [2.4]). Au niveau de la fovéa, les différents types de cônes sont positionnés aléatoirement au sein de sites formant un arrangement hexagonal régulier et très serré. Par contre, à la périphérie leur organisation devient plus irrégulière ; les cônes grossissent et des bâtonnets s'insèrent entre eux (cf. figure [2.5]).

Indépendamment de son organisation spatiale, la rétine possède aussi plusieurs couches (cf. figure [2.6]). La couche la plus profonde est composée des cellules photosensibles (bâtonnets et cônes). Ces photorécepteurs reçoivent l'information optique par l'intermédiaire de pigments visuels (iodopsine pour les cônes et rhodopsine pour les bâtonnets) et la transmettent au cerveau par influx nerveux grâce à plusieurs types de cellules. Lors de cette transmission une compression de l'information est nécessaire car les 130 millions de photorécepteurs sont reliés à seulement 1 million de fibres nerveuses. Au niveau de la fovéa, la connexion cône/fibre nerveuse est directe mais en dehors plusieurs récepteurs influencent une même fibre. Pour cela, les cellules horizontales et amacrines sont chargées de propager l'information latéralement. Les cellules bipolaires peuvent

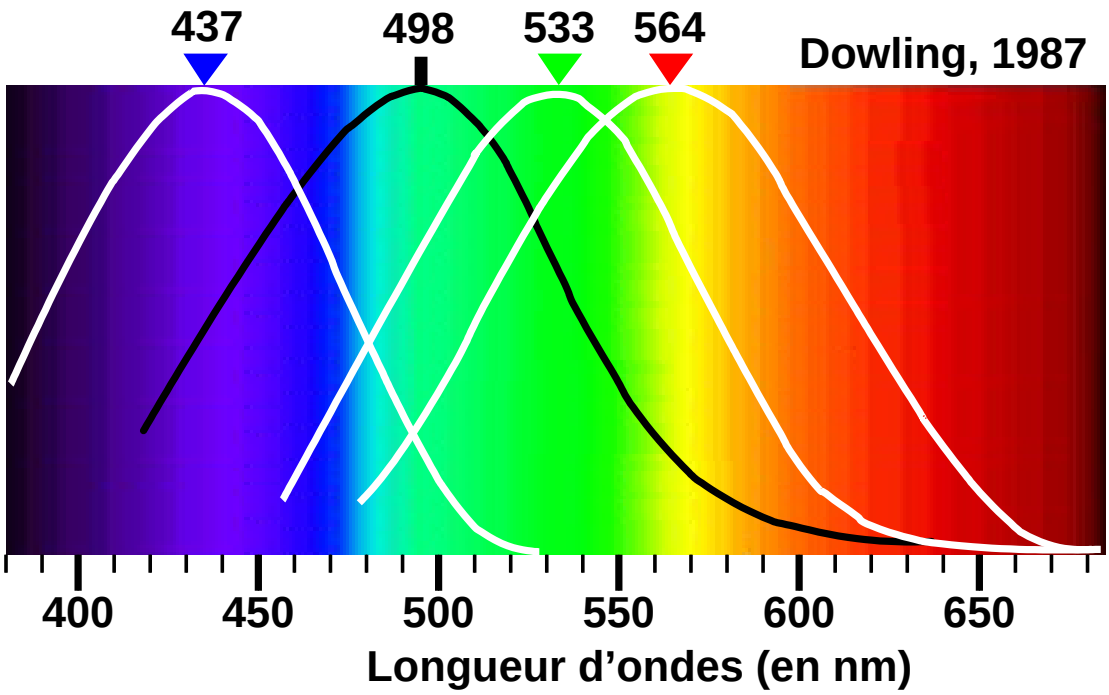


FIG. 2.3 – Fonctions d'absorption relative des photorécepteurs[DOWLIN87]
Les trois types de cônes apparaissent en blanc et les bâtonnets en noir.

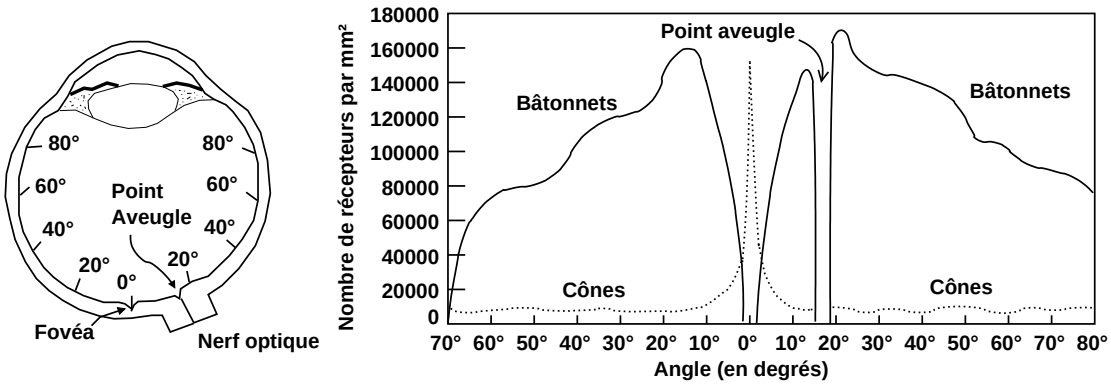
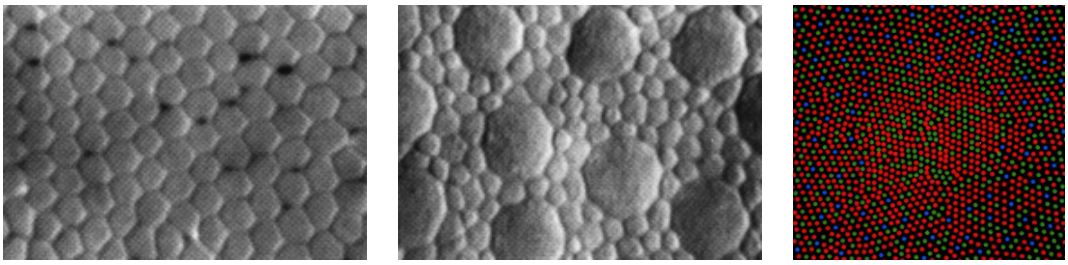


FIG. 2.4 – Répartition des photorécepteurs au niveau de la rétine

avoir deux rôles : inhibitrices ou excitatrices et font le lien avec les cellules ganglionnaires. Enfin, l'impulsion visuelle électrique est convoyée vers les fibres du nerf optique par les cellules ganglionnaires.

2.2.1.2 Les principaux phénomènes de la vision

L'accommodation L'accommodation est la faculté d'adapter l'œil en fonction de la distance aux objets observés. C'est l'équivalent du système de mise au point pour un système imageur artificiel. Celle-ci est réalisée tout d'abord par la cornée qui assure les deux tiers de la convergence des



a) b) c)

FIG. 2.5 – Mosaïque des photorécepteurs sur la rétine[CURCIO90]
a) La fovéa : arrangement régulier et serré de cônes.
b) La périphérie : la taille des cônes augmente et des bâtonnets plus petits s'insèrent entre eux.
Les images a) et b) montrent une zone de $35 \times 25\ \mu m^2$.
c) Arrangement aléatoire des cônes au niveau de la fovéa.

rayons lumineux. De par sa faculté de déformation, le cristallin complète ensuite la mise au point de l'image projetée sur la rétine.

Les maladies courantes des yeux telles que l'hypermétropie, la myopie et la presbytie sont des troubles issus du dysfonctionnement de l'accommodation. La presbytie vient en particulier d'une diminution de l'amplitude de déformation du cristallin.

L'adaptation à la lumière La luminosité de notre environnement varie fortement (nuit noire, soleil). L'œil doit donc être capable de capter les photons dans l'obscurité mais aussi de se protéger d'une lumière trop vive. Il dispose de plusieurs systèmes pour s'adapter aux variations de luminosité ambiante. Ce phénomène est visible, par exemple, lors d'un passage d'une pièce fortement éclairée à une pièce sombre ou l'inverse. Dans les deux cas, il faut quelques secondes pour que la vision se stabilise. Pour cela, trois systèmes entrent en jeu :

- *la régulation de la quantité de lumière* atteignant la rétine est réalisée grâce à l'iris. Il est capable de modifier la forme et la taille de la pupille pour laisser passer plus ou moins de lumière tout comme le fait l'ouverture du diaphragme d'un objectif photographique. Ce phénomène très rapide n'est cependant que temporaire et permet aux autres dispositifs plus lents de se mettre en place ;
- *le pigment visuel* est composé de rhodopsine pour les bâtonnets et d'iodopsine pour les cônes ; il a pour charge d'absorber les photons et d'exciter les cellules visuelles. Ces molécules possèdent, de plus, la propriété de blanchir avec l'intensité lumineuse et ainsi de réduire leur pouvoir absorbant. 7 minutes sont nécessaires aux pigments visuels des cônes pour se régénérer entièrement et 40 minutes pour ceux contenus dans les bâtonnets. Ces délais expliquent le fait que nous sommes moins gênés par une lumière forte soudaine que par l'effet inverse ;
- *le traitement neural des informations* est encore mal compris. Cependant, il semble que les neurones en contact avec les cellules rétiniennes jouent également un rôle dans leur contrôle du gain.

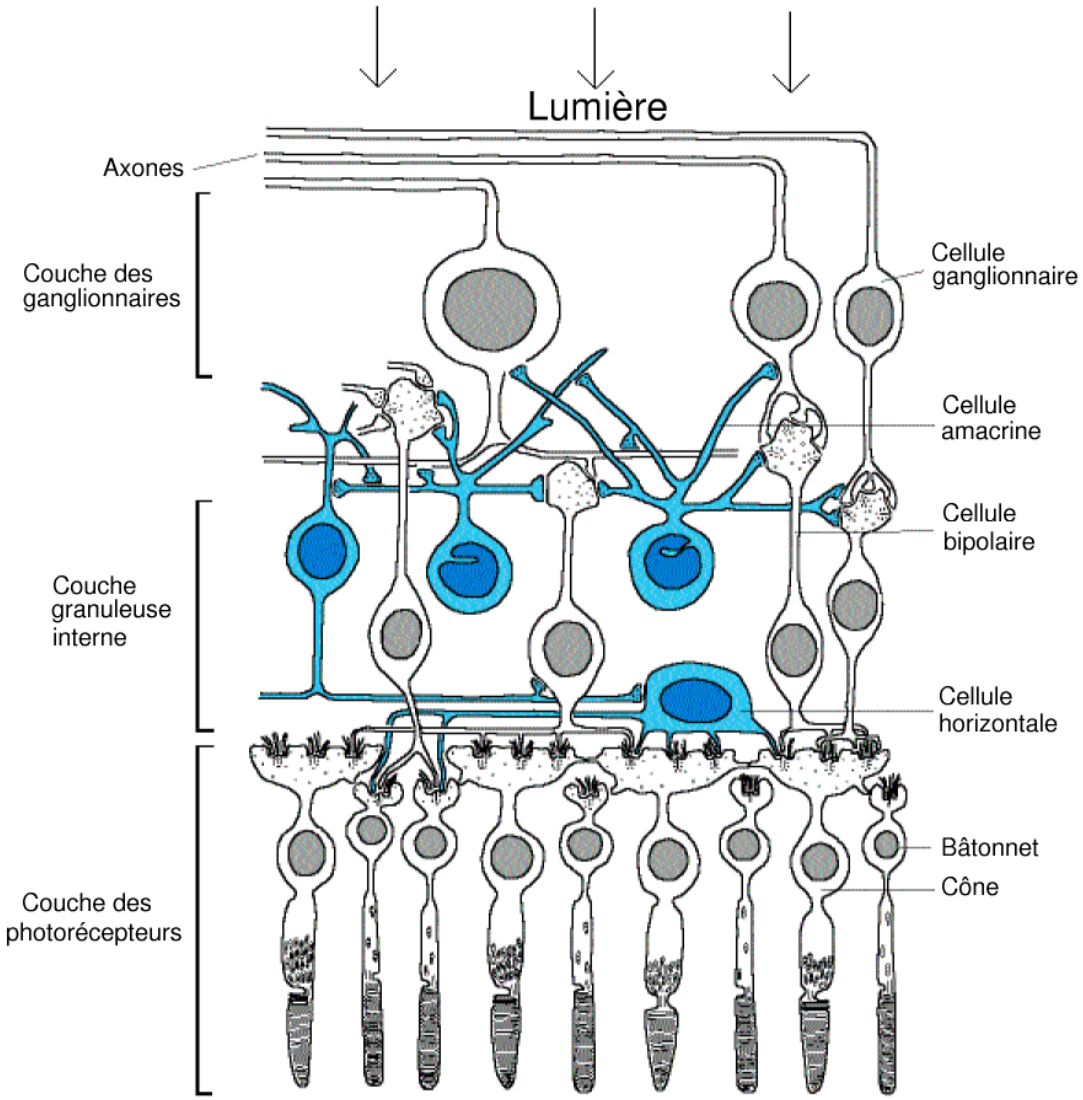


FIG. 2.6 – Coupe schématique de la rétine

L'acuité visuelle L'acuité est le pouvoir de résolution, c'est-à-dire la limite de taille des objets décelables. Généralement, l'acuité minimum de l'œil est d'une minute d'arc ce qui correspond à un point d'une imprimante 300 ppp vu à une distance de 30 cm (cela dépend aussi du contraste). Pour cela, l'image projetée sur la rétine par le système optique doit être de bonne qualité. Les défauts optiques réduisent donc la précision de la vision. Un autre facteur jouant sur l'acuité visuelle est intrinsèque à la rétine. Ainsi, elle compte globalement une centaine de millions de photorécepteurs mais le nerf optique n'est composé que d'un million de fibres. L'information est donc compressée avec pertes.

La sensibilité à la lumière La sensibilité absolue est l'aptitude à détecter de faibles quantités de lumière. Celle de l'œil est liée à la probabilité pour un photon d'être absorbé. Étant de taille

plus importante que les cônes et sensibles sur une bande plus large du spectre, les bâtonnets offrent une meilleure sensibilité à la lumière. Seuls quelques photons suffisent à provoquer une sensation[LE GRA72].

La sensibilité différentielle quant à elle correspond à l'écart de luminance juste discernable ; c'est le contraste minimum visible à une luminance donnée. Cette caractéristique est aussi fonction de la longueur d'onde.

La vision des couleurs La vision des couleurs se rapporte à la capacité à différentier les longueurs d'ondes du spectre du flux lumineux arrivant à l'œil. Il faut tout de même faire attention à ne pas confondre la couleur physique portée par la lumière et la perception qui en est faite. L'homme est trichromate mais de nombreux êtres vivants ne sont que monochromates ou dichromates car ne disposant respectivement que d'un ou deux types de cônes.

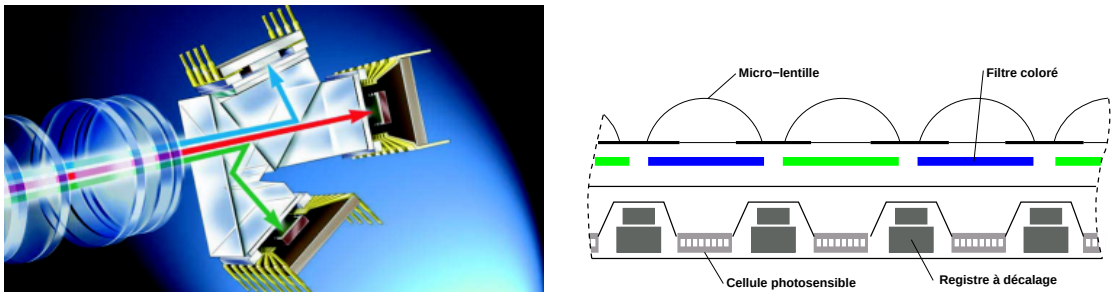
En 1802, YOUNG avançait (suivi par HELMHOLTZ) l'idée que l'œil est composé de trois types de cellules de sensibilités spectrales différentes (rouge, vert et bleu)[YOUNG07, VONHEL67]. Les premières critiques à cette approche furent émises en 1878 par HERING qui proposait plutôt que l'homme est sensible aux différences de couleurs[HERING78]. Ainsi, les couleurs seraient perçues via les oppositions noir/blanc, rouge/vert et bleu/jaune. En fait, de récentes études menées par PADGHAM et SAUNDERS montrent que les deux théories sont présentes[PADGHA75]. Les trois types de cônes sont bien sensibles à des longueurs d'ondes différentes, et donc à des couleurs différentes, mais l'information sur les couleurs antagonistes serait aussi transmise au cerveau. Cette dernière est calculée au niveau de la rétine par les cellules horizontales par un simple agencement d'inversions et de sommations des signaux issus des cônes.

Globalement, l'apparence colorée ne résulte pas d'une transcription brute des signaux lumineux mais plutôt d'une manipulation des signaux reçus par les photorécepteurs au niveau de la rétine puis de leur interprétation par le cerveau. La couleur est donc finalement une vue de l'esprit !

2.2.2 Systèmes imageurs

Bien évidemment, des solutions artificielles ont été mises en place pour imiter l'œil et en particulier pour permettre la capture d'images. Le but principal de tels dispositifs est la restitution de la prise de vue effectuée dans le but de recréer le même stimulus visuel lors de sa restitution. La visualisation la plus fidèle possible de l'image captée est ainsi recherchée.

Tout comme pour l'œil, la première partie d'un imageur artificiel est composée d'un système optique permettant de focaliser les rayons lumineux sur une zone photosensible. Dans le cadre d'un appareil photographique, cela correspond à l'objectif. Ce dispositif peut être plus ou moins complexe. Un système de diaphragme jouant le rôle de l'iris est généralement présent (obturateur). Il permet de laisser passer plus ou moins de lumière. De plus, la focalisation des rayons est réalisée par un ensemble de lentilles et de miroirs plus ou moins complexe suivant la qualité du système optique. La mise au point est réalisée quand le flux lumineux est focalisé de manière à obtenir une image nette sur la zone photosensible. C'est un système d'accommodation artificiel. Dans certaines caméras numériques, des prismes dichroïques peuvent aussi être utilisés pour séparer le flux lumineux et l'amener vers plusieurs (généralement 3) capteurs indépendants améliorant de



a) b)

FIG. 2.7 – Les capteurs *CCDs*
a) Schéma de principe d'un capteur tri-CCDs.
b) Un ensemble de capteurs CCDs dominés de micro-lentilles focalisantes et de filtres couleurs.

cette manière le pouvoir de résolution (cf. figure [2.7]). La qualité de l'image en est donc améliorée une fois résolus les problèmes liés au recalage des différentes images.

Dès que la lumière arrive sur la zone photosensible, elle est transformée en une information exploitable pour la restitution. Sur les films argentiques des appareils photographiques, la sensibilité aux radiations des cristaux d'halogénures d'argent (combinaison de l'argent avec un halogène) permet la capture de l'information lumineuse. Le développement du film révèle ensuite les zones exposées à la lumière. Les systèmes numériques sont formés quant à eux par un ensemble de composants électroniques photosensibles qui transforment la lumière en une excitation électrique fonction de l'énergie reçue. Les plus utilisés sont les *CCDs* (*Charged Coupled Device*) qui sont essentiellement achromatiques. Pour produire une image couleur, des filtres couleurs sont placés au dessus des capteurs (cf. figure [2.7]) : l'ensemble forme un *CFA* (*Color Filter Array*). Le plus répandu des *CFAs* est le modèle $GRGB$ de Bayer (cf. figure [2.8]) mais des arrangements basés sur $CYGM$ (*Cyan, Yellow, Green, Magenta*) sont aussi utilisés. Sony a également proposé dernièrement un arrangement $RGBE$, avec E pour *Emerald*. D'après Sony, l'ajout de l'*Emerald* permettrait de se rapprocher plus du système de perception humaine et de réduire les erreurs de reproduction des couleurs. La sortie de la matrice de *CCDs* est alors composée d'un sous-échantillonnage des différentes composantes chromatiques. Un post-traitement des données, le *demosaicing*, est ensuite nécessaire pour obtenir l'image finale (cf. figure [2.8]). Les imageurs classiques utilisent des filtres sensibles aux fréquences du visible mais il est possible d'en employer d'autres pour créer des images de rayonnements non visibles (radar, infrarouge...). La restitution du signal reçu est alors réalisée en fausses couleurs.

Une fois les signaux lumineux captés, leur restitution est un problème complexe. Comment interpréter les données acquises ? En argentique, l'exposition plus ou moins longue du film au révélateur ne donnera pas du tout le même rendu visuel. En numérique, les signaux électriques émis par les capteurs doivent être interprétés en terme de couleur en tenant compte de la réponse spectrale des différents filtres. Ce problème est très complexe car il fait intervenir un grand nombre de paramètres. Ainsi, le rendu visuel (en considérant un système de visualisation parfait) des systèmes imageurs artificiels est très souvent imparfait et fortement variable.

Les systèmes analogiques fournissent en sortie une image papier (photographie...) contrairement aux systèmes numériques qui stockent les images sous forme informatique. De nombreux formats de codage des images sont utilisés pour compresser plus ou moins le volume d'informa-

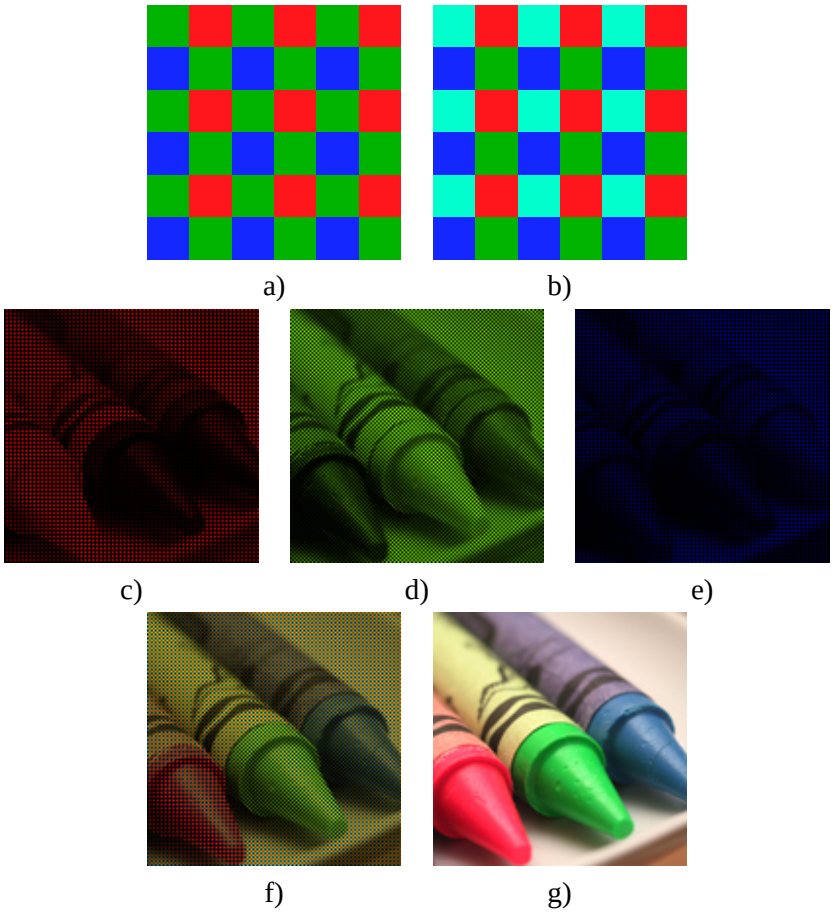


FIG. 2.8 – Principe de *demosaicing* à partir d'un *CFA* de Bayer
a) Arrangement des différentes couleurs de filtres sur le CFA de Bayer.
b) Arrangement des différentes couleurs de filtres sur le nouveau CFA de Sony.
c), d) et e) : Canaux rouge, vert et bleu issu du CCD avec un CFA de Bayer.
f) L'image brute en sortie de la matrice de CCDs (mélange des trois canaux).
g) L'image obtenue après demosaicing.

tions. Pour l'analyse d'images, c'est plutôt la manière de représenter une couleur qui est importante.

2.3 Représentation sous forme d'images couleurs

Au niveau de l'œil, le signal lumineux est reçu puis transformé en un signal neuronal interprétable directement par le cerveau. Par contre en vision artificielle, la transmission de l'information image n'est pas directe ; il est donc nécessaire de la stocker. Il existe aujourd'hui de nombreuses techniques de codage de l'information couleur. Le but de cette section est de les présenter rapidement en faisant ressortir leurs domaines d'applications et leurs caractéristiques.

L'œil dispose de trois types de récepteurs chromatiques. De plus, les diverses expériences menées sur l'appariement de couleurs ont permis de démontrer le principe de trivariance visuelle de la couleur. L'idée de ces études est de réaliser le mélange pondéré de diverses sources lumineuses pour produire une couleur à égaliser, tout comme le ferait un peintre à partir d'un ensemble de tubes de couleur pour obtenir une teinte particulière. De cette manière, il a été montré que toute couleur peut être obtenue par combinaison linéaire de 3 couleurs dites *primaires*. Plusieurs combinaisons différentes peuvent d'ailleurs amener à la même perception colorée : les couleurs sont alors dites *metamères*. Par conséquent, il apparaît naturel de coder la couleur avec trois composantes numériques.

Nous présenterons donc brièvement, les espaces couleurs les plus fréquemment utilisés en commençant par ceux utilisés en acquisition et en restitution d'images. Nous exposerons ensuite les espaces *standards* définit par la CIE (Commission Internationale de l'Éclairage) pour tenter d'uniformiser les espaces couleurs utilisés. Enfin, nous verrons l'intérêt d'utiliser des espaces perceptuellement uniformes ou fondés sur le système visuel humain pour le traitement d'images.

Ne voulant pas alourdir ce manuscrit et étant peu utiles à sa suite, nous ne rentrerons pas ici dans les détails des différents espaces colorimétriques mais exposerons plutôt leurs différentes caractéristiques. Pour de plus amples informations, vous pourrez vous reporter à la thèse de NICOLAS VANDENBROUCKE et en particulier à son deuxième chapitre très complet intitulé « Représentation de la couleur »[VANDEN00]. Vous pourrez également vous reportez au livre d'ALAIN TRÉMEAU, CHRISTINE FERNANDEZ-MALOIGNE et PIERRE BONTON[TRÉMEA03].

2.3.1 Espaces liés à l'acquisition et à l'affichage

Le plus répandu des espaces couleurs est sans aucun doute l'espace RVB (cf. figure [2.9]). Il est utilisé en particulier pour les télévisions couleurs équipées de tubes cathodiques dont les trois types de luminophores émettent à des longueurs d'ondes différentes. Il n'est en fait pas correct de parler de l'espace RVB mais plutôt des espaces RVB. En effet, ce type d'espaces est fondé sur la définition de trois primaires (R, G et B) qui peuvent varier. Ainsi, les télévisions européennes et américaines n'utilisent pas les mêmes primaires (d'où des formats de transmission différents : NTSC, PAL et SECAM, voir plus loin). Le blanc de référence utilisé pour étalonner l'échelle de couleurs peut aussi être différent ce qui modifie encore l'espace couleur. Malgré ces différences, il est toujours possible de passer simplement d'un espace à un autre par une transformation linéaire[1]. De plus, les composantes RVB sont dépendantes du matériel d'acquisition. Effectivement, la prise de vue d'un même objet sous un éclairage constant sera restituée différemment suivant l'imageur utilisé (avec le même système de visualisation).

En imprimerie, l'espace $CMYK$ (*Cyan, Magenta, Yellow, black* où le K désigne *Key black*) est majoritairement utilisé. Il est l'extension de CMY, le complémentaire de RVB, qui est un système *soustractif* bien adapté au papier. En effet, les systèmes de visualisation tels que les écrans travaillent en émission contrairement au papier sur lequel la lumière est réfléchie. L'encre déposée doit donc permettre de sélectionner les couleurs au sein du flux lumineux incident pour que sa réflexion soit de la couleur désirée. Le cyan permet d'enlever les teintes rouges, le magenta les vertes et le jaune les bleues. Par conséquent, l'ajout de cyan, de magenta et de jaune sur une feuille

[1] L'adresse suivante http://www.brucelindbloom.com/index.html?Equations.html fournira d'ailleurs la plupart des équations de changement d'espaces utiles en colorimétrie.

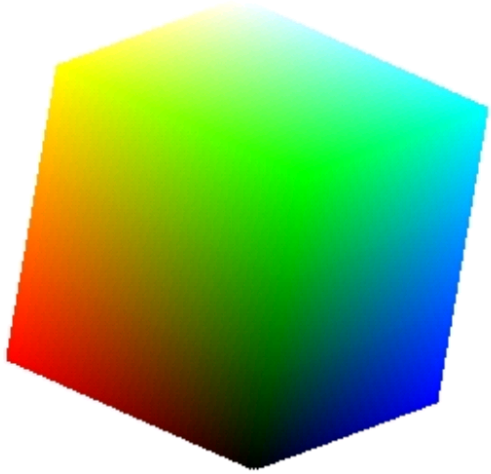


FIG. 2.9 – Cube des couleurs RVB

de papier produira du noir et non du blanc comme dans un système additif. Malheureusement, le filtrage réalisé par les différentes encres n'est pas parfait et un noir pur est quasiment impossible à obtenir avec CMY. L'utilisation d'une teinte noire spécifique permet ainsi d'obtenir un noir très propre (et aussi de foncer une couleur à moindre coût car les encres colorées sont plus chères que la noire).

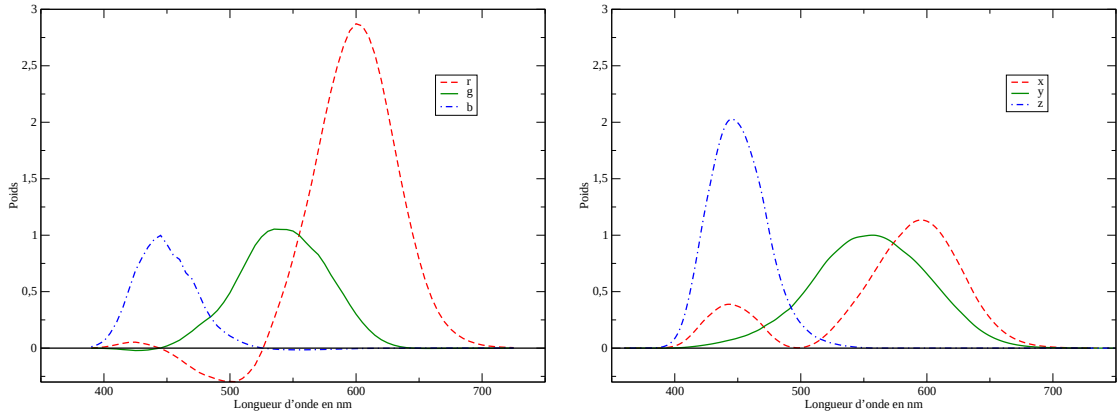
Pour le codage des signaux couleurs de télévision, deux problèmes se posent : pouvoir visualiser un signal noir et blanc sur une télévision couleur mais surtout pouvoir visualiser un signal couleur sur les anciennes télévisions noir et blanc. Dans ces conditions, il apparaît naturel d'utiliser un système séparant l'information de luminance et celle de chrominance. Ainsi, la norme NTSC utilise l'espace YIQ alors que PAL adopte Yuv. La composante Y code la luminance et les deux autres la chrominance. Les flux vidéos numériques emploient plutôt l'espace YC_bC_r. Cet espace est en particulier utilisé pour la télévision numérique ou le DVD ; il est aussi inclus dans le nouveau standard JPEG2000.

2.3.2 Espaces standards de la CIE

Pour pallier le problème de multiplication des espaces couleurs utilisés, la CIE (Commission Internationale de l'Éclairage) a défini un certain nombre d'espaces standards. RVB fut standardisé en 1931 en se basant sur les travaux d'égalisation de couleurs pour un champ visuel d'angle 2° de WRIGHT et GUILD[WRIGHT29, GUILD31]. Dans le même temps, la CIE a établi le *système de référence colorimétrique* XYZ à partir des études de JUDD[JUDD30]. Ce nouveau système permet principalement de s'affranchir de deux contraintes du système RVB :

- Les trois composantes de RVB dépendent de la luminance qui n'est pas elle-même une composante ;
- Pour obtenir certaines couleurs, les composantes trichromatiques peuvent prendre des valeurs négatives (cf. figure [2.10]).

Les composantes de XYZ sont dites virtuelles car sans réalité physique. Malgré cela, les appareils de colorimétrie utilisent généralement cet espace pour la mesure des couleurs. La composante Y



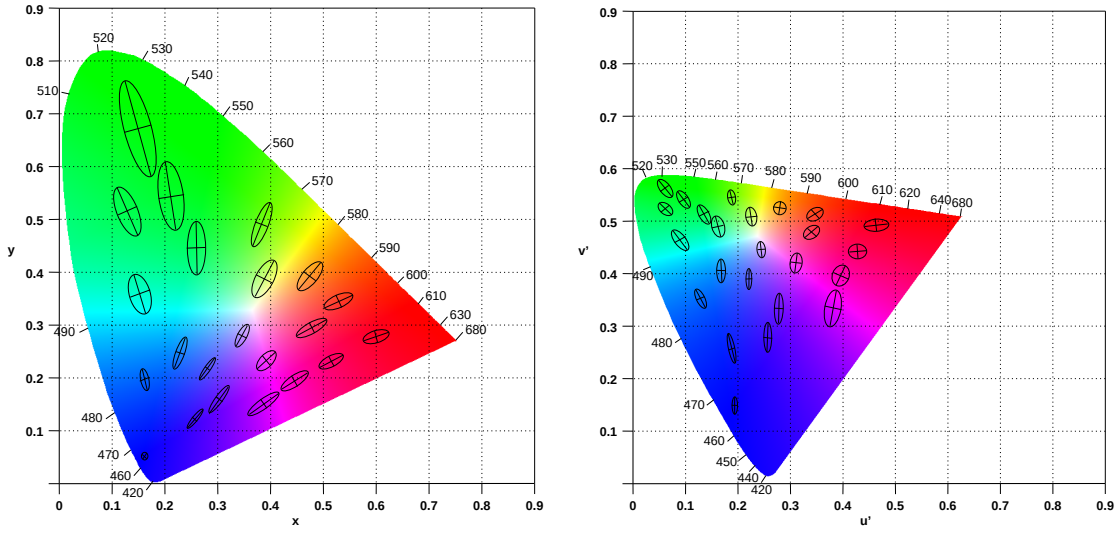
RVB *XYZ*

FIG. 2.10 – Fonctions d'égalisation de couleur pour les espaces CIE RVB et XYZ (1931)
Pour l'espace RVB, les valeurs négatives correspondent au fait que la couleur doit être soustraite à la couleur à égaliser.

peut être assimilée à la luminance car elle est égale à la fonction d'efficacité de l'œil. Il est à noter que la transformation de RVB à XYZ est linéaire et se définit donc simplement par une matrice de passage.

En 1964, la CIE adopte un nouveau système XYZ nommé $X_{10}\,Y_{10}\,Z_{10}$ pour lequel le champ visuel utilisé pour les expériences d'égalisation est de 10°. L'augmentation du champ visuel empêche les bâtonnets de fausser les résultats pour de faibles longueurs d'onde. La normalisation de ces espaces réalisée par l'opération $(x, y, z) = \frac{(X,Y,Z)}{X+Y+Z}$ permet d'obtenir $x + y + z = 1$. Il est alors possible de représenter les couleurs dans un plan ; en effet, connaissant x et y nous pouvons déduire z. La figure obtenue est appelée diagramme de chromaticité (x, y) (cf. figure [2.11].1931). Ce type de diagramme est très utilisé car il permet de déterminer de nombreux résultats par simple construction géométrique (couleur complémentaire, mélange de deux couleurs...).

Ces espaces que ce soit RVB ou XYZ sont perceptuellement non uniformes. Ainsi, une même différence d'apparence colorée ne correspondra pas forcément à la même distance dans l'espace des composantes mais dépendra de la zone où se trouvent les couleurs. Ce problème est très important en traitement d'images car les distances entre couleurs sont régulièrement utilisées (pour un calcul de gradient par exemple). Dans ce type d'espaces non uniformes, il est donc incorrect de comparer deux distances. Ce phénomène a été mis en évidence par MACADAM qui présente sur le diagramme de chromaticité (x, y) des zones où les couleurs ne sont pas discernables (cf. figure [2.11].1931)[MACADA42]. Ces zones ont des formes elliptiques et sont de tailles et d'orientation complètement différentes suivant la zone du diagramme où elles se trouvent. Pour résoudre ce problème d'uniformité, la CIE propose de 1960 à 1976 divers espaces : UVW en 1960, $U^*V^*W^*$ en 1964 pour aboutir à $L^*u^*v^*$ et $L^*a^*b^*$ en 1976. Les trois premiers sont obtenus par une simple transformation linéaire de XYZ ; en revanche, $L^*a^*b^*$ nécessite une transformation non linéaire. Ce dernier espace est d'ailleurs une adaptation d'un espace couramment utilisé à cette époque dans l'industrie : celui d'ADAMS-NICKERSON. Dans les espaces $L^*a^*b^*$ et $L^*u^*v^*$, il est possible d'utiliser une métrique euclidienne de manière « plus correcte » même s'ils ne sont toujours pas complètement uniformes et linéaires (cf. figure [2.11].1976). Ils sont tout de même assimilés à des espaces perceptuellement uniformes. Il est intéressant de noter que dans les deux cas L^* correspond



1931 - (x, y) 1976 - (u^*, v^*)

FIG. 2.11 – Diagrammes de chromaticité (x, y) et (u^*, v^*) avec leurs ellipses de MACADAM
Les ellipses tracées sont les ellipses de MACADAM *représentées à 10 fois leur taille réelle. Les couleurs du spectre se retrouvent en périphérie des diagrammes et sont représentées par leur longueur d'onde en* nm.

à la luminance, qu'une composante chromatique code l'opposition vert/rouge (a^* et u^*) et l'autre l'opposition bleu/jaune (b^* et v^*).

2.3.3 Espaces perceptuellement uniformes

Il apparaît sur la figure [2.11] que même les espaces recommandés par la CIE ne sont pas complètement uniformes et linéaires. Ainsi, de nombreux autres espaces plus ou moins uniformes ont vu le jour : $L_H a_H b_H$ de HUNTER en 1958, $L'a'b'$ de GLASSER en 1958, $Lu_f v_f$ de FARNSWORTH en 1957 ou encore PQS entre 1961 et 1971 pour ne citer que les premiers. L'espace de FARNSWORTH (cf. figure [2.12]) est d'ailleurs presque parfait mais comporte le gros désavantage de nécessiter une transformation non-linéaire très coûteuse à calculer.

Actuellement, aucun organisme ne préconise l'utilisation de tel ou tel espace car aucun n'est complètement perceptuellement uniforme ; le choix de l'espace est donc fortement lié à l'application visée. En fait, il a été démontré, en particulier par MACADAM [MACADA85], que seul un espace de dimension supérieure à six permettrait de résoudre le problème d'uniformité.

2.3.4 Espaces fondés sur le système visuel ou à la perception humaine

Trois types d'espaces sont distinguables dans cette partie : ceux liés aux trois types de cônes, ceux fondés sur le principe d'HERING d'antagonisme des couleurs et enfin ceux liés aux notions de teinte, luminance et saturation.

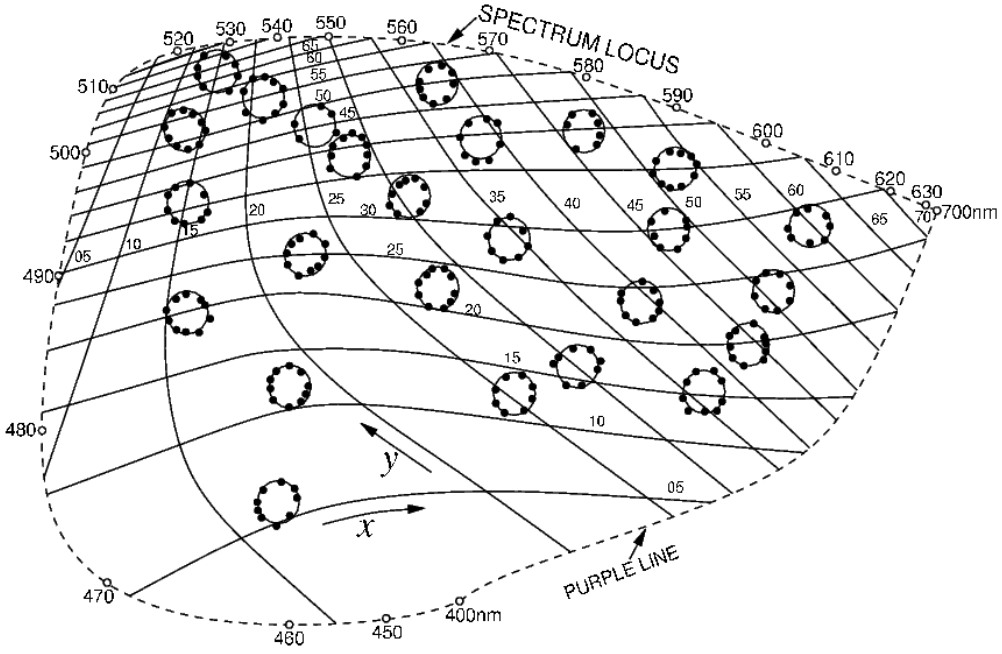


FIG. 2.12 – Espace couleur de FARNSWORTH obtenu par transformation non linéaire de XYZ

L'espace LMS définit les trois composantes colorimétriques comme proches des réponses spectrales des trois types de cônes. Cette opération s'effectue de manière linéaire à partir de XYZ. L'idée ici est de coder l'information couleur comme celle présente en sortie des photorécepteurs de l'œil.

D'autres systèmes vont plutôt se tourner vers un codage de l'information telle qu'elle est trouvée en sortie des différentes cellules de l'œil (horizontales et amacrines). Ils codent l'opposition des différentes couleurs et sont composés d'une composante achromatique (opposition noir/blanc) et de deux composantes chromatiques (opposition vert/rouge et bleu/jaune). De nombreux systèmes ont été proposés (par exemple, par FAUGERAS, GARBAY ou BALLARD). Les espaces $L^*u^*v^*$ et $L^*a^*b^*$ peuvent aussi être considérés comme faisant partie de cette catégorie.

Certains espaces utilisent aussi les notions de teinte, de luminance et de saturation pour définir les couleurs (cf. figure [2.13]). Ces trois notions sont celles généralement employées par l'homme pour décrire une couleur : « bleu foncé délavé », « rouge clair vif ». Habituellement, la première chose que nous notons au sujet d'une couleur est sa teinte qui décrit la nuance de couleur (rouge, vert, bleu...). Ensuite, celle-ci peut être agrémentée d'un qualificatif lié à la saturation correspondant à la pureté de la couleur par rapport au blanc ; un rouge complètement saturé sera un rouge sans ajout de blanc ; l'ajout de blanc rendra le rouge plus pastel et tendra vers le rose. Enfin, une couleur dispose aussi d'un éclat qui correspond à la luminosité émise par la couleur. De très nombreux espaces utilisent ces trois notions : ISH, IST, HSV, TLS, LCH... Les grandeurs utilisées représentent toujours le même aspect des couleurs mais se différencient par leur mode de calcul et de représentation (coordonnées polaires, cône hexagonal...).

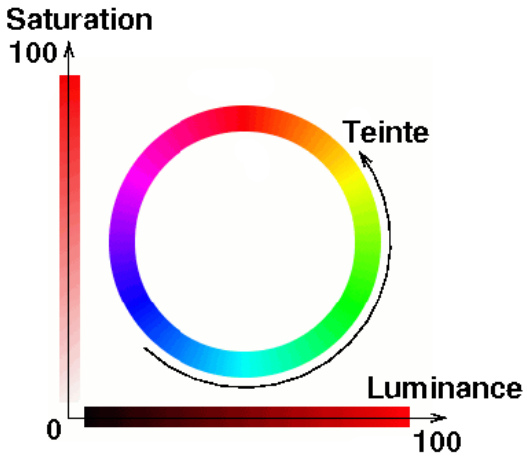


FIG. 2.13 – Les couleurs définies par les notions de teinte, de luminance et de saturation

2.3.5 Et les autres...

Nous sommes passés très rapidement sur la présentation des divers espaces couleurs généralement utilisés. En fait, il est très difficile d'être exhaustif sur ce point car il existe de très nombreux systèmes de représentation dont certains sont parfois uniquement utilisés pour une application spécifique.

Nous pouvons quand même remarquer ici que d'autres systèmes n'entrent pas dans notre classification et en particulier les systèmes d'axes indépendants fondés sur une analyse en composantes principales de l'espace couleur. Cette analyse permet de définir les composantes principales présentes dans les données ; elles peuvent ensuite être utilisées comme axes colorimétriques. OHTA effectue en 1980 cette étude sur huit images différentes. L'axe le plus discriminant (la composante principale) qu'il détermine de cette manière correspond à la luminance. Les deux autres composantes représentent les oppositions de couleur bleu/rouge et magenta/vert mais apportent moins d'information que la luminance. D'après son étude la troisième composante est réellement négligeable par rapport aux deux autres. En se basant sur cette analyse, il définit l'espace $I_1\,I_2\,I_3$ qui, de par la définition de ses composantes, peut être considéré comme un espace luminance/chrominance antagoniste.

Pour finir nous pouvons souligner que la CIE travaille actuellement sur la standardisation d'un espace de type LMS et surtout sur $CIECAM$ qui tente de prendre en compte des phénomènes tels que l'influence du voisinage dans la perception colorée ou la variabilité de la couleur de l'éclairage. De cette manière, il essaye de fournir un outil performant pour la comparaison des couleurs. La dernière version $CIECAM02$ gomme ainsi les trois effets indésirables de $CIECAM97$[LI02]. Récemment, FAIRCHILD propose également $Meet\ iCAM$ pour continuer à améliorer ce type d'espaces couleur[FAIRCH02]. Les voies à explorer au sein de ce travail sont principalement la prédiction des phénomènes liés à l'apparence colorée et spatiale des images.

2.4 Traitement et analyse d'images

Une fois l'image perçue par l'homme, son cerveau l'analyse en prenant en compte les diverses images précédemment observées. Le comportement du cerveau au niveau de cette fonction est encore un mystère ; son activité est intense mais encore obscure. En traitement d'images, le principe est plus clair ; l'image ou les images successives sont étudiées pour en extraire les informations utiles à l'application visée. Ces données peuvent ensuite être examinées pour prendre une décision (ex : rejet de pièce...) ou détecter des objets (ex : guidage autonome de robots...). Pour mettre en évidence le chemin qui a déjà été réalisé dans cette voie, un petit historique permet d'appréhender la jeunesse très active de cette discipline.

Les années 1920 peuvent être considérées comme l'origine du traitement numérique des images avec les premiers systèmes électroniques de visualisation et de transmission d'images. Même si divers systèmes mécaniques étaient déjà disponibles depuis quelques années pour la transmission d'images de télévision, VLADIMIR ZWORYKIN peut être considéré comme le père de la télévision moderne. En 1923, il invente l'iconoscope un tube analyseur permettant de diviser électroniquement une image en de nombreux éléments. Il expose alors en 1929 à Pittsburgh son système de télévision fondé sur l'iconoscope et un tube à rayons cathodiques pour la visualisation des images. Les années 1920 voit aussi la première transmission d'une image par câble entre New-York et Londres réalisé par PHILO FARNSWORTH à peine âgé de 21 ans en 1927 (un signe $ en quelques heures). Après cet enchaînement d'inventions et la naissance des premières chaînes et émissions de télévision (BBC en 1932), de nouveaux systèmes de télévision voient le jour. Il n'est pas encore possible de parler réellement de traitement ou d'analyse d'images mais plutôt d'acquisition, de transmission et de visualisation.

À partir des années 1950, certaines applications telles que les chambres à bulles nécessitent d'entrer dans une phase de traitement permettant l'amélioration de la qualité des images. Malgré tout, le traitement numérique des images prend son véritable essor dans les années 1960 avec l'apparition des systèmes informatiques utilisés pour les applications spatiales. En 1964, un des premiers traitement par ordinateur est utilisé par le *Jet Propulsion Laboratory* de la NASA pour l'étude des vidéos de *Ranger-7*. Les travaux ultérieurs se limitent aux projets spatiaux en raison des coûts (et de l'encombrement) des systèmes informatiques et des systèmes d'affichage. Ainsi, pour les missions *Surveyor* et *Mariner* les traitements effectués sont principalement de la restauration (correction des erreurs de transmission et d'acquisition), de la compression et de l'amélioration pour rendre la visualisation des images plus « jolie ». Pour marquer ces débuts, ROSENFELD écrit un des premiers livres portant sur le traitement numérique des images en 1969[ROSENF69].

Par la suite, les années 1970 sont le lieu de grandes avancées dans ce domaine grâce à la forte évolution du matériel informatique et à la baisse de son coût. Des nouvelles techniques telles que le seuillage, la segmentation ou la détection de contours rendent possible l'extraction automatique de certaines caractéristiques des images. Pour organiser ces diverses informations la notion de structure de description apparaît. La puissance des ordinateurs ne cessant de croître et leur miniaturisation de continuer, l'interprétation des images devient un enjeu majeur. L'engouement pour les systèmes experts en est un exemple intéressant ; la déception est à la hauteur de l'enthousiasme d'alors. L'échec repose principalement sur le fait que la connaissance est généralement trop complexe pour être modélisée simplement.

Les années 1980 sont marquées par l'expansion des systèmes de vision dans l'industrie grâce à la démocratisation de la micro-informatique et des capteurs. Le milieu des années 1980 voit également l'arrivée des stations graphiques facilitant l'accès aux modèles tridimensionnels. Le niveau de miniaturisation est tel que la vision robotique devient accessible. De nouveaux domaines de recherche s'ouvrent alors : analyse du mouvement 2D et 3D, reconnaissance d'objets et d'environnements...

En un peu plus de 30 ans, les fondations de la vision par ordinateur ont donc été érigées. Un grand nombre de problèmes ont pu être identifiés mais ils sont encore, pour la majorité, non entièrement résolus. Globalement, il apparaît que les questions simples ont trouvé une solution même partielle mais celles nécessitant une interprétation poussée de l'image posent encore de nombreuses difficultés. La segmentation d'une image en objets (au sens humain), par exemple, est encore un problème non résolu ; des techniques apportent de bons résultats pour des applications spécifiques mais la notion d'entité visuelle pour l'homme est encore trop complexe et variée. La nécessité d'interprétation du contenu peut d'ailleurs marquer la limite entre traitement et analyse d'images. L'échec relatif des systèmes experts montre bien qu'il paraît impossible actuellement de pouvoir modéliser la connaissance et l'expérience d'un être humain pour des problèmes complexes. En fait, l'analyse d'images est plutôt une tentative de reproduction ou d'imitation du comportement humain pour obtenir un résultat similaire. Cette imitation est réalisable à l'heure actuelle pour des problèmes simples ne nécessitant pas ou peu l'utilisation combinée d'expérience et de connaissances.

2.5 Conclusion

Ce chapitre a donc introduit rapidement les diverses analogies réalisables entre vision humaine et artificielle. De la perception à l'analyse, le système visuel humain est encore grandement méconnu. Malgré cela, l'imagerie couleur tente de mettre en place des systèmes artificiels imitant le comportement humain. Les systèmes d'acquisition sont fortement inspirés du fonctionnement de l'œil. Les données obtenues par de tels outils doivent alors être représentées numériquement pour être utilisables à des fins de restitution, de traitement ou d'analyse.

Il apparaît aussi que le domaine du traitement et de l'analyse d'images est très jeune. De très nombreux problèmes sont encore loin d'être résolus et en particulier ceux liés à l'interprétation des images. Actuellement, les applications très en vogue avec la généralisation des bibliothèques numériques sont celles liées à l'accès aux données : indexation, classification et recherche d'éléments dans de vastes bases d'images. Le chapitre suivant se propose de présenter rapidement les différentes solutions apportées à ces problèmes. Il ressort de cette étude que seule une interprétation plus fine des images va permettre de réellement progresser dans cette discipline. Il faut donc aller vers une modélisation complexe sans pour autant vouloir aboutir à celle de l'homme.

CHAPITRE 3

INDEXATION ET RECHERCHE D'IMAGES

Sommaire

3.1 Introduction

Actuellement les moteurs de recherche sur le Web fournissent de très bons outils pour trouver une information dans l'ensemble des pages qu'ils recensent (qu'ils indexent). De tels systèmes pour les images apparaissent petit à petit mais ils sont pour la plupart basés sur une description textuelle de l'image. Dans ces conditions, le problème se borne toujours à la recherche de motifs alphanumériques. *Google* propose, par exemple, son moteur de recherche de pages Web[1] et son moteur de recherche d'images[2]. Ce dernier caractérise une image par son nom de fichier et par la zone de texte l'entourant sur la page Web dont elle est tirée.

Aujourd'hui, la recherche d'images ressemblantes dans des bases de données ne se limite plus à l'indexation textuelle car cette approche est assez contraignante. En effet, la description textuelle d'une image demande une intervention humaine ; et celle-ci a deux limites principales : le temps et le jugement sans parler des limitations liées à la langue (ou au jargon) utilisée pour l'indexation. D'une part, présenter le contenu d'une image demande beaucoup de temps (« voiture bleue dans

[1] http://www.google.fr/.
[2] http://images.google.com/.

la forêt » ou « plage de sable et mer bleue turquoise » par exemple). Pour des ensembles d'une centaine d'images cela est encore imaginable mais actuellement les volumes étudiés sont plutôt de l'ordre du millier voire de la centaine de milliers d'images. Une intervention humaine pour décrire de tels ensembles semble donc impossible.
D'autre part, la présentation qui est faite d'une image peut varier énormément suivant la personne. Les différences peuvent porter sur :

- la langue utilisée (ou le jargon) : anglais, français, allemand... ;
- le vocabulaire utilisé : une Porsche sera décrite par « Porsche 911 noire » par un connaisseur mais par « voiture de sport noire » dans la plupart des cas ;
- le niveau de description : « maison » et « maison avec une cheminée, un toit de chaume et six fenêtres aux volets verts » permettent de désigner la même maison mais à des niveaux de détails différents.

Par conséquent, ne pouvant pas se baser entièrement sur une description textuelle, la recherche d'images tente aujourd'hui d'extraire les informations directement des images. Ce type de travaux se regroupent sous le terme : indexation d'images basée sur le contenu (ou *CBIR* en anglais pour *Content Based Image Retrieval*). Dans ce domaine, les images sont décrites grâce à leur propre contenu. Elles sont analysées afin d'en extraire les éléments permettant de les caractériser le plus fidèlement possible.

Ce chapitre se propose de présenter, tout d'abord, les différentes questions qui se posent en indexation d'images. Elles amènent ensuite à trouver un système de représentation d'images adapté au problème. Nous exposerons alors un rapide aperçu des différentes techniques les plus employées en recherche d'images par le contenu : tout d'abord dans un schéma où l'image est décrite dans sa globalité puis en prenant en compte l'organisation spatiale des objets au sein de l'image. Enfin, nous présenterons les diverses méthodes utilisées pour quantifier la qualité de tels systèmes de recherche.

3.2 Représenter pour indexer et rechercher

3.2.1 Modes de recherche possibles

Globalement, l'indexation d'images par le contenu a pour but de retrouver des images d'un contenu bien précis. Tout l'enjeu est que l'information recherchée peut être de nature très diverse. La liste ci-dessous donne un panel des problèmes qui peuvent être posés :

Description simple : l'utilisateur désire retrouver des images à partir d'un critère simple et global à l'image. Pour trouver des images de plages, la demande pourra être formulée en terme de couleurs : « Trouver les images contenant 50% de bleu, 30% de vert et 20% de jaune ».

Description composée simple : les images sont recherchées suivant de multiples critères combinés. L'obtention d'images de forêt pourra être réalisée par : « Trouver les images à dominante verte et contenant une texture semblable à celle d'un arbre. Le critère de couleur étant 3 fois plus important que l'autre ».

Description localisée : l'utilisateur spécifie la position des critères qu'il propose. La demande « Trouver les images contenant du bleu en haut et du jaune en bas » pourra être définie pour rechercher des images de plages.

Recherche par l'exemple : dans ce cas, une image est utilisée comme référence pour effectuer la recherche. Ainsi, les demandes effectuées seront fonction des caractéristiques de l'image de référence voire même fonction d'une partie de celle-ci. Par exemple, il est possible de rechercher les images contenant des arbres en sélectionnant un arbre dans une image et en demandant « les images de composition colorimétrique et de texture semblables à celle de cet arbre ».

Recherche par esquisse : la recherche s'effectue de la même manière que dans le cas précédent mais la référence est un dessin réalisé par l'utilisateur.

Description d'un objet : un problème plus délicat est la recherche d'objets au sein des images. Dans ce cas, l'utilisateur peut vouloir obtenir « les voitures rouges positionnées au centre de l'image ». Il spécifie alors l'objet qu'il désire obtenir en le sélectionnant dans une image ou en l'esquissant et en définissant des critères sur cet objet et/ou sur son positionnement : « forme de voiture », « couleur rouge » et « au centre de l'image » pour l'exemple précédent. Ce type de demandes est appelé *requête partielle*.

Description d'un ensemble d'objets : l'évolution de la requête précédente est de vouloir obtenir un ensemble d'objets organisés d'une certaine façon. Par exemple : « Trouver les images d'une route sur laquelle une voiture rouge suit une voiture bleue ».

Il apparaît que les problèmes posés sont très variés et que les solutions apportées peuvent être très diverses suivant les contextes, en particulier le type de bases (généraliste ou spécialisée) et le type d'utilisateurs (expert du domaine ou non).

3.2.2 Nécessité de représenter

3.2.2.1 À quel niveau décrire ?

L'homme perçoit une scène et l'analyse grâce à son expérience et à ses connaissances. Un système d'acquisition numérique fournit une image se composant de centaines de milliers voire de millions de pixels colorés. Ils forment une énorme matrice contenant les différentes couleurs. Le but de l'imagerie couleur dans son ensemble est de donner un sens à cette énorme masse de données, comme le fait l'homme de manière intuitive en s'appuyant sur un énorme apprentissage. Pour cela, il est nécessaire d'extraire de l'image des caractéristiques permettant d'une part de réduire le volume à traiter et d'autre part de pouvoir donner un sens à ses données. Le contenu visuel d'une image peut ainsi être modélisé à différents niveaux d'abstraction, du niveau physique jusqu'à la représentation humaine (cf. figure [3.1]).

1. Du point de vue le plus fin, l'image est décrite comme un arrangement de pixels ; chacun portant une information sur sa couleur (ou sa luminance dans le cas d'une image en niveaux de gris). Sur l'exemple proposé, l'image est composée de pixels de diverses couleurs formant l'image d'une maison.
2. Des traitements appropriés peuvent permettre d'extraire des caractéristiques géométriques de l'image. Ce deuxième niveau peut, par exemple, être composé de contours, de lignes, de points d'intérêt, de régions... Cette étape tente d'imiter le système visuel humain qui extrait les éléments caractéristiques d'une scène et sépare les différents objets qui la compose. Les frontières entre objets peuvent ainsi être détectées pour obtenir les bords de la maison, du nuage, du toit et de la porte.

3. Ces éléments peuvent ensuite être combinés et interprétés comme des objets ayant chacun des propriétés propres (de couleur, de texture...). Les régions issues de l'étape précédente peuvent être caractérisées par leur couleur : le toit est jaune, le ciel bleu...
4. Pour l'instant, l'image est vue comme un ensemble d'objets indépendants. En prenant en compte les interactions entre objets, l'image peut alors être considérée comme un arrangement d'objets. La porte est entourée par les murs qui se trouvent eux-mêmes sous le toit...
5. Le premier niveau d'interprétation humain permet ensuite de prendre en compte la composition des objets. Les régions obtenues sont regroupées sous forme d'objets qui sont composés et en relation les uns avec les autres. La maison est ainsi constituée du toit, des murs et de la porte ; elle est adjacente avec le ciel. Les régions de la maison sont également organisées.
6. Le niveau de conceptualisation humain complet considère l'image comme un ensemble d'objets nommés et composés en relation les uns avec les autres. C'est une description sémantique totale de l'image. Un objet se décompose en sous-objets qui sont disposés suivant un arrangement spatial. De plus, chaque objet est décrit non plus par des caractéristiques mais par son nom, ou plus précisément par son concept : les murs de la maison pourraient être nommés « murs » ou bien « façade », où comme des « éléments verticaux délimitant l'intérieur de l'extérieur de la maison ». Le ciel n'est plus considéré comme une zone bleue mais comme une entité du concept « ciel ». L'image exemple se décompose en deux : le ciel et la maison. La maison est constituée du toit, des murs et de la porte...

Ces différents niveaux apportent donc une description plus ou moins fine et complexe de l'image. Suivant les problèmes (cf. section 3.2.1) auxquels il est nécessaire de répondre, il faudra aller plus ou moins loin dans le niveau de représentation pour comparer les images.

3.2.2.2 Mythe de la représentation totale

Idéalement, la description des images serait réalisée au niveau de conceptualisation humain. La composition sémantique d'une image serait extraite grâce à son contenu. L'image figure [3.2] pourrait, par exemple, être perçue comme « une scène d'extérieur à la campagne où une personne habillée en rouge se promène avec ses deux chiens au milieu d'arbres de divers types ». Cette description, même si elle est encore incomplète et/ou incorrecte, apparaît irréaliste d'un point de vue informatique car c'est l'interprétation humaine qui permet d'atteindre un tel niveau de détail. Dire que la personne se promène avec ses deux chiens n'est pas évident à première vue. Tout d'abord, la notion de promenade est déduite de l'atmosphère générale de l'image. Ensuite, les deux chiens sont quasiment invisibles sur l'image et c'est toujours par déduction qu'ils sont identifiés.

Globalement, l'informatique ne peut fournir aujourd'hui un système doué des qualités d'interprétation valant celles de l'homme. Son interprétation sera toujours bornée au domaine qui a été programmé. Il est possible de faire acquérir à un système artificiel de l'expérience et des connaissances très diverses mais il apparaît très difficile de mixer les deux pour obtenir une machine réellement intelligente. C'est pourquoi, l'indexation d'images se limite principalement à une description statistique des images. Des élèments sont extraits des images pour les caractériser mais ils ne sont pas interprétées pour remonter à une description sémantique des images. Il existe tout de même certains travaux qui essayent d'identifier des objets à partir de leurs caractéristiques[DUYGUL02] ; ceux-ci sont pour l'instant très limités que ce soit du point de vue du nombre d'objets caractérisés (i.e. du volume du vocabulaire pris en compte) ou de l'utilisation de l'environnement dans lequel

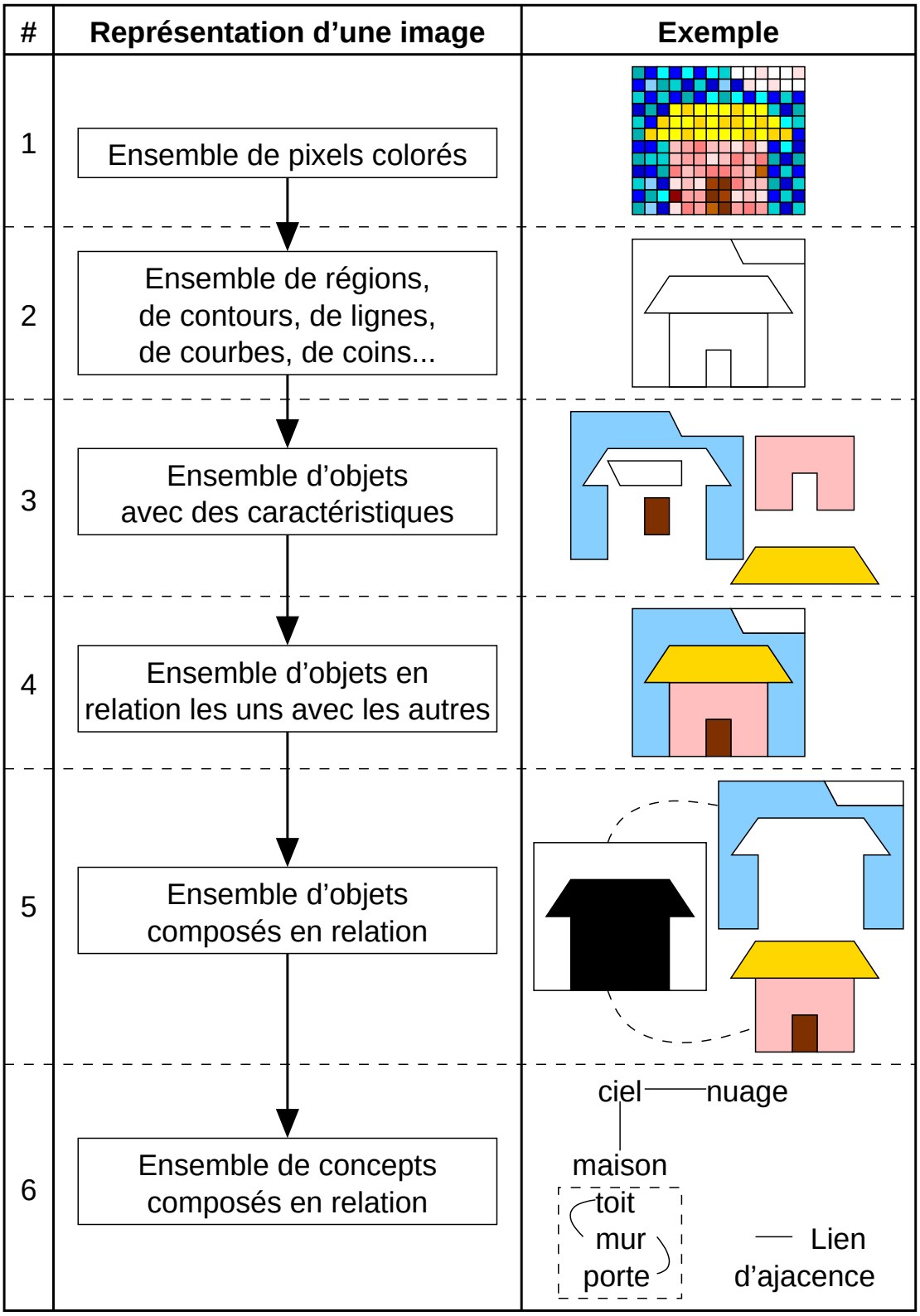


FIG. 3.1 – Niveaux d'abstraction auxquels une image peut être vue

FIG. 3.2 – Image issue de la base de l'université de Washington (UW-groundtruth)

ils sont plongés. De plus, il paraît très difficile dans ces approches de différencier statistiquement deux objets différents mais visuellement proches (un chat et un tigre par exemple).

Les premières approches développées en indexation décrivaient les images dans leur ensemble. Des statistiques simples sont alors calculées pour représenter les images. Généralement, les informations prises en compte sont la couleur, la texture et/ou la forme. Le troisième niveau d'abstraction est ainsi atteint même si l'image est considérée comme un objet unique et non comme un ensemble d'objets. Malheureusement, ces techniques décrivent les images globalement. En utilisant la couleur moyenne de l'image, celle-ci pourra être présentée à dominante bleue. En allant plus loin avec les histogrammes couleurs (cf. section 6.2.2), elle pourra être vue comme composée de 50% de bleu, 30% de vert et 20% d'orange. Mais, il paraît difficile avec ce type d'approche d'obtenir une description plus fine de l'image et notamment de rechercher des objets.

Depuis quelques années, des approches dites spatiales essayent de fournir une description plus précise en considérant l'image comme composée d'un ensemble d'objets (de régions). La représentation de l'image est alors portée par l'ensemble des descriptions des éléments la composant mais aussi par les relations existant entre eux. La conceptualisation de l'image a ainsi gagné un niveau.

Enfin, de récents travaux (en partie ceux de cette thèse et [XU00]), tentent d'aller encore à un niveau de plus dans la description. L'image est considérée à plusieurs échelles ce qui permet de la décomposer tout d'abord en objets, puis en parties d'objets et ainsi de suite. La caractérisation est alors portée par l'ensemble des régions obtenues, de leur relations de voisinage et d'inclusion

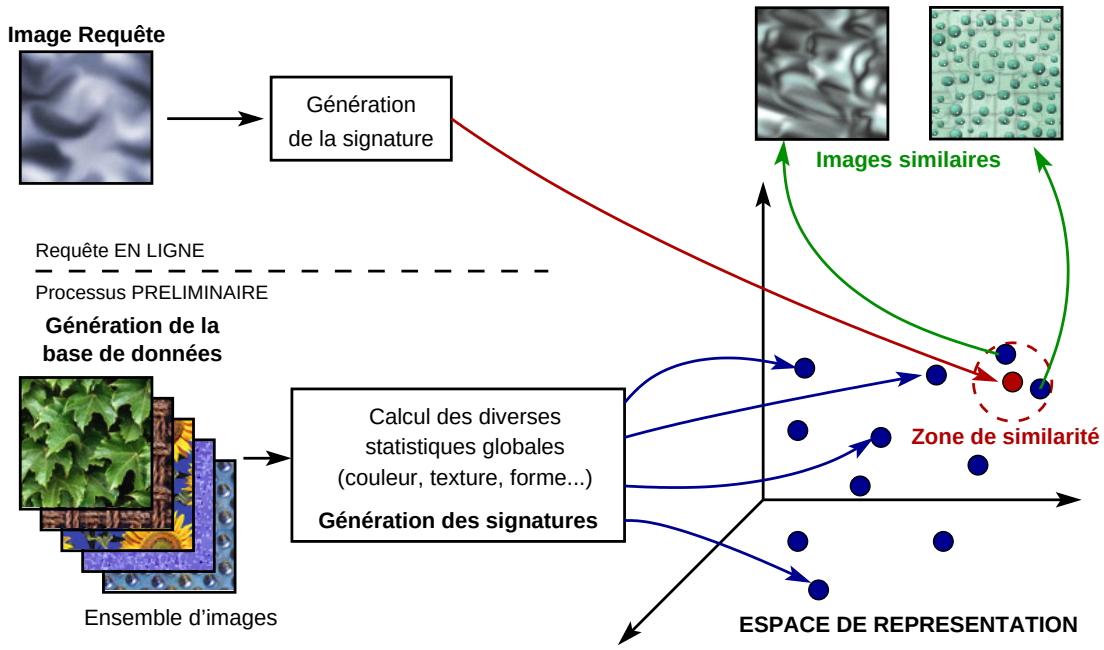


FIG. 3.3 – Schéma d'indexation classique

(cf. section 4.4). Sans remonter à la sémantique même des images, ces structures tentent de les décrire le plus fidèlement du point de vue de la perception humaine.

Nous détaillons maintenant un peu plus l'approche globale puis l'approche spatiale.

3.3 Approche globale

Cette approche considère l'image dans son ensemble et la caractérise en utilisant des statistiques calculées sur l'image entière.

Les différentes approches globales portent généralement sur trois critères : la couleur, la texture et la forme. Pour chacun d'eux, un grand nombre d'attributs ont été développés. Ils peuvent ensuite être combinés pour obtenir un descripteur complet et robuste. Cet ensemble de statistiques est généralement appelé *signature* de l'image.

Ensuite, dans le cadre d'une recherche par l'exemple, la démarche est généralement la suivante (cf. figure [3.3]) :

- **Phase préliminaire :** calculer les descripteurs de chaque image de la base de données ;
- **Phase en ligne :** calculer les descripteurs de l'image requête ;
- **Phase de recherche :** rechercher les images proches de l'image requête dans l'espace du (des) descripteur(s) utilisé(s).

Dans cette partie, les trois principaux types de statistiques sont présentés rapidement. Nous ne voulons donner ici qu'un aperçu des techniques principales utilisées ; de plus amples détails seront proposés au chapitre 6 portant sur la description des régions dans une approche spatiale.

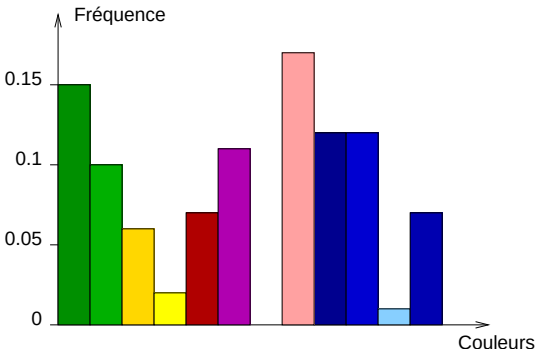


FIG. 3.4 – Un histogramme couleur

3.3.1 Aspect couleur

La couleur est sûrement le critère le plus important psychovisuellement parlant dans la vision d'une image. Le premier regard porté sur une image permet d'appréhender très rapidement sa couleur dominante et ses différentes teintes.

La distribution des couleurs au sein d'une image est très souvent représentée sous la forme d'un histogramme couleur. Celui-ci associe à chaque couleur sa fréquence d'apparition (cf. figure [3.4]). Ainsi, pour une image définie par $I : x \in \Omega \rightarrow I(x) \in \Theta$ avec Ω le domaine de l'image et Θ celui des couleurs ; la fonction histogramme de l'image I peut s'écrire $H_I : c \in \zeta_I \rightarrow \mathbb{R}$ où $\zeta_I = \{I(x) \,/\, x \in \Omega\}$ est l'ensemble des couleurs de l'image (ce qui entraîne que $nbc_I = \text{Card}(\zeta_I)$ est le nombre de couleurs présentes au sein de l'image I). Les valeurs de l'histogramme sont alors :

$$H_I(c) = \frac{\text{Card}(\{x \,/\, x \in \Omega,\ I(x) = c\})}{\text{Card}(\Omega)}, \forall c \in \zeta_I.$$

Il est à noter ici qu'avec cette définition l'histogramme obtenu est normalisé, $\sum_{c \in \zeta_I} H_I(c) = 1$. De plus, le choix des couleurs représentatives est très important et est un problème très complexe (cf. section 6.2.1).

SWAIN et BALLARD ont les premiers proposé une méthode d'intersection d'histogrammes permettant de comparer des images par leur distribution colorimétrique[SWAIN91]. Considérant deux images I_1 et I_2 et leurs histogrammes respectifs H_1 et H_2 définis sur le même domaine couleur ζ (cette hypothèse est assez contraignante car les images doivent utiliser la même palette de couleurs), la distance proposée est :

$$D_{SwainBallard}(H_1, H_2) = \frac{\sum_{c \in \zeta} min(H_1(c), H_2(c))}{\sum_{c \in \zeta} H_1(c)}$$

Il est à noter ici que pour une utilisation correcte de cette distance, les deux histogrammes ne doivent pas être normalisés. De nombreuses autres distances entre histogrammes ont ensuite été proposées pour des images à supports colorimétriques identiques ou non. Ces différentes approches seront présentées plus longuement dans le chapitre 6.2 de même que les moments colorimétriques,

les *color sets* ou le vecteur de cohérence couleur qui sont trois autres méthodes de représentation de l'information couleur.

3.3.2 Aspect forme

Une fois l'analyse colorimétrique effectuée, le cerveau a décomposé la scène en plusieurs régions. La forme peut, par la suite, être utile pour les décrire. Elle est d'ailleurs très utilisée en vision industrielle et notamment pour du tri de pièces ou de la vérification de placement de composants électroniques. En prenant l'image *peppers* (cf. figure [3.5]), les deux poivrons sont colorimétriquement très proches alors que leur forme est très différente ; c'est donc ce critère qui doit être utilisé pour les distinguer.

La reconnaissance des formes a débuté en appliquant des méthodes issues de la caractérisation de variables aléatoires et en particulier les moments. Le moment d'ordre $p+q$ sont définis par :

$$m_{pq}(I) = \sum_{(x,y)\in\Omega} x^p y^q I(x,y),\ \forall (p,q) \in \mathbb{N}^2$$

où I est l'image au sein de laquelle la forme à étudier a été isolée. Ainsi, les moments d'ordre 1 normalisés (par m_{00}), $\frac{m_{01}}{m_{00}} = \bar{y}$ et $\frac{m_{10}}{m_{00}} = \bar{x}$, correspondent à la position du centre de gravité de la forme. Les moments centrés peuvent alors être définis en utilisant $\bar{x}$ et $\bar{y}$:

$$\mu_{pq}(I) = \sum_{(x,y)\in\Omega} (x-\bar{x})^p (y-\bar{y})^q I(x,y),\ \forall (p,q) \in \mathbb{N}^2$$

Les différentes valeurs de μ_{pq} permettent de caractériser une forme ; μ_{01} mesure par exemple l'allongement suivant y. Les formes de deux objets sont alors comparables simplement grâce à leurs moments respectifs.

HU propose en 1962 un ensemble de combinaisons des moments centrés définissant 7 nouveaux moments invariants par translation, rotation et changement d'échelles[HU62].

De nombreuses autres techniques ont aussi été mises en place pour caractériser une forme et la différentier d'une autre : descripteurs de FOURIER, moments de ZERNIKE, représentation *CSS* (pour *Curvature Scale Space*)... Plus d'attention sera portée à ces méthodes dans le chapitre 6.3.

3.3.3 Aspect texture

La texture est un concept très lié à la perception mais assez difficile à caractériser du point de vue de la vision. Cette caractéristique est plutôt associée au toucher ou au goût. La notion de rugosité est ainsi difficile à analyser sur une image 2D. Par contre, le contraste paraît un peu plus naturel. La notion de texture peut sommairement être assimilée à la description du (des) motif(s) formé(s) par les différentes couleurs : uniforme, alterné une ligne sur deux... La figure [3.6][3] propose quelques exemples de textures. Leur point commun est une information visuelle fortement redondante à l'intérieur d'une fenêtre de taille minimale.

Dans ce cadre, HARALICK propose d'utiliser la matrice de cooccurrences[HARALI73] (cf. figure [3.7]). Celle-ci contient les informations de voisinage entre les différentes couleurs composant

[3] VISTEX est téléchargeable sur http://www-white.media.mit.edu/vismod/imagery/VisionTexture/.

FIG. 3.5 – Deux formes présentes dans l'image *peppers*

FIG. 3.6 – Quatre textures issues de la base VISTEX

l'image. Son utilisation est motivée par la conjecture de JULESZ selon laquelle l'œil humain utilise une statistique d'ordre 2 pour discriminer deux textures et qu'il est très difficile de différentier deux textures dont les statistiques sont égales jusqu'à l'ordre 2[JULESZ75] (en 1978, JULESZ nuance cette affirmation car certains cas exceptionnels la contredisent à l'ordre 2 et même 3).

Plus précisément, la fonction de cooccurrences au sein d'une image I est définie telle que $MC_I : \zeta_I^2 \rightarrow \mathbb{R}$. À un couple de couleurs (c_1, c_2) est associé la probabilité qu'elles soient voisines. Il faut donc définir une fonction de voisinage : $V : \Omega^2 \rightarrow \mathbb{B}$ qui définit si $(p_1, p_2) \in \Omega^2$ sont voisins suivant le critère choisi : 4-connexité, 8-connexité ou autre (pour une direction et une longueur donnée par exemple).
Ainsi, $\forall (c_1, c_2) \in \zeta_I^2$:

$$MC_I(c_1, c_2) = \frac{\text{Card}(\{(p_1, p_2) \ / \ (p_1, p_2) \in \Omega^2,\ I(p_1) = c_1,\ I(p_2) = c_2,\ V(p_1, p_2) = vrai\})}{\text{Card}(\{(p_1, p_2) \ / \ (p_1, p_2) \in \Omega^2,\ V(p_1, p_2) = vrai\}}$$

La matrice de cooccurrences est égale à la matrice carrée de taille $nbc_I = \text{Card}(\zeta_I)$:

$$[MC_I] = [MC_I(c_1, c_2)]$$

Si la notion de voisinage est symétrique alors MC_I l'est aussi.

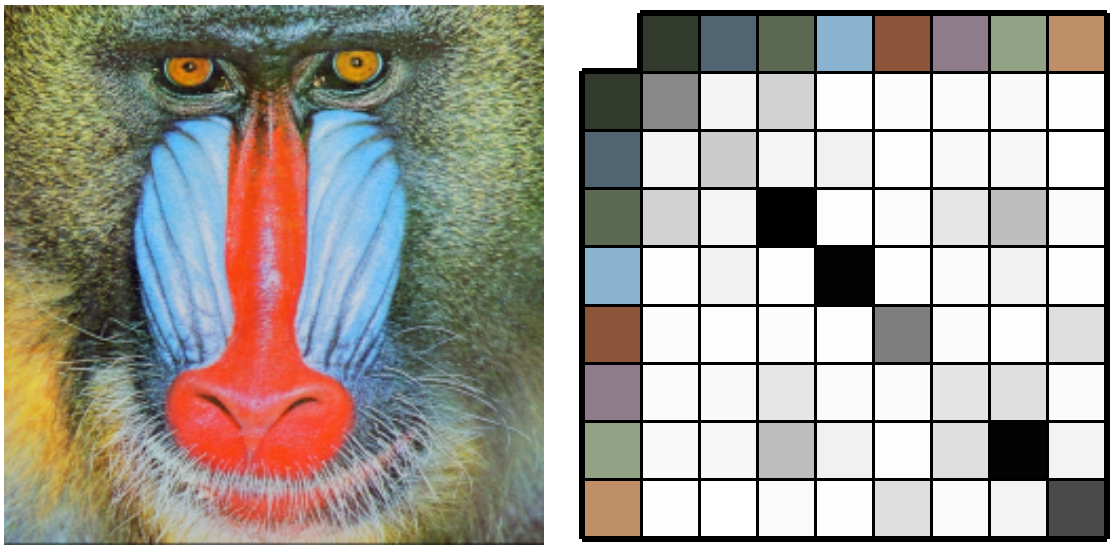


FIG. 3.7 – Matrice de cooccurrences de l'image *baboon* réduite en 8 couleurs
Les associations en noire sont les plus fréquentes et celles en blanc inexistantes.
La fonction de voisinage utilisée ici est celle correspondant à la 4-connexité.

À partir de celle-ci, HARALICK définit 14 mesures pour caractériser une texture telles que le contraste, la corrélation, la variance...

D'autres statistiques ont été mises en place à partir de cette même matrice de cooccurrences mais aussi à partir des coefficients d'ondelettes ou de GABOR ou encore avec la définition de matrices de longueurs de plages. Plus d'attention sera portée à ces méthodes dans le chapitre 6.3

3.4 Approche spatiale

L'idée pour cette approche est de considérer l'image comme un ensemble d'objets et non plus comme une entité unique. Une fois l'image segmentée en plusieurs régions, ces dernières peuvent être caractérisées de la même manière que les images dans la partie précédente. De cette façon, des détails plus fins peuvent être retrouvés au sein des images car les statistiques sont localisées ce qui induit une diminution du problème de masquage d'une donnée par une autre (la composition d'un objet ne sera plus noyée dans celle de l'image mais sera isolée). De plus, il est possible de rechercher des arrangements spécifiques d'objets en utilisant l'information liée à l'organisation des régions dans l'image.

Ensuite dans le cadre d'une recherche par l'exemple, la démarche est généralement la suivante (cf. figure [3.8]) :

- **Phase préliminaire :** segmentation des images de la base de données et calcul des descripteurs pour chaque région pour obtenir les représentations des images ;
- **Phase en ligne :** segmentation de l'image requête et calcul des descripteurs des régions pour obtenir la représentation de l'image. Les régions à utiliser pour la recherche peuvent aussi être sélectionnées ;

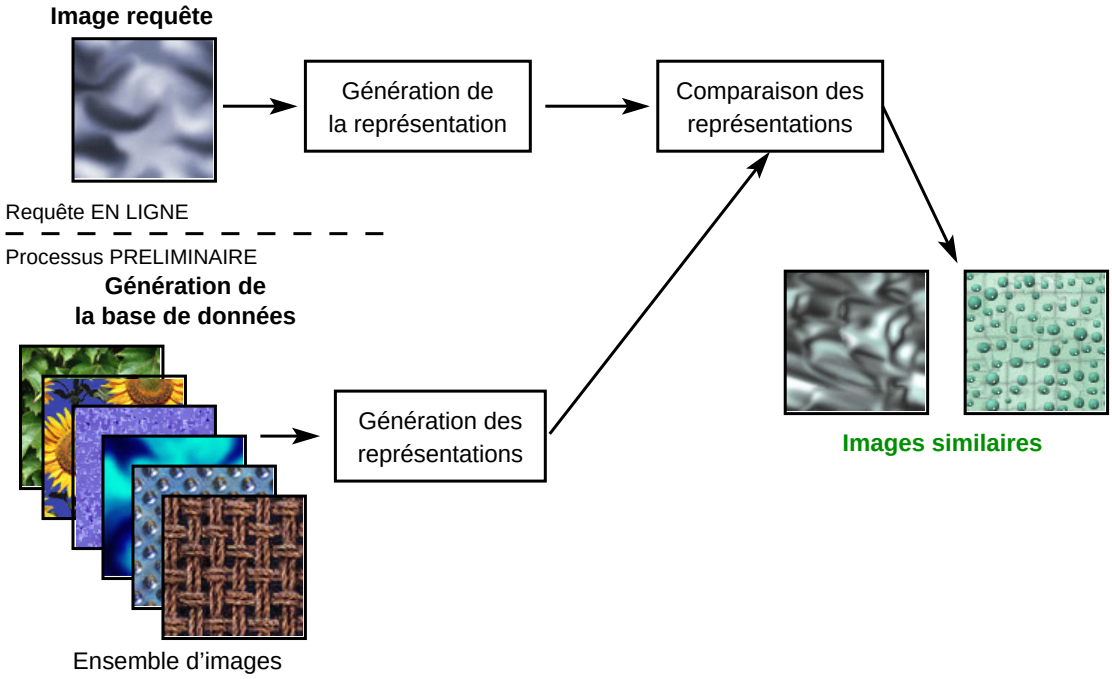
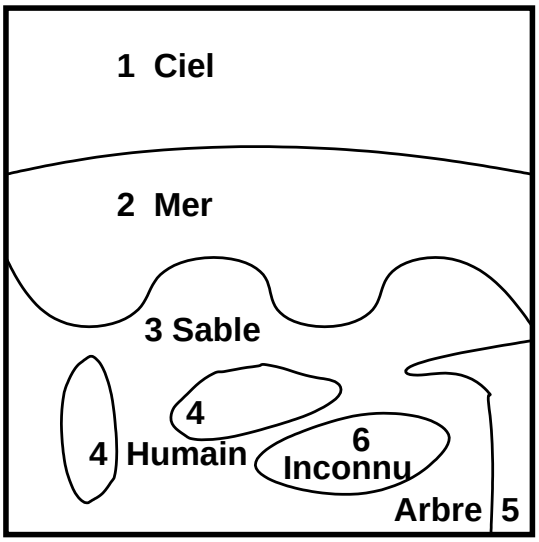


FIG. 3.8 – Schéma d'indexation spatiale

– **Phase de recherche :** rechercher les images proches de l'image requête en comparant les descriptions.

La problématique de l'indexation dite *spatiale* prend naissance dans la fin des années 1980 et se développe rapidement à la fin du siècle dernier. Dans ce type de méthodes nous ne considérons pas celles qui concernent les requêtes partielles, c'est-à-dire celles qui n'effectuent qu'une recherche de régions. En effet, dans le cadre des requêtes partielles, une fois les images segmentées et les régions décrites, le système de recherche est exactement le même que celui de l'approche globale mais les éléments recherchés ne sont pas les images entières mais certaines régions particulières [MA99, CARSON02]. Nous considérons donc les systèmes d'indexation *spatiale* comme ceux prenant en compte l'organisation des différents objets au sein des images pour effectuer la recherche.

CHANG propose en 1987 une méthode de représentation d'images fondée sur les *2-D strings* [CHANG87]. Les *2-D strings* apportent une représentation efficace du contenu d'une image et permettent de réduire la complexité de la comparaison (polynomiale dans ce cas). Ils supposent que chaque objet est identifié avant le stockage, un identifiant unique devant leur être affecté. Les positions relatives entre les objets sont alors représentées par deux chaînes de caractères monodimensionnelles. L'idée est de projeter la position des objets (i.e. leurs centres de gravité) sur les deux axes de l'image et de les prendre dans le même ordre que leurs projections (cf. figure [3.9]). Le problème de recherche d'images est alors transformé en une comparaison de chaînes de caractères. Pour accélérer la recherche, de nombreuses méthodes d'indexation des *2-D strings* ont été proposées[CHANG91, PETRAK93]. Pour gérer les problèmes de formes complexes et de recouvrement, les *2-D C-strings*[LEE90] et les *2-D G-strings*[CHANG89] ont vu le jour. En contrepartie de l'information supplémentaire apportée, leur appariement est plus complexe (exponentiel pour les *2-D C-strings*). Au lieu d'utiliser les projections pour ordonner les régions, il est également pos-



Ordre	*2-D string*
Horizontal	(4 4 2 3 1 6 5)
Vertical	(1 2 4 3 4 6 5)
Aire	(4 4 6 5 1 2 3)
Périmètre	(4 4 5 6 1 3 2)

À gauche l'image d'exemple et dans le tableau ci-dessus des exemples de *2-D strings* qui peuvent être utilisés. Les régions de l'image ont été au préalable segmentées et identifiées.

FIG. 3.9 – Exemples de *2-D strings*

sible d'utiliser des caractéristiques autres que spatiales telles que l'aire ou le périmètre des régions. De plus, la définition de nouvelles relations spatiales a permis à PETRAKIS de définir les *expanded 2-D strings* dont la comparaison est polynomiale[PETRAK96]. Ce type de représentation pose trois problèmes principaux :

- l'existence d'erreurs lors de la recherche (fausses alarmes, ou non retour de réponses évidentes) dues au caractère strict de la recherche de chaînes. Ce problème peut être évité partiellement par l'utilisation de techniques de recherche inexacte[PETRAK96] qui sont bien évidemment d'une complexité plus importante ;
- le contenu de l'image doit pouvoir être identifié. Cela peut être réalisé pour des problèmes ciblés (structures du corps humain[PETRAK93]...) mais difficile dans un cadre général ;
- une telle représentation est invariante par translation et changement d'échelles mais pas par rotation.

Pour obtenir encore plus de souplesse dans la représentation spatiale des images et des régions les composant, des structures telles que le graphe d'adjacences de régions ont été employées à partir de la moitié des années 1990. Dans ce cadre, les régions ne sont plus nécessairement identifiées et peuvent être décrites par les diverses statistiques utilisées en indexation globale ; les arrangements spatiaux décrits peuvent ainsi être plus complexes qu'un simple ordonnancement vertical et horizontal.

Une première étape vers un tel système a été proposé par SMITH au sein du système VISUALSEEK[SMITH96] qui se dégage du problème d'identification en séparant une requête en deux parties :

- l'ensemble des régions similaires à celles de l'image requête est tout d'abord déterminé à partir de critères simples (couleur, position, aire et rectangle vertical englobant) ;
- une fois la liste des régions similaires obtenue, l'agencement spatial des régions est utilisé pour discriminer les images. Pour cela, le système mis en place est fondé sur les *2-D strings*.

Dans un même temps, GUDIVADA propose de remplacer les *2-D strings* par un graphe spatial pour décrire l'organisation des différents éléments de l'image[GUDIVA95]. Dans cette approche

les objets doivent toujours être identifiés. Cette méthode est comparée en particulier au *2-D C-strings* définis par LEE[LEE89] ; expérimentalement les résultats sont meilleurs d'un point de vue qualitatif. De plus, cette méthode est invariante par rotation.

Une synthèse de ces deux approches (*2-D strings* et graphes d'adjacences) peut être trouvée dans les travaux récents de PETRAKIS. Initialement, il réalisait les recherches de similarités entre structures du corps humain avec des *2-D strings*[PETRAK93, PETRAK96]. Mais, dans ses derniers travaux[PETRAK97, PETRAK02], PETRAKIS montre que les graphes d'adjacences de régions apportent de meilleurs résultats. Ayant segmenté les différentes images, un graphe d'adjacences attribué est alors construit en donnant des caractéristiques aux régions et aux arêtes. Les régions sont caractérisées par leur aire, leur périmètre, leur orientation et leur identification si elle est disponible (tumeur, foie...). Les arêtes quant à elles contiennent des informations telles que les distances entre régions et leurs orientations relatives. Les mesures de similarité entre structures sont alors effectuées par un algorithme de comparaison de graphes attribués. Ce type de méthodes est beaucoup moins rapide que les *2-D strings* (environ 10 fois moins) mais apporte une précision inversement proportionnelle à cette lenteur.

En marge d'une description des images par un ensemble de régions, il est à noter que MATAS propose d'utiliser un graphe des couleurs, le CAG (*Colour Adjacency Graph*). Chaque nœud représente alors une couleur (définissant un amas de l'histogramme) et non plus une région. Les arêtes symbolisent les adjacences des couleurs au sein de l'image et contiennent les coefficients de réflectance mutuelle. Les différences entre les représentations sont obtenues à partir d'un formalisme de type *graph-matching* d'où une complexité équivalente à celle des graphes d'adjacences de régions. Pour contourner l'énorme complexité des calculs lors de la comparaison des graphes, PARK propose le MCAG (*Modified Colour Adjacency Graph*)[PARK99B]. Après filtrage et quantification de l'image, chaque nœud contient la fréquence d'apparition d'une couleur et chaque arête code la longueur de la frontière entre couleurs. Ces simples informations peuvent alors être stockées dans une matrice de faible taille (égale au nombre de couleurs utilisées). La comparaison de deux représentations est alors grandement facilitée. De plus, pour prendre en compte la géométrie de la répartition des diverses couleurs, Park met en place un autre descripteur, le SVG (*Spatial Variance Graph*). Les nœuds de ce nouveau graphe contiennent la variance spatiale de chaque classe colorimétrique et les arêtes, les variances relationnelles entre classes. La combinaison de ces deux graphes permet donc de prendre en compte la distribution des différentes couleurs au sein de l'image, leurs relations et leur répartition géométrique propre.

Enfin, il est à noter qu'il existe un certain nombre de statistiques globales prenant en compte la composition spatiale des images de manière partielle. JACOBS a conçu un système et un critère de comparaison d'images qui utilise l'information spatiale et des caractéristiques visuelles fondées sur les coefficients ondelettes principaux[JACOBS95]. PASS propose aussi une technique (*Color Coherence Vectors*) qui divise un histogramme global d'une image en composants cohérents et distincts[PASS96]. La mesure de cohérence colorimétrique identifie l'existence des régions colorées connectées. Cette technique bien qu'ayant des performances supérieures aux histogrammes simples ne permet pas d'effectuer des recherches utilisant la position des régions colorées. Cet ensemble de méthodes utilise donc partiellement l'information spatiale et ne permet pas de l'utiliser efficacement pour une requête précise portant sur un agencement donné d'objets.

3.5 Évaluer l'indexation

La performance d'un système d'indexation est très subjective. Généralement, un système peut être considéré comme efficace s'il retourne les informations désirées. La quantification de cette notion est très difficile car le jugement en lui même est subjectif. Les images retournées sont-elles réellement pertinentes ? La réponse sera différente d'une personne à l'autre. Il faudrait donc pour évaluer correctement un système que de nombreuses personnes effectuent de nombreuses requêtes. Le terme « nombreux » apparaissant au moins deux fois de trop dans la phrase précédente, un tel système d'analyse paraît irréaliste à grande échelle. En considérant une base de 10 000 images, la performance réelle du système pourrait être évaluée de manière correcte si 100 personnes (pour l'analyse statistique) effectuaient chacune 1000 recherches (10% de la taille de la base) et qu'elles déterminaient la pertinence des résultats obtenus selon leurs propres critères.

Une solution est de prendre le problème à revers. Au lieu de déterminer si les images retournées sont correctes du point de vue d'un utilisateur, la base d'images est organisée en catégories et seules les images d'une même catégorie sont considérées comme pertinentes. De cette manière, les recherches peuvent être effectuées automatiquement et les résultats analysés sans intervention humaine. Cette approche est bien évidemment aussi biaisée que la précédente car la génération des différentes catégories est réalisée de manière manuelle et donc subjective. Les deux avantages de cette méthode sont la diminution du temps d'évaluation de la performance d'un système et surtout la possibilité de la comparer à d'autres utilisant la même organisation de la base.

Dans ce cadre, les méthodes les plus répandues pour évaluer la recherche d'images sont les notions de précision et de rappel. La précision correspond au pourcentage d'images pertinentes retrouvées au sein de l'ensemble des images retournées par le système. Le rappel donne le pourcentage d'images pertinentes retournées par rapport au nombre total d'images pertinentes présentes dans la base. En utilisant les notations de la figure [3.10], la précision et le rappel sont définis par :

$$\text{Précision} = \frac{B}{B+C}, \text{ Rappel} = \frac{B}{A+B}.$$

Il est possible de tracer le graphe Précision/Rappel (cf. figure [3.11]). Ce type de courbes correspond à la performance du système pour une requête donnée en fonction du pourcentage d'images pertinentes retrouvées. Un système parfait aura alors une courbe constante de valeur 1. Plus les courbes tendent vers cet extrême plus les méthodes sont performantes. Ce type de graphique ne portant pas toute l'information désirée[SALTON92], d'autres mesures basées sur la précision et le rappel peuvent être utilisées :

- $P(N)$, $P(N_R)$: la précision après avoir retrouvé N (valeur quelconque) ou N_R images avec N_R le nombre d'images pertinentes pour la requête ;
- la précision moyenne ;
- le rappel à une précision donnée (0.5 par exemple).

Il est aussi possible de définir de nombreuses mesures de qualité d'un système de recherche d'images par le contenu :

- le rang de la première image pertinente ;
- le rang moyen des images pertinentes ;

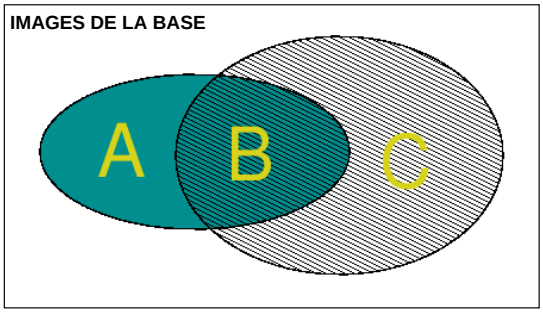


FIG. 3.10 – Ensembles d'images obtenus après une requête
A : images pertinentes non retrouvées.
B : images pertinentes retrouvées.
C : images non pertinentes retrouvées.
(A+B) : images pertinentes.
(B+C) : images retrouvées.

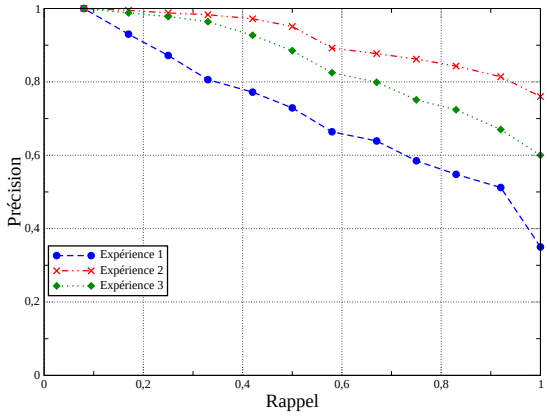


FIG. 3.11 – Exemple de courbes Précision/Rappel

– le taux d'erreur $\frac{C}{B+C}$ défini par HWANG et al. qui correspond en fait au complémentaire à 1 de la précision [HWANG99].
– MÜLLER[MULLER99] propose d'utiliser une normalisation du rang moyen des images pertinentes définie par :

$$\widetilde{Rang} = \frac{1}{NN_R}\left(\sum_{i=0}^{N_R} Ri - \frac{N_R(N_R-1)}{2}\right)$$

où R_i est le rang de la i^e image pertinente et N le nombre d'images total de la base. $\widetilde{Rang}$ vaut 0 pour une performance parfaite et tend vers 1 dans le cas opposé. Une recherche aléatoire aura un $\widetilde{Rang}$ de 0.5.

Pour cette étude nous nous limiterons à l'évaluation de la performance en utilisant le graphe précision/Rappel et le calcul de $\widetilde{Rang}$. La première est la méthode la plus classique et la seconde apporte sur les résultats une deuxième vision utile pour infirmer ou confirmer les conclusions effectuées.

3.6 Conclusion

Comme nous l'avons montré dans ce chapitre, le(s) problème(s) de l'indexation d'images est (sont) très vaste(s) que ce soit au niveau des questions auxquelles il faut répondre ou que ce soit au niveau des solutions apportées.

Le principal problème lié à la recherche d'images est de définir une méthode de caractérisation de celles-ci qui soit assez simple pour que les traitements soient rapides. Mais elle doit également être la plus complète possible pour permettre de répondre à une vaste gamme de problèmes. Il faut donc trouver un compromis entre vitesse et précision.

Les approches globales, les premières mises en place dans ce cadre, sont très rapides mais ne permettent pas d'effectuer des recherches fines au sein des images. Ainsi, pour pallier cette limitation et imiter encore plus fidèlement le système visuel humain, des techniques dites *spatiales* ont été mises en place depuis le début des années 1990.

Dans le chapitre suivant, nous allons présenter les différentes structures de données fréquemment utilisées en traitement d'images. À l'issue de ce chapitre, nous proposerons notre structure de données pour l'indexation. Celle-ci est de type spatiale mais prend aussi en compte la notion de composition des différents objets inclus dans les images. Celles-ci sont décrites par une représentation multi-échelles adaptée à la recherche d'objets.

CHAPITRE 4

REPRÉSENTATIONS D'IMAGES

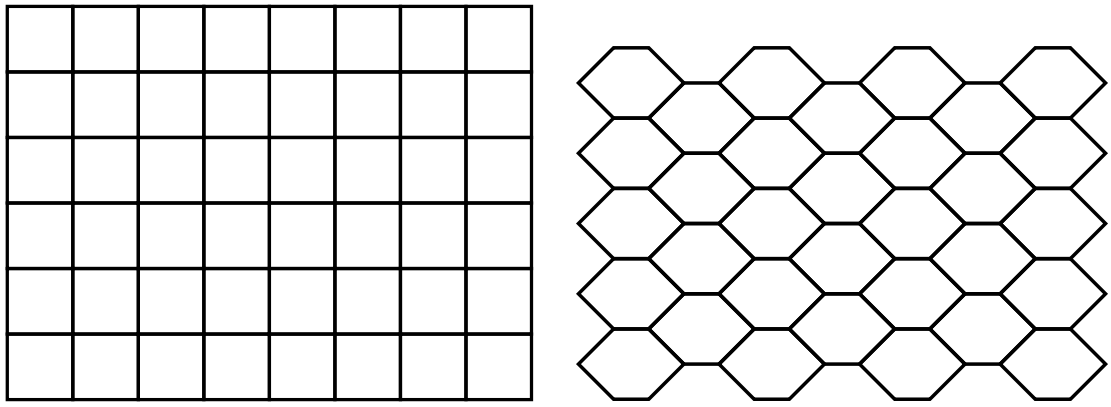
Sommaire

4.1 Introduction

Un programme est généralement composé de deux parties dépendantes : l'algorithme et les données. En traitement d'images, les processus développés sont fortement liés au système de stockage des images. La structure même de stockage de l'information peut considérablement influer sur la simplicité et l'implantation des algorithmes.

De très nombreux systèmes de représentation peuvent ainsi être utilisés. Ce chapitre a pour objectif d'exposer les différentes structures de données généralement utilisées en traitement d'images et en particulier en indexation d'images.

Nous exposons tout d'abord les modèles qui décrivent l'image par une structure plane. Ces approches considèrent globalement l'image comme un ensemble d'éléments de base organisés au sein de la représentation (pixel, région...).



rectangulaire hexagonale

FIG. 4.1 – Deux grilles d'échantillonage permettant d'obtenir la matrice image

Ces méthodes voient l'image avec un niveau de détail, très fin dans le cas des matrices, ou à un niveau de découpage de l'image pour les régions. Or, selon l'approche multi-résolutions, chaque élément d'une image possède une résolution à laquelle les divers traitements envisagés (recherche, extraction, description...) sont les plus adaptés. Ainsi, depuis les années 70, le traitement d'images bénéficie des travaux liés à la multi-résolutions. Différentes structures issues de ces études sont donc présentées dans la seconde partie de ce chapitre

La dernière partie traite quant à elle de la structure que nous avons mise en place pour décrire des images dans un but d'indexation. Celle-ci essaye d'imiter le plus fidèlement possible le système visuel humain en s'inspirant des approches multi-résolutions et arborescentes et en particulier des pyramides irrégulières.

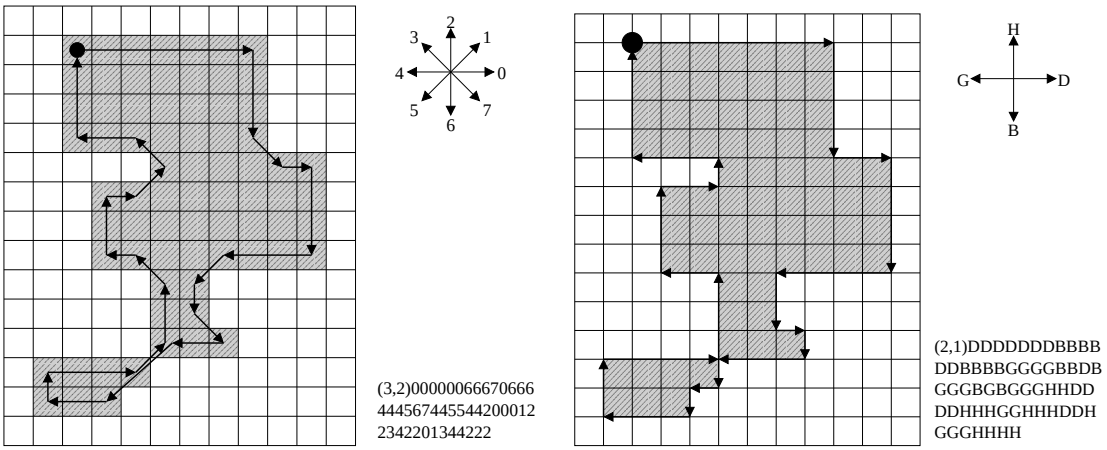
4.2 Structures de données planes d'images

4.2.1 Matrices

La matrice est la représentation brute d'une image. Chaque élément de la matrice peut être un entier, un réel ou même un complexe quantifiant une caractéristique du pixel de la grille d'échantillonage (luminosité, composante couleur ou spectrale...). Cette grille peut avoir des formes diverses comme celles présentées à la figure [4.1].

Ce type de données image est souvent le résultat direct du système d'acquisition. Ainsi, cette représentation de l'image est totale et ne tient pas compte du contenu de l'image ; elle n'est liée qu'à la mesure physique réalisée par le capteur.

Il est possible de distinguer certains types de matrices spéciales. Une image binaire est composée exclusivement de 0 et de 1. Une image segmentée est souvent représentée par une matrice composée d'entiers caractérisant l'appartenance des pixels à une région donnée. Les images multispectrales peuvent être stockées en utilisant une matrice pour chaque bande spectrale. Pour obtenir une structure de données hiérarchique, des matrices de différentes tailles peuvent être construites.



Codes de FREEMAN *Crack-codes*

FIG. 4.2 – Exemple de parcours pour les codes de FREEMAN et les *crack-codes*
Le point de départ est symbolisé par un rond noir.

Cette structure contient la totalité de l'information de l'image. N'importe quelle donnée peut ainsi être extraite de la matrice. Suivant l'information désirée, les processus utilisés sont plus ou moins complexes. Au sein de la matrice, toutes les informations sont présentes mais elles ne sont pas organisées ; les autres modèles sont généralement issus d'un post-traitement de la matrice permettant d'extraire et de structurer les informations pertinentes pour l'application souhaitée.

En indexation d'images, de très nombreuses statistiques peuvent être calculées directement à partir de cette représentation (histogramme couleur, matrice de cooccurrences...). L'ensemble des valeurs calculées forment alors la caractérisation de l'image.

4.2.2 Chaînes

Les chaînes sont formées à partir de primitives de l'image. Chacune d'entre elles est codée par un symbole au sein de la chaîne. Cette méthode de représentation est très intéressante pour des structures qui peuvent être arrangées en séquence de symboles. Le voisinage dans la chaîne est souvent lié au voisinage des primitives dans l'image elle-même.

Deux exemples classiques sont les chaînes de codes (ou codes de FREEMAN[FREEMA61]) et les *crack-codes*[BRICE70] qui sont très utilisés pour la description des frontières de régions ou tout autre ensemble d'un pixel de large. Dans les deux cas, le contour d'une région est décrit à partir d'un point de départ en donnant successivement les directions prises par le contour. La différence principale entre les deux approches est la notion même de bord : les codes de FREEMAN considèrent le contour comme l'ensemble des segments reliant les centres des pixels du bord (approche pixel) alors que les *crack-codes* utilisent une approche inter-pixels en considérant le parcours extérieur des pixels du bord (cf. figure [4.2]).

Historiquement, les *crack-codes* sont apparus après les codes de FREEMAN dans le but d'éliminer quelques problèmes topologiques tels que la possible appartenance d'un même élément structurant (pixel ou frontière entre pixels suivant la méthode) à deux segments du bord. Ce phénomène rend difficile des tâches *a priori* triviales comme le suivi de contour ou le remplissage de région.

De plus, en utilisant un ensemble de pixels comme bords, la frontière de la région et de son complémentaire ne coïncident pas.

Dans les deux exemples précédents, les chaînes sont donc composées du point de départ et d'une suite de caractères codant le parcours du bord de l'objet ; chaque caractère correspondant à une direction. La dépendance de ces méthodes vis-à-vis du point de départ peut être éliminée en considérant la chaîne comme cyclique. Dans ce cas précis, le point de départ est omis. Ces systèmes sont aussi invariants par translation mais souffrent d'une variance forte par symétrie d'axe ce qui pénalise fortement ces structures dans un cadre de comparaison. En particulier, en reconnaissance de formes, si un objet est retourné au moment de l'acquisition, il sera difficile de le reconnaître avec de telles structures. Il est à noter qu'il existe aussi d'autres systèmes de codage de chaînes invariants par symétrie d'axes tels que les codes de BRIBIESCA[BRIBIE99].

4.2.3 Structures de données topologiques

La topologie étudie les propriétés de l'espace d'un point de vue qualitatif (adjacences, voisinages, inclusion...). Une image peut être vue comme un ensemble d'éléments (de régions) et des relations qui les unissent. Les structures dites « topologiques » tentent de décrire le plus complètement possible l'organisation des divers éléments au sein de l'image.

4.2.3.1 Graphes

Les graphes sont souvent utilisés pour décrire la composition et la topologie des images. Par définition, un graphe est une structure algébrique $G = \{V, E\}$ composée de nœuds $V = \{v_1, v_2, \ldots, v_n\}$ et d'arêtes $E = \{e_1, e_2, \ldots, e_m\}$. Chaque arête e_k est connectée à une paire de nœuds $\{v_i, v_j\}$ non nécessairement distincts. Le nombre d'arêtes incidentes à un nœud est appelé degré. Un graphe attribué est un graphe au sein duquel des données sont associées aux nœuds et/ou aux arêtes.

La graphe d'adjacences de régions (ou *RAG* pour *Region Adjacency Graph*) est un exemple classique de ce type de structure (cf. figure [4.3]). Chaque nœud correspond à une région de l'image et les couples de nœuds adjacents sont liés par une arête. Le *RAG* possède un certain nombre de propriétés intéressantes et dispose de tous les développements théoriques et algorithmiques de la théorie des graphes depuis le milieu du 19^{e} siècle. Cela rend cette représentation intéressante notamment pour la comparaison ou la simplification.

Par contre, le *RAG* n'apporte pas une description topologique complète. La structure en elle même ne prend pas en compte les multi-adajcences et les inclusions. Pour obtenir ces propriétés il faut utiliser les graphes duaux (utilisation combinée du *RAG* et de son dual) qui ne sont pas encore topologiquement complets[KROPAT95] : ils ne codent pas l'ordonnancement des adjacences contrairement aux cartes généralisées (cf. section 4.2.3.2). La figure [4.4] propose un exemple de deux configurations différentes ayant même représentation avec les approches par graphes.

En indexation, les approches dites spatiales utilisent très souvent ce type de description pour formaliser l'organisation des objets au sein de l'image[PETRAK02]. De plus, cette structure permet de stocker très simplement diverses caractéristiques des objets et de leurs relations au sein des nœuds et des arêtes.

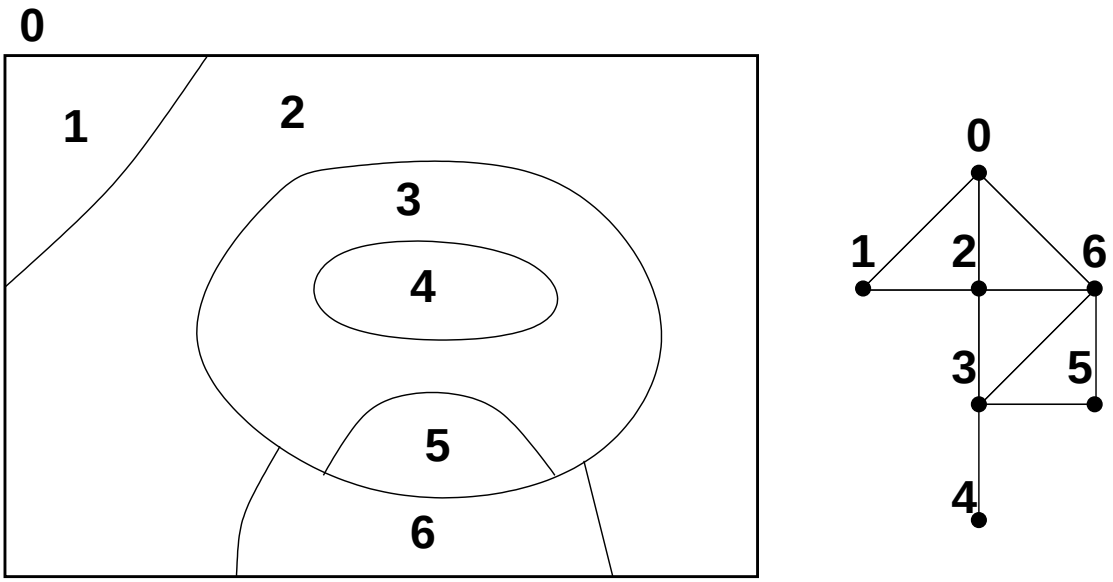


FIG. 4.3 – Exemple de graphe d'adjacences de régions pour une carte de segmentation
La région infinie 0 correspond à l'extérieur de l'image.

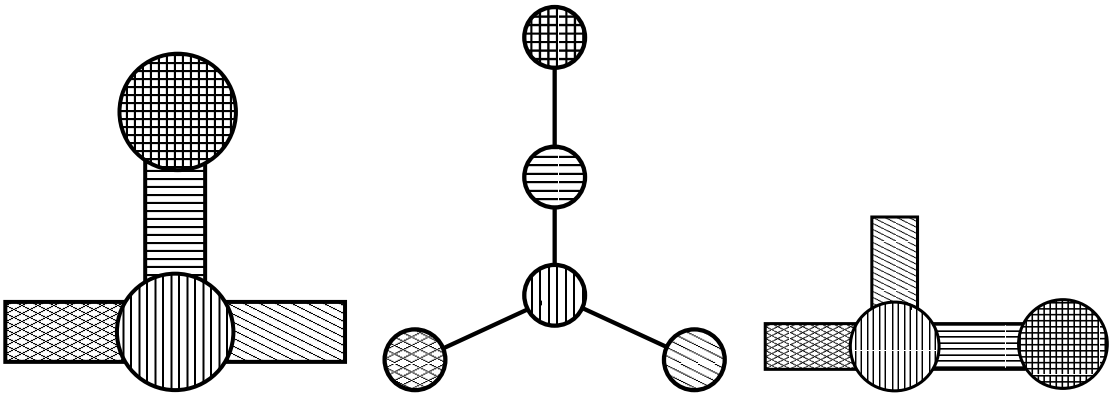


FIG. 4.4 – Deux structures différentes (a et c) et leur *RAG* équivalent (b)

4.2.3.2 Cartes généralisées

Les cartes généralisées [LIENHA91] apportent quant à elles une description topologique complète. Plus précisément, une subdivision d'un espace topologique de dimension n (une image segmentée si $n = 2$) est une partition de cet espace en cellules de dimension i (ou i-cellules) pour $0 \leq i \leq n$ (si $n = 2$, les cellules sont la région, la frontière et le point). Deux cellules sont incidentes si l'une appartient au bord de l'autre (les frontières sont incidentes avec les points extrêmes de celles-ci). Deux i-cellules adjacentes sont incidentes à une même cellule (deux régions adjacentes sont incidentes à leur frontière commune).

Les cartes généralisées de dimension n, ou n-G-cartes, sont un modèle combinatoire défini pour représenter la topologie d'une subdivision de l'espace. Plus précisément, une n-G-carte représente

la topologie d'une n-quasi-variété, orientable ou non, avec ou sans bords. Ainsi, en dimension 2, une 2-G-carte peut être vue comme un multi-graphes planaire (multi-adjacences entre régions possibles) codant intrinsèquement l'ordonancement des adjacences. Par contre, en dimension supérieure il est très difficile de faire un lien avec les graphes qui ne sont pas adaptés aux problèmes en dimensions supérieures à 2.

Nous pouvons ici citer les travaux de LUC BRUN sur les aspects segmentations d'images utilisant les G-cartes comme structure de représentation[BRUN03, BRAQUE98]. Dans le cadre de la caractérisation d'images, les travaux de GUILLAUME DAMIAND sont aussi très intéressants[DAMIAN04].

4.3 Structures de données hiérarchiques ou multi-résolutions

4.3.1 Notion de multi-résolutions

La vision par ordinateur est par nature très coûteuse en temps de calcul de par le volume important de données à traiter. Les systèmes sophistiqués traitent un volume d'information allant de la centaine de kilo-octets aux dizaines de Méga-octets ce qui est encore très peu comparé au système visuel humain dont le flot de données est estimé à 3000 Méga-octets par seconde. Des solutions de traitements utilisant des architectures parallèles permettent de résoudre certains problèmes de manière brutale. Par contre, un certain nombre d'applications sont difficilement parallélisables et cette solution n'est donc pas envisageable. Pour remédier à cela, la multi-résolutions a été introduite depuis les années 1970 en traitement d'images. Une structuration hiérarchique des images rend ainsi possible l'adaptation de la stratégie de traitement à partir de peu de données. Il est également possible d'utiliser la résolution la plus adaptée pour répondre à un problème tel que la caractérisation, la recherche ou l'extraction d'objets.

Ces méthodes paraissent naturelles au regard du système visuel humain de par leurs nombreuses analogies structurelles et fonctionnelles. Nous allons, par exemple, focaliser notre regard sur un objet de notre champ de vision uniquement si nous recherchons une information plus précise sur celui-ci ; un objet peut aussi être plus ou moins loin de l'observateur et n'être pas perçu avec le même niveau de détail. Pratiquement, une représentation multi-résolutions comporte un ensemble d'images dérivées de l'originale qui contiennent de moins en moins d'informations.

Du point de vue spatial, la taille des images dérivées diminue et provoque donc une baisse de la résolution (cf. figure [4.5]). Une autre *forme* de la multi-résolutions conserve la taille de l'image originale au cours des échelles mais réduit les détails (par filtrage notamment) (cf. figure [4.6]). Ces techniques cherchent ainsi à simplifier de plus en plus l'image selon un certain critère que ce soit en diminuant les détails ou en diminuant le nombre de régions d'une carte de segmentation. Dans ce cas, on parle plutôt de processus multi-échelles. C'est aussi le principe de certains algorithmes de décomposition qui vont permettre d'afficher progressivement une image à différentes échelles comme dans JPEG2000.

Il paraît assez simple de construire différentes vues d'une même image à des résolutions ou des échelles diverses mais la détermination *a posteriori* du meilleur niveau pour réaliser un traitement est une tâche non triviale. Certaines applications particulières où le contenu de l'image est connu (le nombre d'objets par exemple) peuvent déterminer le niveau optimal mais leur nombre est assez

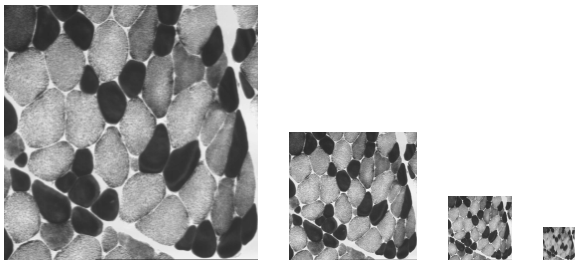


FIG. 4.5 – Une image à différentes résolutions spatiales

FIG. 4.6 – Simplifications des détails d'une même image[TREMBL02]
De gauche à droite, les échelles croissantes obtenues par lissage par diffusion anisotropique (en haut) et isotropique (en bas).

restreint. L'exemple classique est lié à la segmentation. Disposant de partitions plus ou moins fines d'une image, il paraît difficile de déterminer automatiquement la meilleure segmentation. Par contre, il est assez facile de choisir laquelle utiliser connaissant le niveau de détail requis (et donc le nombre de régions minimum souhaité) (cf. figure [4.7]).

4.3.2 Pyramides matricielles - gaussiennes et laplaciennes

Une des premières classes de méthodes de représentation multi-résolutions mise en place fut les pyramides matricielles. Elles sont composées d'une séquence d'images $\{M_0, \ldots, M_{N-1}, M_N\}$ où M_0 est de la taille de l'image originale. L'image M_{i+1} dérive de M_i par réduction de la résolution.

Il est ainsi possible de réduire la taille des images d'un facteur 2 à chaque itération (pour obtenir des images 2 fois plus petites) ; dans ce cas, M_N n'a qu'un seul pixel et les traitements sur M_{i+1} seront 4 fois plus rapides que sur M_i. La figure [4.8] donne un exemple d'un telle pyramide.

Ce type de pyramide est utilisé principalement lorsqu'il est nécessaire de travailler simultanément sur une image à différentes résolutions (codage ou segmentation par exemple).

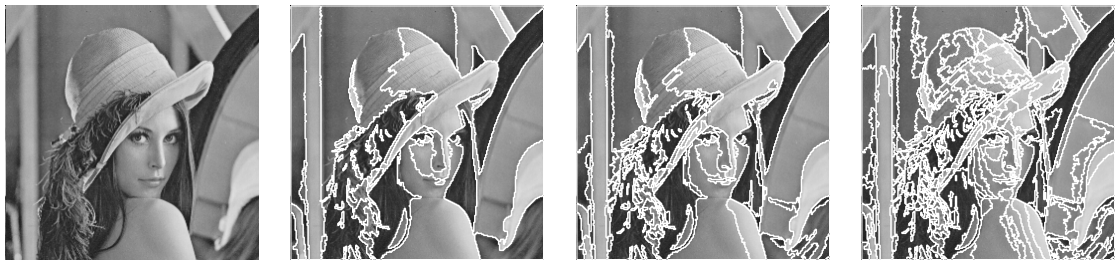


FIG. 4.7 – Différents niveaux de segmentation pour une même image

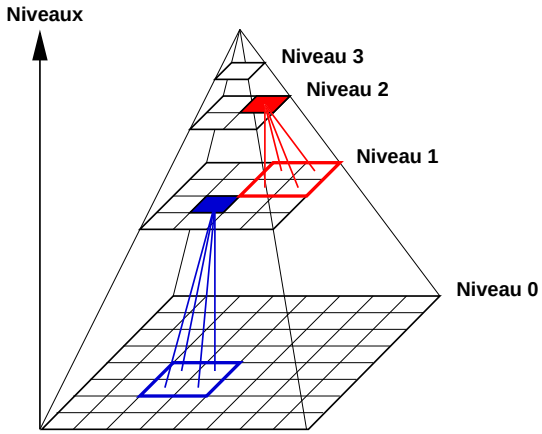


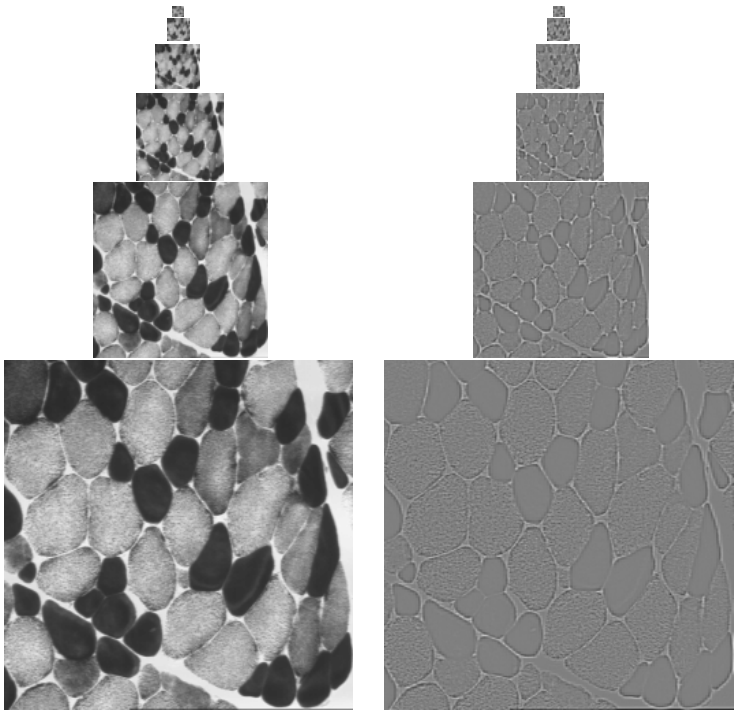
FIG. 4.8 – Pyramide par réduction

BURT a développé l'une des premières approches multi-résolutions en traitement d'images avec les pyramides gaussienne et laplacienne[BURT83]. La pyramide gaussienne est construite par répétition de moyennages locaux pondérés où chaque image de niveau $k+1$ a une surface représentant le quart de celle du niveau k. La figure [4.8] montre un exemple d'un telle pyramide dans le cas où le moyennage local est réalisé sur un voisinage 2×2 (un voisinage 5×5 est plus généralement utilisé). La pyramide gaussienne peut aussi être vue comme une série de versions filtrées passe-bas et rééchantillonnées de l'image originale. Ce type de pyramides tire son nom du fait que le filtrage correspond pratiquement à un filtrage gaussien.

La pyramide laplacienne correspond au dual de la pyramide gaussienne. Elle est calculée simplement par différence des images successives de la pyramide gaussienne G_k et G_{k+1} : $L_k = G_k - G_{k+1}$ (G_{k+1} doit donc être étendue à la taille de G_k). Par conséquent, L_k contient les hautes fréquences éliminées lors de la création de G_{k+1}.

La figure [4.9] donne un exemple de telles pyramides. Dans le cas d'une image carrée de largeur N, une de ces pyramides comprend alors $N^2(1 + \frac{1}{4} + \frac{1}{16} + \ldots) \simeq 1.33N^2$ pixels.

Il est d'ailleurs possible de faire évoluer l'image sans pour autant réduire sa taille. Le *scale-space* en est l'exemple le plus répandu[WITKIN83, LINDEB90]. Dans ce cas, les images dérivées peuvent être calculées, par exemple, par convolution de l'image originale avec un noyau gaussien de taille croissante et gardent alors la taille de l'image originale. C'est donc un système multi-échelles.



Pyramide gaussienne Pyramide laplacienne

FIG. 4.9 – Pyramides gaussienne et laplacienne

4.3.3 Pyramides arborescentes ou géométriques

En traitement d'images, il est fréquent de travailler sur plusieurs représentations d'une même image à une échelle donnée. Chacune d'elles contient alors une composante de l'information utile pour l'analyse de l'image (texture, couleur ou toute autre caractéristique). De ce point de vue, les pyramides matricielles sont peu intéressantes ; il est beaucoup plus aisé de travailler sur une représentation pyramidale arborescente utilisant également une approche par division récursive tout en prenant en compte les particularités géométriques.

Deux types de construction sont généralement utilisées : la génération ascendante (ou *bottom-up*) et la construction descendante (ou *top-down*). Le *bottom-up* considère tous les pixels de l'image comme des régions. Au cours de la construction, les régions adjacentes sont fusionnées pour obtenir le niveau suivant de la pyramide. La taille moyenne des régions augmente donc au cours de la construction à l'inverse de leur nombre. Le niveau supérieur de la pyramide est obtenu au moment ou aucune fusion n'est possible. De manière inverse, le *top-down* considère l'image entière comme une seule région. À chaque niveau les régions sont divisées ce qui engendre une augmentation du nombre de régions.

4.3.3.1 Quadtree

Un *quadtree* ou arbre quaternaire est une construction *top-down* utilisant la propriété de récursivité du maillage carré[HOROWI74, SAMET80]. L'image représentée doit être carrée et de

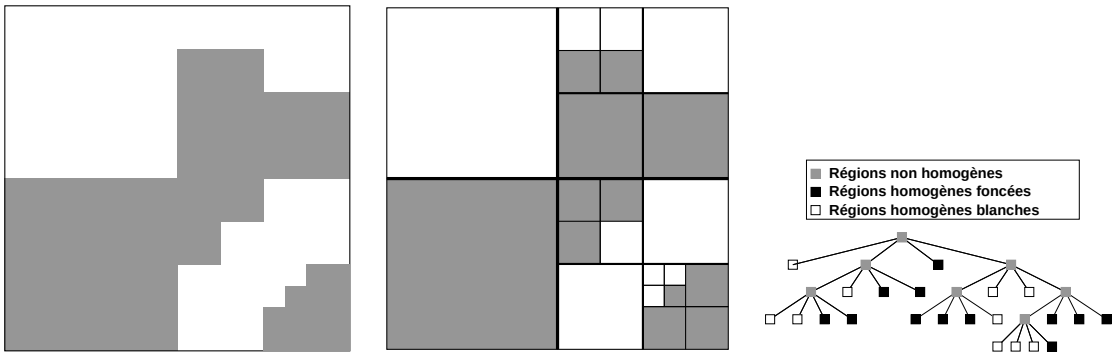


Image binaire Découpage quadtree Quadtree

FIG. 4.10 – Principe de découpage *quadtree*

dimension égale à une puissance de 2. Le bloc initial M_0 correspondant à l'image entière est donc de dimension $2^m \times 2^m$ avec $m \in \mathbb{N}$.

Le principe consiste, à partir de M_0, à diviser récursivement tout bloc non homogène selon le critère choisi (couleur, texture...) en 4 sous-blocs (cf. figure [4.10]). De plus, il n'est pas nécessaire de stocker tous les nœuds de la pyramide car :

- si une région est homogène, ses 4 quadrants fils seront inutiles car également homogènes suivant le critère choisi ;
- si une région est non homogène, elle ne sera pas stockée car ses 4 quadrants fils le seront.

Cette remarque permet de limiter les calculs en ne construisant pas l'arbre complet.

Cette structure est souvent utilisée en segmentation d'images. Les blocs adjacents correspondant à des régions homogènes mais qui ont été divisées par le partitionnement doivent alors être regroupés. En effet, le *quadtree* est complètement dépendant du partitionnement géométrique quaternaire qui ne peut pas être adapté au contenu de l'image et qui, au contraire, introduit des contours horizontaux et verticaux. De plus, cette représentation n'est robuste ni aux translations ni aux rotations. L'intérêt de cette structure est surtout lié à sa simplicité de mise en œuvre et à sa rapidité.

4.3.3.2 Pyramides régulières

La pyramide régulière standard est la pyramide quaternaire qui associe le coté structurel de la pyramide gaussienne avec la subdivision par maillage carré du *quadtree* (cf. figure [4.8]). La construction est réalisée en *bottom-up* ; 4 pixels voisins (les fils) sont représentés au niveau supérieur par un pixel (le père). Tout comme le *quadtree*, l'image doit donc être carrée et de dimension $2^m \times 2^m$ avec $m \in \mathbb{N}$. Ainsi, le sommet de la pyramide (ou apex) et de taille 1 et se trouve au niveau m. Cette structure simple a été utilisée en filtrage[PARK91] mais l'aspect quaternaire rigide ne permet pas de mettre en œuvre un algorithme de segmentation satisfaisant.

Pour relâcher cette contrainte liée au maillage quaternaire, BURT met en place une pyramide liée[BURT81]. Pour rendre plus flexible la représentation, un élément de la pyramide n'est plus lié à un seul père mais à quatre ; chaque père a donc au maximum 16 fils (cf. figure [4.11]). Les caractéristiques des pères sont tout d'abord calculées à partir de celles de ses fils. Chaque fils peut ensuite choisir son *meilleur* père. Si aucun ne convient alors c'est qu'il ne peut pas faire partie d'une

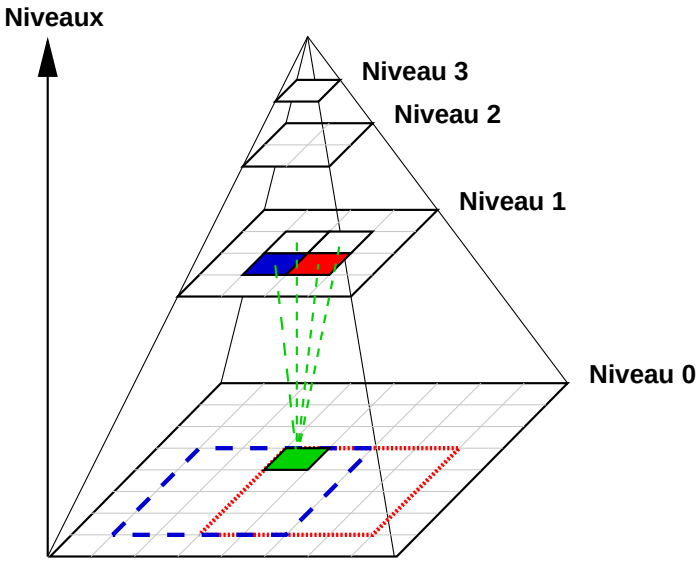


FIG. 4.11 – Liaisons inter-niveaux au sein d'une pyramide liée
Le pixel du niveau bas à quatre pères et les deux pères ont chacun 16 fils. Les champs récepteurs de pères voisins ont donc une zone commune.

région plus importante et qu'il caractérise à lui seul une région. Cette région est alors constituée de son champ récepteur (i.e. ses descendants). Dans le cas contraire, seul le lien au père *favori* est conservé permettant de reconstruire facilement la composante connexe de chaque région. Cette méthode permet de répondre en partie au problème de rigidité des pyramides quaternaires mais reste sensible aux rotations et aux translations. C'est une première évolution vers les pyramides irrégulières.

4.3.3.3 Pyramides irrégulières ou pyramides de graphes

Pour obtenir une représentation n'imposant pas de contraintes structurelles, il faut se tourner vers une représentation où les cellules sont de tailles et de formes quelconques et dont le nombre de voisins n'est pas limité. Les graphes sont très utilisés pour cela car structurellement libres[MONTAN91, JOLION92]. Les graphes d'adjacences de régions, par exemple, symbolisent les régions et leurs relations à chaque niveau de la pyramide. Ainsi, dans une approche *bottom-up* tous les pixels de l'image sont d'abord considérés comme les cellules du premier niveau de la pyramide. Chaque niveau est alors contracté suivant un critère donné pour obtenir le suivant (cf. figure [4.12]).

Cette approche supprime de nombreuses contraintes des solutions précédentes : image de taille $2^m \times 2^m$, non robustesse aux translations et aux rotations, limitation des formes des régions représentées. De plus, l'utilisation de graphes apporte un cadre mathématique déjà fortement formalisé dans lequel de nombreux développements ont déjà été réalisés.

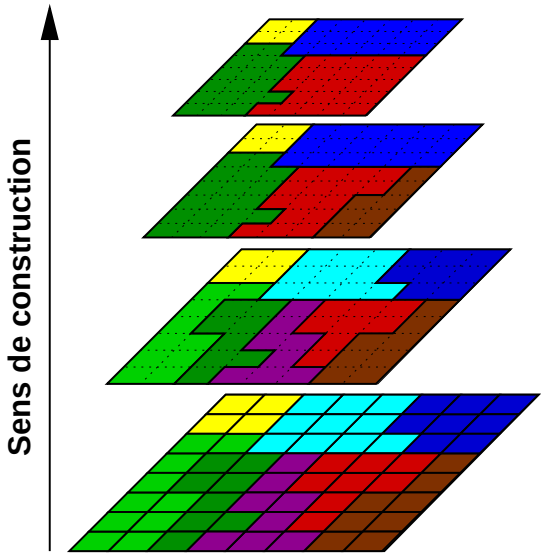


FIG. 4.12 – Construction par fusion d'un pyramide irrégulière
Les traits gras représentent les limites de régions et donc les zones d'adjacence.

4.4 Notre approche : le graphe pyramidal

4.4.1 Principe

Cette approche se situe dans le cadre des pyramides irrégulières. Les idées motrices de cette structure de données sont :

– tout objet de l'image est composé de parties ;
– les objets sont en interaction les uns avec les autres.

Ces deux aspects apparaissent comme très pertinents dans le cadre de la vision humaine. Le premier regard porté sur une image va permettre de déterminer grossièrement les différents objets la composant, puis une vue prolongée de l'image apportera plus de détails sur la composition de ces objets. De plus, les interactions d'un objet avec le reste de l'image mettent en place le contexte général dans lequel il baigne. Ces informations complémentaires peuvent alors confirmer, infirmer ou préciser la nature même de l'objet. La nature d'une voiture dans la main d'un enfant peut par exemple être précisée en « voiture miniature » (cf. figure [4.13]).

Ainsi, si l'image est vue comme un objet, elle se décompose en parties qui sont elles-mêmes formées de sous-parties. Ce processus pouvant se reproduire autant de fois que nécessaire. Il apparaît que la notion d'interaction peut être codée très facilement par un graphe d'adjacences et celle de composition par une pyramide. Une pyramide de graphes offre donc un lien naturel de représentation des images. Cette structure ressemble fortement à une pyramide irrégulière mais la différence majeure en est le sens de construction. Les pyramides irrégulières sont construites de manière ascendante alors qu'ici le processus est descendant. Le sommet d'une pyramide irrégulière peut être tronqué alors que dans cette approche c'est la base de la pyramide qui est élaguée (aux endroits où la décision est prise de ne plus diviser une région).

FIG. 4.13 – Exemple d'image où le contexte précise l'interprétation
La présence de l'ours en peluche précise la caractérisation de la voiture en « voiture miniature ».

La figure [4.14] présente le schéma d'un graphe pyramidal issu de segmentations manuelles de l'image. L'image entière est considérée comme une seule région ; c'est le sommet. La segmentation la plus grossière fournit le niveau suivant. L'image se redécompose alors en plusieurs régions. Elles sont représentées par les nœuds du graphe et leurs adjacences par les arêtes. Elles possèdent tous la même région mère : l'image. Successivement, les diverses segmentations fournissent ainsi un nouveau niveau. Cette structure peut se comparer à un arbre d'inclusion au sein duquel les graphes d'adjacences des niveaux auraient été incorporés.

Dans cette structure, les régions peuvent être décrites au niveau des nœuds et les frontières au sein des arêtes et cela à chaque niveau de la pyramide. De par la structure pyramidale, il apparaît intéressant d'utiliser des caractéristiques multi-échelles qui pourront être calculées récursivement aussi vite que possible. Les histogrammes rentrent par exemple dans ce schéma ; ayant les histogrammes des régions filles, il suffit de les sommer pour obtenir celui du père.

Une approche similaire a été mise en place par XU qui propose d'utiliser conjointement un arbre d'inclusion et une matrice d'adjacence pour décrire une image[XU00]. Tout d'abord, une segmentation multi-échelles des images est obtenues par un processus de regroupement de régions. Une approche descendante est alors utilisée pour déterminer si les objets ou les images sont similaires en comparant les ensembles de régions caractérisées par leur forme et leur couleur. Par rapport au graphe pyramidal le manque de cette structure est la caractérisation des arêtes qui n'est pas possible. De plus, la manipulation de deux structures en parallèle est complexe et peut engendrer quelques problèmes mais globalement l'approche employée est très semblable à la notre.

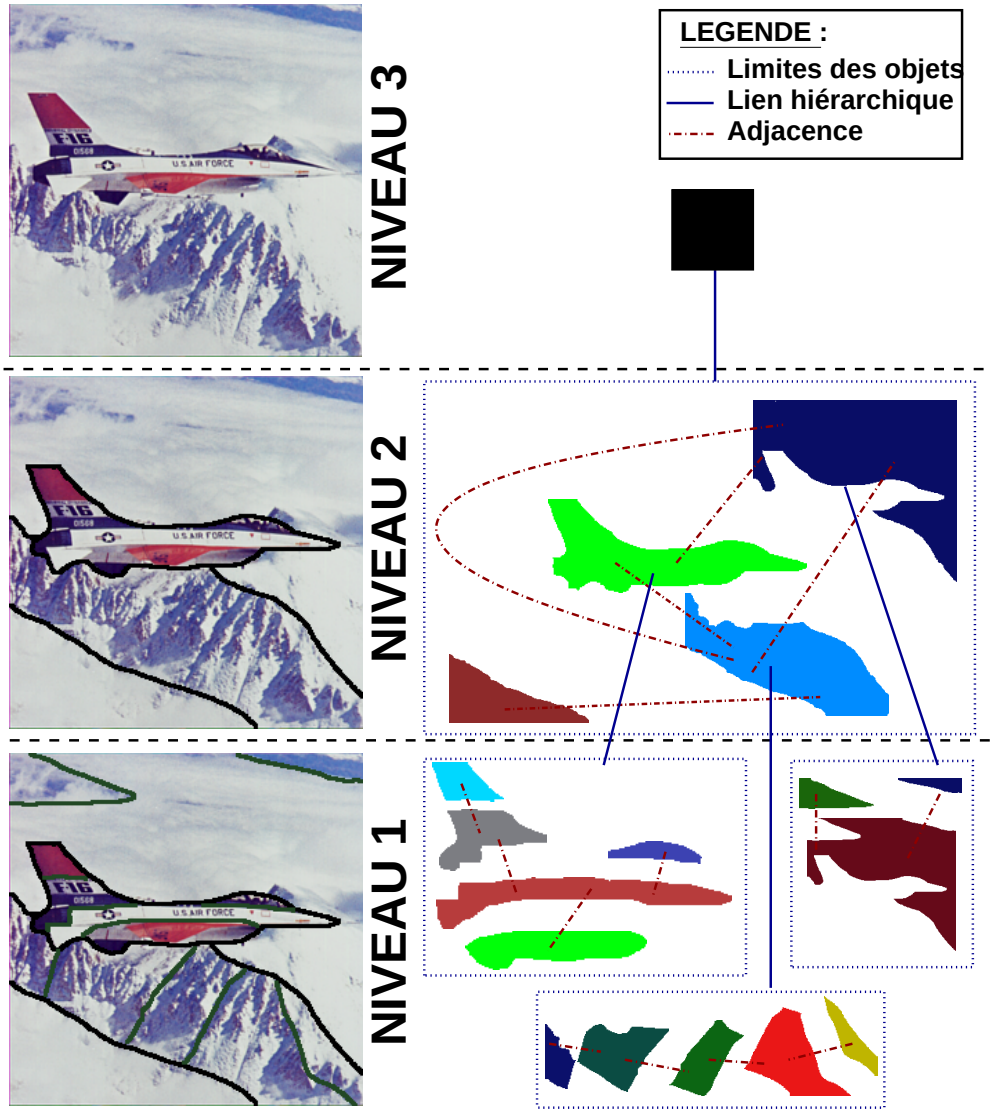


FIG. 4.14 – Graphe pyramidal obtenu à partir de segmentations manuelles

4.4.2 Formalisation

Soit $\mathcal{S}^k$ une partition de l'image I : $\mathcal{S}^k = \{\mathcal{R}_n^k / \bigcup_{n=1}^{N_R^k} \mathcal{R}_n^k = I$ avec $\mathcal{R}_n^k \bigcap \mathcal{R}_m^k = \emptyset$ si $n \neq m\}$ où N_R^k est le nombre de régions de la partition. Nous disposons alors d'une série de partitions de I cohérentes entre elles, i.e. que les segmentations les plus fines sont des sursegmentations des plus grossières :

- $\mathcal{S}_k,\ k = 1 \ldots n$
- $\forall k < l,\ \forall p \in [1 \ldots N_R^k],\ \exists m \in [1 \ldots N_R^l]\ /\ \mathcal{R}_p^k \subset \mathcal{R}_m^l$

La deuxième condition indique donc qu'une région d'une partition est complètement incluse dans une région des partitions plus grossières.

Pour chaque segmentation il est alors possible de construire son graphe d'adjacences de régions constitué d'un ensemble de nœuds V et un ensemble d'arêtes E :

$$\forall k \in [1 \dots n],\ V^k = \{v_1^k, v_2^k, \dots, v_{N_R^k}^k\}\ et\ E^k = \{e_1^k, e_2^k, \dots, e_z^k\}$$

où z est le nombre de relations de voisinage entre les régions symbolisées par des arêtes.

De plus les différents graphes d'adjacences de régions des partitions sont liés entre eux par des arêtes de compositions :

$$\forall k \in [1 \dots n-1],\ L^k = \{l_1^k, l_2^k, \dots, l_{N_R^k}^k\}$$

où $\forall i \in [1 \dots N_R^k]$, l_i^k est l'arête de composition liant le nœud e_i^k correspondant à la région $\mathcal{R}_i^k$ avec le nœud e_j^{k+1} correspondant à la région $\mathcal{R}_j^{k+1}$ tel que $\mathcal{R}_i^k \subset \mathcal{R}_j^{k+1}$.

Le graphe pyramidal d'une image est donc construit à partir de la série $\mathcal{S}_k,\ k = 1 \dots n$ et se compose de :

- $V^k,\ k = 1 \dots n$
- $E^k,\ k = 1 \dots n$
- $L^k,\ k = 1 \dots n-1$

4.4.3 Intérêts et limitations

Les intérêts principaux d'une telle structure sont les notions de multi-échelles et de graphes d'adjacences. L'aspect pyramidal permet d'approcher la méthode d'analyse du système visuel humain en considérant un objet comme composé de plusieurs parties. L'image est ainsi perçue à plusieurs niveaux de détails, du premier regard distinguant uniquement les différents objets jusqu'à la vision en détail cherchant à déterminer la composition fine des objets. De plus, l'aspect pyramidal dans un but d'indexation va rendre possible la mise en place de critères simples et rapides de rejet engendrant une diminution des temps de calculs. Les graphes de chaque niveau représentent très fidèlement l'organisation des différentes régions au sein de l'image. Il est donc aisé de rechercher un arrangement d'objets ou d'utiliser l'environnement d'un objet pour le retrouver. D'autre part, la théorie des graphes apporte un grand nombre d'algorithmes et en particulier pour la mise en correspondance (ou *matching*) ainsi que le calcul de similarité entre structures.

Par contre, la limitation majeure de cette approche est son lien fort avec la segmentation. En traitement d'images c'est un problème récurent car actuellement aucun algorithme n'est capable de segmenter une image au sens de la vision humaine à partir de critères mathématiques. Des objets perçus par l'œil peuvent ne pas être détectés ou mal délimités. L'aspect pyramidal permet en partie de réduire cette contrainte par l'utilisation de plusieurs niveaux de segmentation en parallèle. Il est aussi à noter le manque des graphes d'adjacences en terme de topologie. La multi-adjacences et l'ordonnancement des adjacences ne sont pas codés directement dans la structure. Ces informations ne seront donc pas utilisables. Pour pallier ce manque, une évolution possible serait d'utiliser une pyramide de G-cartes. Actuellement, ces informations ne sont pas utilisées dans le cadre de l'indexation mais pourraient à terme être importantes pour discriminer deux configurations proches mais différentes.

4.5 Conclusion

Nous avons présenté rapidement dans ce chapitre les différents types de représentations d'images habituellement utilisés en traitement d'images. Les structures planes telles que la matrice, les chaînes ou encore les graphes d'adjacences sont utilisés dans de nombreux travaux en indexation d'images. Par contre, ils ne tiennent pas compte de l'aspect multi-échelles du système visuel humain.

Nous avons alors présenté diverses approches multi-résolutions et multi-échelles pour le traitement d'images qui sont malheureusement pour la plupart peu adaptées à l'indexation d'images et à la recherche d'objets.

Ainsi, en s'inspirant des pyramides irrégulières nous avons mis en place une structure de données qui sera utilisée au chapitre 7 pour de l'indexation de base d'images couleur et de la recherche d'objets au sein d'images. Le chapitre 8 présentera également une application spécifique portant sur la recherche d'objets composés au sein d'images de fresques médiévales. Les méthodes utilisées pour décrire les différentes régions de la pyramide seront quant à elles présentées au chapitre 6.

Cette nouvelle structure est fondée sur une segmentation d'objets multi-échelles. Le problème de segmentation est un problème complexe. De nombreuses solutions ont été apportées pour des applications spécifiques mais très peu encore pour une division des images en terme d'objets au sens de la vision humaine. Le chapitre suivant se propose d'aborder cette question.

CHAPITRE 5

SEGMENTATION MULTI-ÉCHELLES

Sommaire

5.1 Introduction

La segmentation consiste à diviser une image en différentes régions (ou le problème dual : à rechercher les frontières les délimitant). Les critères pour générer ces régions sont très divers. Généralement ces zones sont obtenues en minimisant un critère d'homogénéité donné. De très nombreuses segmentations d'une image peuvent ainsi être obtenues suivant le critère utilisé. La notion de segmentation idéale est donc difficilement justifiable de manière générale car le *meilleur* partitionnement des images est fonction de l'application visée.

Jusqu'à il y a quelques années, les critères proposés pour la segmentation étaient majoritairement étudiés pour les images en niveaux de gris[HARALI85, PAL93]. Bien que l'information couleur apporte une représentation plus complète des images et par conséquent une segmentation plus fiable, les traitements en couleur étaient beaucoup plus coûteux que ceux en niveaux de gris. Les dix dernières années ont vu une baisse significative des coûts calculatoires et l'apparition de systèmes d'acquisition couleur à des prix très abordables provoquant l'élaboration de très nombreux algorithmes de segmentation couleur.

Dans ce travail, la segmentation est la base de la structure de représentation hiérarchique mise en place (cf. section 4.4). Il est donc nécessaire de disposer d'un algorithme fournissant divers niveaux de segmentation cohérents entre eux mais aussi en terme d'objets au sens de la vision humaine. Les régions obtenues doivent permettre une description précise de la scène de l'image.

Ce chapitre se propose donc de présenter les diverses solutions de segmentation d'images couleur disponibles à l'heure actuelle. Il permet aussi d'exposer l'algorithme retenu et d'en expliquer son choix.

5.2 Méthodes existantes

De manière plus formelle, la définition d'une segmentation peut s'écrire en se basant sur la description d'une *bonne* segmentation d'une image en niveaux de gris par HARALICK et SHAPIRO[HARALI85] : « Les régions d'une image segmentée doivent être uniformes et homogènes suivant différents critères tels que le niveau de gris ou la texture. L'intérieur des régions doit être simple et comporter peu de petits trous. Relativement à l'indice d'uniformité choisi, les régions adjacentes d'une segmentation doivent avoir des valeurs significativement différentes. Les frontières de chaque région doivent être simples, continues et positionnées précisément. » Formellement, en prenant une image I et un prédicat d'homogénéité $\mathcal{H}$, une segmentation de I est une partition $\mathcal{P}$ de I en un ensemble de N_R régions $\mathcal{R}_n$ où $n = 1, \ldots, N_R$ telle que :

- $\bigcup_{n=1}^{N_R} \mathcal{R}_n = I$ avec $\mathcal{R}_n \bigcap \mathcal{R}_m = \emptyset$ si $n \neq m$
- $\mathcal{H}(\mathcal{R}_n)$ est vrai $\forall n \in [1 \ldots N_R]$
- les pixels de $\mathcal{R}_n$ sont connexes $\forall n \in [1 \ldots N_R]$
- $\mathcal{H}(\mathcal{R}_n \bigcup \mathcal{R}_m)$ est faux $\forall \mathcal{R}_n$ et $\mathcal{R}_m$ adjacentes

Ces quatre conditions indiquent que la partition doit couvrir toute l'image, que les pixels formant une région doivent être connexes, que les régions doivent être homogènes suivant le prédicat donné et qu'aucun couple de régions adjacentes ne peut être regroupé pour former une région satisfaisant le prédicat.

Cet état de l'art expose divers systèmes de segmentation couleur les plus fréquemment utilisés. La catégorisation retenue nous amène à présenter d'une part ceux travaillant dans le domaine du critère retenu et, d'autre part, ceux opérant dans le domaine de l'image. Nous verrons ainsi que le premier groupe de méthodes s'adapte mal au contenu de l'image. Le travail de classification des différents pixels s'effectuant dans l'espace du critère, il ne prend pas du tout en compte l'information spatiale de l'image ce qui génère souvent des régions homogènes mais non compactes.

5.2.1 Méthodes utilisant l'espace du critère

Les propriétés majoritairement utilisées sont liées aux notions de couleur et/ou de texture. L'image est alors transformée dans l'espace du critère. Dans le cas d'un histogramme couleur par exemple, tous les pixels de l'image sont projetés dans l'espace colorimétrique à 3 dimensions considéré. L'étude est alors réalisée sur le nuage de points ainsi formé.

De très nombreuses techniques fonctionnent par groupement des pixels au sein de l'espace couleur sans aucun *a priori*. Les clusters formés correspondent alors à une classe. La classification des différents pixels conduit à la formation de régions homogènes en terme de couleur mais génère

logiquement des classes non compactes au sein de l'image. La méthode la plus classique est sans doute l'algorithme des centres mobiles (ou ses variantes les nuées dynamiques et les *k-means*) qui est grandement utilisé tant dans la quantification vectorielle que dans la compression de données. Il est applicable dans n'importe quel espace colorimétrique : PARK[PARK98] l'emploie dans l'espace RVB alors que WEEKS et HAGUE[WEEKS97] le font dans l'espace HSI. Ces méthodes de classification sont non hiérarchiques. Leur principe est la décomposition d'un ensemble d'individus en un nombre n de classes choisi *a priori*. Pour cela, un processus itératif fait converger la sélection des représentants de chaque classe qui ont été préalablement initialisés aléatoirement ou manuellement. La différence principale entre ces algorithmes est la définition même des représentants utilisés et leur processus d'évolution. Une extension floue de cette méthode, le *fuzzy c-means*, est aussi très utilisée en classification[RUI96B]. Le problème de ce type de méthodes est la nécessité de connaître *a priori* le nombre de classes de l'image pour initialiser l'algorithme. La méthode $ISODATA$ (*Iterative Self-Organizing Data Analysis Techniques A*), issue de l'approche des *k-means*, lève cette contrainte est permet l'évolution du nombre de classes au cours du traitement [TAKAHA95].

Une autre technique très répandue pour les images en niveaux de gris est le seuillage (ou partitionnement) d'histogramme[PREWIT66]. Il est facile sur l'histogramme 1D de déterminer les maximum et les minimum locaux à partir desquels les différents ensembles de niveaux de gris sont générés. En couleur, les choses sont un peu plus compliquées car cette détection doit se faire de manière combinée sur trois histogrammes mono-dimensionnels ou au sein d'un histogramme $3D$. De plus, le problème des histogrammes est leur sensibilité au bruit qui génère souvent des fausses crêtes et des faux pics perturbant le processus de segmentation. Diverses techniques sont ainsi utilisées pour déterminer les différents modes présents au sein de l'histogramme telles que la ligne de partage des eaux[WATSON87, SHAFAR98] ou une adaptation d'un modèle des classes comme le réalise SABER en caractérisant les objets par une loi de probabilité gaussienne[SABER95] ou bien encore grâce à des opérateurs flous[GILLET01].

Ces deux premiers types de méthodes n'utilisant pas l'information spatiale des images, elles produisent des segmentations en régions homogènes mais non compactes. Un post-traitement (généralement un étiquettage en composantes connexes) permet de former des régions compactes qui peuvent être intéressantes pour des applications telles que la compression mais qui sont souvent peu adaptées dans le cadre de la description de l'image en objets au sens de la vision humaine. Une première adaptation de ces techniques est basée sur l'algorithme des *k-means* auquel une contrainte spatiale est ajoutée[PAPPAS88, GEMAN84]. En ce sens, ce type de méthodes se situe entre les deux catégories que nous avons définies.

5.2.2 Méthodes utilisant le domaine de l'image

Le fait de vouloir une segmentation en objets nous amène à nous tourner plutôt vers des traitements intégrant l'information spatiale de l'image. Les méthodes de *split-and-merge*[BRAQUE98] considèrent initialement une partition quelconque de l'image. L'image entière est souvent la région initiale unique mais la partition formée de tous les pixels est aussi un point de départ possible. Les régions sont alors divisées (*split*) ou regroupées (*merge*) de multiples fois suivant un critère d'homogénéité donné pour obtenir la segmentation finale. Une structure classique pour ce type de processus est le *quadtree* (cf. section 4.3.3.1). Une fois la phase de division accomplie, de nombreuses régions sont produites ; il suffit alors de les regrouper en garantissant l'homogénéité des

zones ainsi produites pour obtenir la segmentation finale. Le graphe d'adjacence de régions (cf. section 4.2.3.1) est la structure standard pour cette phase de fusion. La rigidité du *quatree* peut aussi être contournée par l'utilisation d'une triangulation de DELAUNAY incrémentale[GEVERS97] ou des diagrammes de VORONOI[SCHETT94] bien que les pyramides semblent être, ces dernières années, fortement plébiscitées pour leur souplesse de manipulation[ZILIAN98].

Une région homogène peut également être obtenue par un système de croissance de régions (*Region growing* en anglais)[TRÉMEA97, CRAMAR97]. À partir de germes présélectionnés, les pixels sont progressivement agglomérés tout en conservant un certain critère d'homogénéité. À la suite d'un tel processus, il existe souvent de petites régions qui n'ont pas réussie à s'étendre ou des régions similaires adjacentes qui ont été obtenues par deux germes différents ; par conséquent, un post-traitement est nécessaire pour corriger ces erreurs.

D'autre part, l'approche duale consistant à détecter les frontières bordant les régions est possible. Il est bien connu en niveau de gris que les contours peuvent être trouvés en utilisant les fonctions gradient ou laplacienne des images. En couleur, il faut donc définir une mesure prenant en compte les trois composantes[DIZENZ86, CHAPRO97] (approche vectorielle) ou bien combiner les trois mesures des différents canaux[CARRON94] (approche marginale). Malheureusement, ces méthodes garantissent rarement la création de contours fermés délimitant les différents objets. Les modèles déformables (*snakes* ou contours actifs) font aussi partie de ce type de méthodes [WONG98] même si celles-ci sont plutôt utilisées pour la détection d'objets bien définis que pour segmenter entièrement une image. Ces méthodes consistent à déformer un contour initial vers les contours de l'objet à détecter. Cette évolution est guidée par la définition de deux énergies : une énergie interne garantissant la cohésion de la forme et une énergie externe définie de manière à attirer le contour actif vers le bord de l'objet.

Enfin, les approches variationnelles telles que les Équations aux Dérivées Partielles (EDP) forment un ensemble intéressant de méthodes permettant de construire des segmentations multi-échelles. Ces algorithmes sont généralement développés en niveaux de gris[PERONA90] mais une partie d'entre eux a été étendue en couleur[TSCHUM02]. Les techniques qui nécessitent un filtrage par diffusion, emploient souvent un espace-échelles linéaire[GAUCH99]. D'un autre côté, une diffusion non linéaire est généralement utilisée pour prétraiter les images avant divers algorithmes de segmentation[WEICKE98]. IRIS VANHAMEL propose ainsi une approche originale basée sur une ligne de partage des eaux du gradient des images couleurs[VANHAM03]. Une fois l'espace-échelles généré à partir d'une mesure de gradient couleur, une ligne de partage des eaux est réalisée sur les différentes échelles pour générer des régions. La notion de multi-échelles permet alors de lier les différents niveaux les uns avec les autres et de permettre la régénération d'une hiérarchie de segmentations cohérente.

5.2.3 Choix de la méthode

L'algorithme de segmentation dont nous avons besoin doit :
- générer un ensemble de niveaux de segmentation d'une image allant d'une partition grossière délimitant les différents objets à une segmentation fine séparant les objets en parties ;
- donner des segmentations les plus cohérentes possible par rapport à la notion d'objets.

La contrainte de cohérence en terme d'objets est sûrement la plus compliquée à satisfaire. D'une part, le critère utilisé doit permettre la détection des objets ce qui élimine toutes les méthodes

travaillant dans l'espace du critère étant donné l'importance de la compacité et de la cohérence spatiale des régions produites. D'autre part, il doit être suffisamment polyvalent pour répondre à la majorité des images. De ce fait, les critères basés uniquement sur une source d'information ne semblent pas assez complets ; il paraît plus intéressant d'opter pour une association de critères comme la couleur et la texture.

Les méthodes de *split-and-merge* et surtout celles basées sur les pyramides semblent adaptées à notre approche car elles sont directement intégrables à notre structure (cf. section 4.4). Malheureusement, il paraît difficile de définir les niveaux optimaux à utiliser dans le graphe pyramidal. La pyramide de segmentation fournit ainsi un grand nombre de partitions différentes. Par contre, le choix des niveaux à utiliser par la suite semble hasardeux. De surcroît, avec ce type de méthodes, la manipulation de petites régions est gênante si un critère texture est utilisé (et même couleur parfois). Sur des régions de quelques pixels les indices textures ne sont plus pertinents et les regroupements éventuels fortement aléatoires.

La méthode mise en place est en fait une méthode de croissance de régions travaillant sur un critère de rupture de textures proposé par DENG[DENG99]. Cet indice caractérise de manière fine les frontières des régions. De plus, ce critère permet de définir plusieurs niveaux représentant chacun les frontières plus ou moins franches entre les régions ce qui rend possible la génération de plusieurs segmentations cohérentes entre elles. Les résultats obtenus sont encore loin d'être parfaits mais fournissent un bon point de départ pour le reste de l'étude qui consiste à décrire les différentes régions (cf. chapitre 6) puis à comparer les représentations obtenues pour mesurer la ressemblance entre images ou pour rechercher des objets (cf. chapitre 7).

5.3 Segmentation JSEG

5.3.1 Présentation

L'algorithme de segmentation JSEG est un algorithme fondé sur la définition d'un critère de rupture de modèle textural proposé par DENG[DENG99]. Cet indice fournit une quantification de l'uniformité de la texture. Une fois ce critère calculé pour l'ensemble de l'image, la génération de la segmentation est réalisée par un système d'accroissement de régions utilisant une ligne de partage des eaux. En utilisant plusieurs tailles de voisinage, il est alors possible de détecter les ruptures de textures plus ou moins marquées. Cet aspect nous permet de mettre en place un système multi-échelles améliorant la qualité des segmentations obtenues et d'obtenir des niveaux de segmentation plus ou moins fins de l'image.

Nous présenterons dans cette section les différentes étapes de cet algorithme : définition du critère de rupture de modèle textural, quantification, accroissement des régions par ligne de partage des eaux, système multi-échelles et regroupement des régions. Les divers algorithmes détaillés peuvent être trouvés en annexe A.

5.3.2 Calcul du coefficient d'uniformité texturale

Les différentes couleurs de l'image sont considérées comme des classes. Une classe est composée d'un ensemble de pixels de même couleur. L'étude texturale est alors réalisée sur cette carte de classe.

DENG[DENG99] propose le critère de rupture de modèle textural suivant. Considérons un ensemble de N points de la carte des classes nommé Z. Comme nous le verrons plus tard, Z correspondra à une fenêtre de voisinage de taille restreinte très faible par rapport à celle de l'image. Soit $z = (x, y)$ avec $z \in Z$ et m le centre de gravité de Z :

$$m = \frac{1}{N} \sum_{z \in Z} z$$

Supposons Z divisé en C classes où Z_i est l'ensemble des N_i pixels de la i^{e} classe avec $i = 1, \ldots, C$. Soit m_i le centre de gravité de Z_i,

$$m_i = \frac{1}{N_i} \sum_{z \in Z_i} z, \forall i \in [1 \ldots C]$$

La dispersion (variance) totale des points de Z est définie par :

$$D_t = \sum_{z \in Z} \parallel z - m \parallel^2$$

De même, la dispersion relative aux centres des classes s'écrit :

$$D_r = \sum_{i=1}^{C} D_i = \sum_{i=1}^{C} \sum_{z \in Z_i} \parallel z - m_i \parallel^2$$

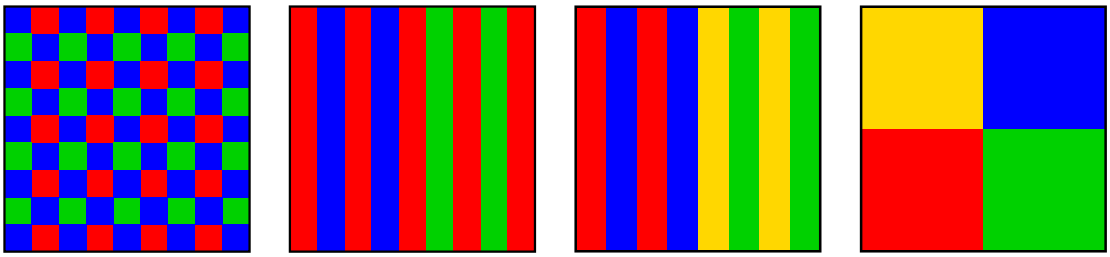
où D_i est la variance de la classe i.

Le critère J défini pour décrire la rupture du modèle textural est alors :

$$J = \frac{D_t - D_r}{D_t}$$

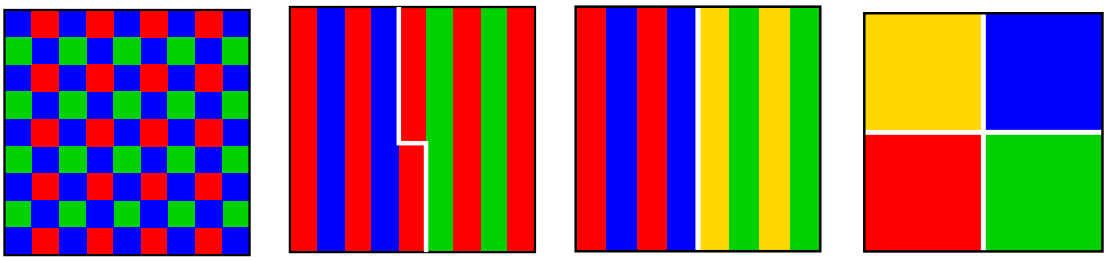
Ce critère sera très faible si la somme des dispersions des classes est égale à la dispersion totale ce qui correspond aux cas où le centre de gravité global et ceux des différentes classes sont très proches. J est d'ailleurs nul si tous les centres sont confondus. La figure [5.1] propose quelques exemples de calculs de J sur une fenêtre carrée. La valeur de J croît de gauche à droite avec la désorganisation texturale.

DENG propose plutôt une normalisation par D_r. L'utilisation de D_t que nous proposons permet de borner J entre 0 et 1 (car $D_r < D_t$ dans tous les cas) et de cette manière de rendre la comparaison possible entre deux valeurs. En effet, si la forme de deux ensembles Z_1 et Z_2 est la même alors les valeurs D_t seront identiques et les valeurs de J pourront être confrontées. La figure [5.6].d montre que ce changement de normalisation n'engendre pas de modification flagrante ; dans les deux cas, les cartes de J sont semblables. La normalisation par D_t fournit d'ailleurs une carte légèrement plus lisse.



$D_t = D_r$ $J = 0$ | $D_t > D_r$ $J = 0.1333$ | $D_t >> D_r$ $J = 0.4048$ | $D_t >> D_r$ $J = 0.7619$

FIG. 5.1 – Calcul de J dans un voisinage



$\bar{J} = 0$ | $\bar{J} = 0.0107$ | $\bar{J} = 0.0385$ | $\bar{J} = 0$

FIG. 5.2 – Calcul de $\bar{J}$ en restreignant le calcul de J aux régions définies par les lignes blanches

Une région est donc uniforme au niveau textural si sa valeur de J est faible. Sur la figure [5.1], l'exemple de gauche est une région à part entière, celles du milieu sont composées de deux régions avec une frontière verticale et celui de droite correspond en réalité à 4 régions distinctes. En restreignant le calcul de J aux régions définies sur la figure [5.2], la somme des J des différentes régions tend alors vers 0. Par cette simple constatation, une segmentation intéressante de l'image apparaît comme celle qui minimise la moyenne des J des différentes régions. De cette manière, si J est calculé pour chaque région et non pour la zone entière, la moyenne pondérée $\bar{J}$ des J, peut être définie par :

$$\bar{J} = \frac{1}{N} \sum_k M_k J_k$$

où J_k est la valeur de J calculée sur la région k et M_k le nombre de pixels de cette région. $\bar{J}$ semble donc un critère d'évaluation intéressant de la qualité de la segmentation.

5.3.3 Quantification

Le critère d'uniformité texturale présenté dans la section précédente est calculé sur une image de classes des couleurs. Or, les images comportent généralement un très grand nombre de couleurs dont certaines sont très proches. Si, l'image brute est utilisée, les couleurs proches seront donc considérées comme des classes différentes et l'analyse texturale ne sera pas visuellement correcte. Il est donc nécessaire de réaliser au préalable une quantification de l'image pour regrouper les couleurs similaires dans une même classe et ainsi rendre le calcul de J plus pertinent.

La notion de couleur n'intervient que lors de cette étape préliminaire. Ce prétraitement est nécessaire à la qualité de l'algorithme général car, par la suite, la notion de couleur n'intervient plus ; les différentes classes générées seront considérées comme équivalentes et ne porteront plus d'information colorimétrique. Il est d'ailleurs important que des couleurs proches soient représentées par une même classe pour éviter la détection de ruptures de textures perceptuellement peu visibles lors de l'analyse. La quantification influe donc fortement sur le processus de segmentation. Par conséquent, une attention toute particulière est à apporter à la qualité des quantifications obtenues. L'idée est donc d'extraire de l'image les quelques couleurs représentatives des régions de l'image. Généralement, 10 à 20 couleurs suffisent pour différencier les objets d'une scène naturelle.

La quantification utilisée est celle d'*ImageMagick*[1]. Elle fonctionne par élagage d'un arbre des couleurs pavant l'espace colorimétrique. Cet algorithme est du même type que les algorithmes *octree*[GERVAU90]. La réduction de l'arbre est réalisée de manière à minimiser l'erreur entre l'image originale et l'image quantifiée jusqu'à obtenir le nombre de couleurs voulu. Les résultats obtenus sont très satisfaisants visuellement (cf. figure [5.3]). Bien entendu, d'autres méthodes de quantification peuvent être utilisées (cf. section 6.2.1). Les différents tests réalisés avec divers algorithmes (*median cut*, méthode des palettes locales...) n'ont pas montré d'améliorations flagrantes et constantes des résultats obtenus. Par rapport à la qualité de la segmentation, les diverses expérimentations ont montré que l'utilisation d'un nombre trop important de couleurs engendrait une analyse de texture non pertinente. Suivant la nature des images traitées, il semble intéressant de choisir un nombre de couleurs compris entre 8 (pour des images spécifiques) et 16 (pour des scènes naturelles).

L'algorithme de quantification choisi peut travailler dans les espaces couleur RVB, YIQ et Yuv. Aucune diffusion n'est appliquée à la suite de cette quantification même si visuellement les résultats en seraient améliorés. Un tel post-traitement créerait des textures artificielles non présentes dans l'image originale.

5.3.4 Génération des régions par ligne de partage des eaux

Obtenir une faible valeur de $\bar{J}$ sur l'ensemble de l'image n'est pas chose facile. Par contre, le calcul de J au sein d'un voisinage Z de la carte des classes fournit un bon indicateur sur la position du voisinage pour déterminer s'il se trouve à l'intérieur ou près d'un bord d'une région. L'idée est alors de construire une carte de J sur toute l'image. Pour cela, le coefficient J est calculé pour chaque pixel de l'image sur un voisinage donné (cf. figure [5.4]). Il apparaît sur ce type d'images que les valeurs importantes de J correspondent à des pixels proches des frontières des régions. Le fait d'utiliser un voisinage de grande taille permet de détecter les ruptures marquées et étendues du modèle textural alors qu'un voisinage plus petit rend compte des variations de texture plus fines et localisées.

Toutefois, un voisinage étendu augmente les temps de calcul de la carte des J. Pour rendre ce temps constant, un échantillonnage du voisinage en fonction de sa taille est possible. Le nombre de points pris en compte peut alors être indépendant de l'échelle étudiée (cf. figure [5.5]). Cette optimisation réduit le temps de création des cartes de J de niveaux supérieurs sans réellement réduire la qualité de l'information calculée (cf. figure [5.6]). Un lissage postérieur de la carte des J est tout de même nécessaire pour diminuer le bruit généré par l'échantillonnage. Nous avons utilisé

[1] voir `http://studio.imagemagick.org/www/quantize.html` pour plus de détails.

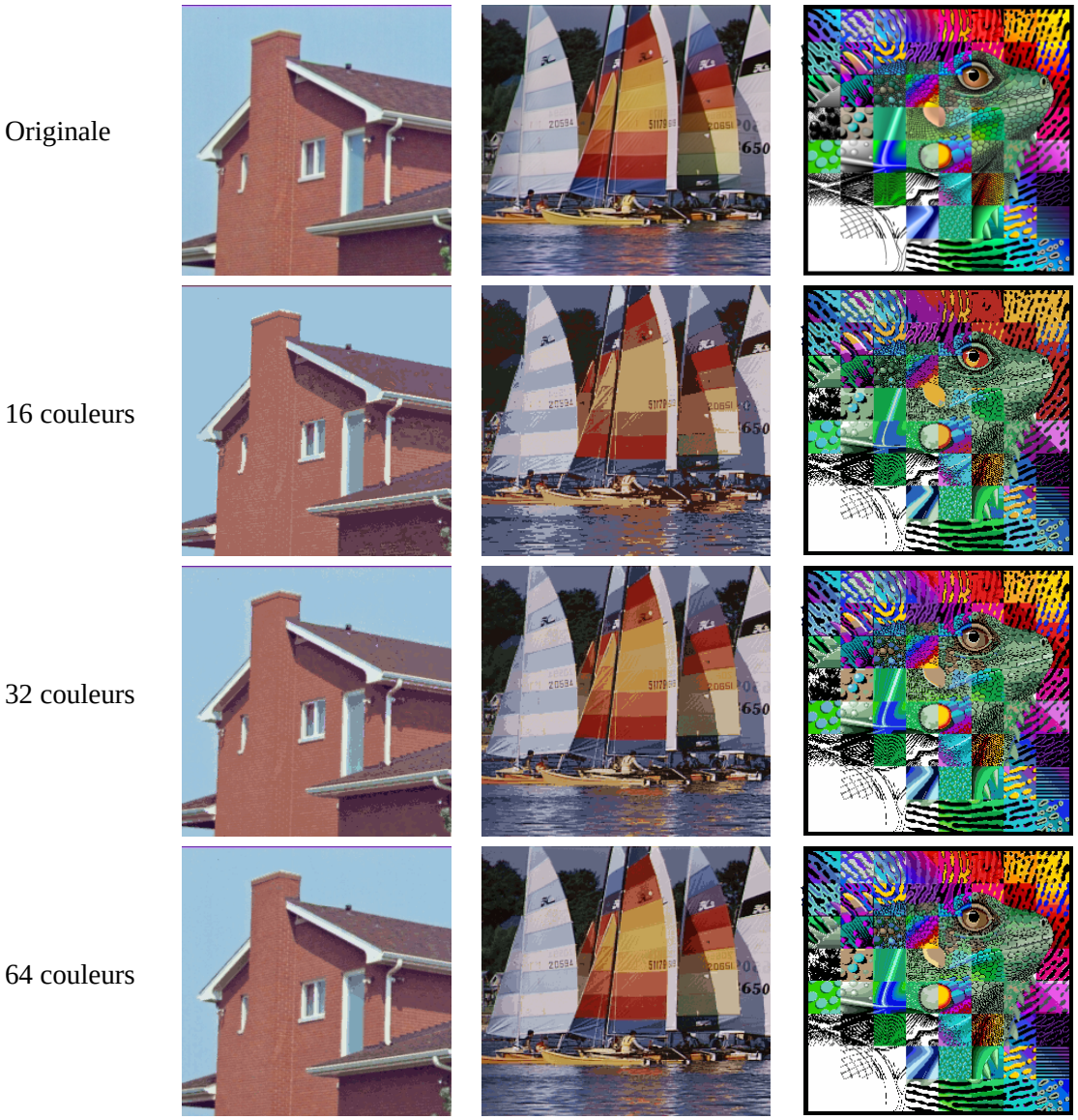
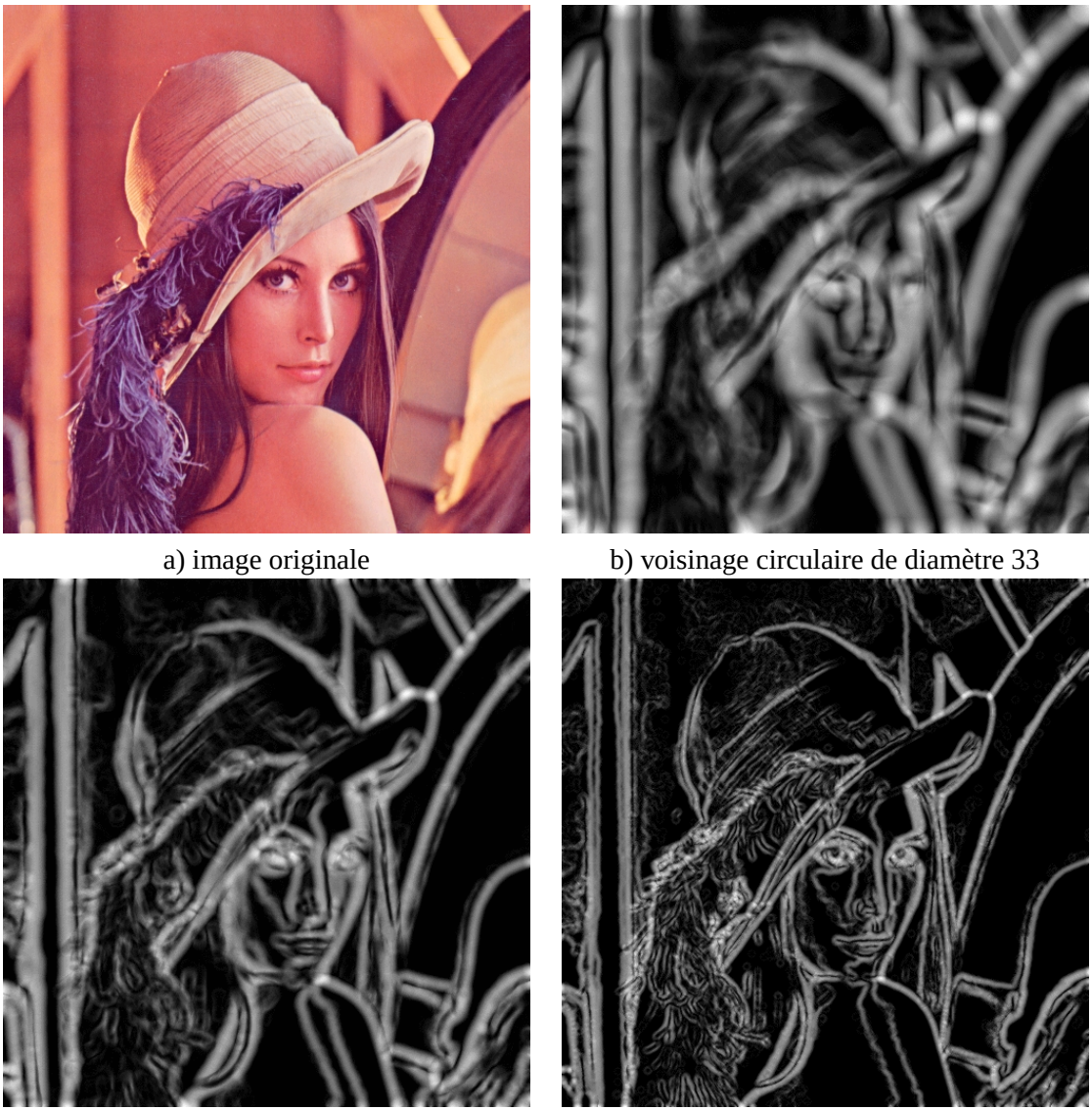


FIG. 5.3 – Quantifications obtenues avec l'algorithme d'*ImageMagick*

pour cela un moyennage en utilisant une fenêtre carrée de taille suffisante nous permettant de nous ramener à une interpolation linéaire du relief (3×3 dans l'exemple de la figure [5.5].b).

Les différentes cartes de J apportent des informations complémentaires nécessaires pour segmenter les images correctement. Ces images peuvent aussi être vues comme des reliefs topographiques contenant des crêtes représentant les frontières entre les régions et des vallées correspondant à l'intérieur de celles-ci. Par conséquent, des méthodes de type croissance de régions peuvent être utilisées pour générer la segmentation. En effet, considérons un voisinage mobile positionné sur une valeur faible de J ; il est donc situé sur une zone de texture uniforme. Un déplacement du masque vers un pixel voisin de faible J assure une texture toujours uniforme et un recouvrement important par rapport à sa position initiale. Ces deux aspects permettent de dire que nous sommes probablement au sein de la même texture pour les deux positions du masque et donc dans la même

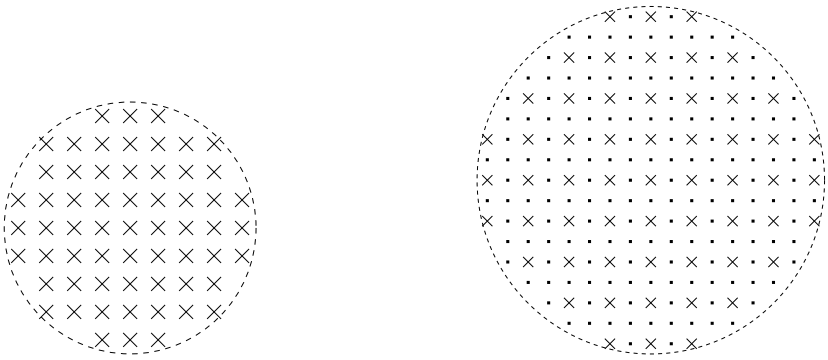


a) image originale b) voisinage circulaire de diamètre 33

c) voisinage circulaire de diamètre 17 d) voisinage circulaire de diamètre 9

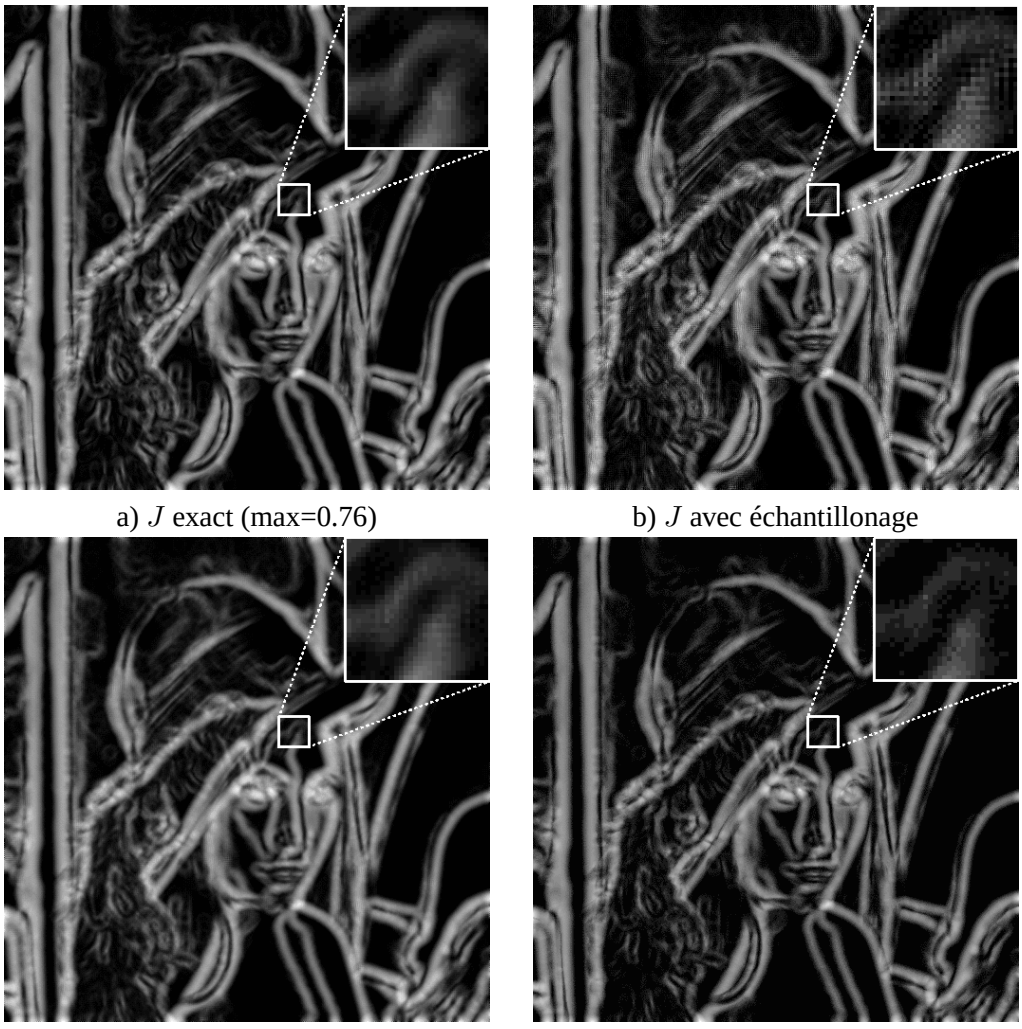
FIG. 5.4 – Calcul de J pour des différents voisinages circulaires (pour 16 couleurs)
Les valeurs sont comprises entre 0 (en noir) et 1 (en blanc).

région ($\bar{J}$ sera donc faible). L'itération de ce processus correspond alors à un système de croissance de régions. Un système de ligne de partage des eaux a été mis en place pour générer les régions à partir d'une carte de J. De plus, l'utilisation combinée des différentes cartes de J permet d'améliorer la qualité de la segmentation finale grâce à un système de raffinement multi-échelles. Ce dernier dispositif permet d'ailleurs de disposer de segmentations plus ou moins fines d'une même image ce qui est intéressant pour notre structure de représentation hiérarchique.



a) Voisinage de référence de taille 9×9 b) Voisinage échantillonné de taille 17×17

FIG. 5.5 – Exemple d'échantillonnage d'un voisinage circulaire
Les croix correspondent aux pixels pris en compte et les points à ceux qui ne le sont pas.



a) J exact (max=0.76) b) J avec échantillonage

c) J avec échantillonage et filtrage d) J avec normalisation par D_r (max=2.76)

FIG. 5.6 – Différentes méthodes de calcul du coefficient J sur une image (voisinage 17×17)

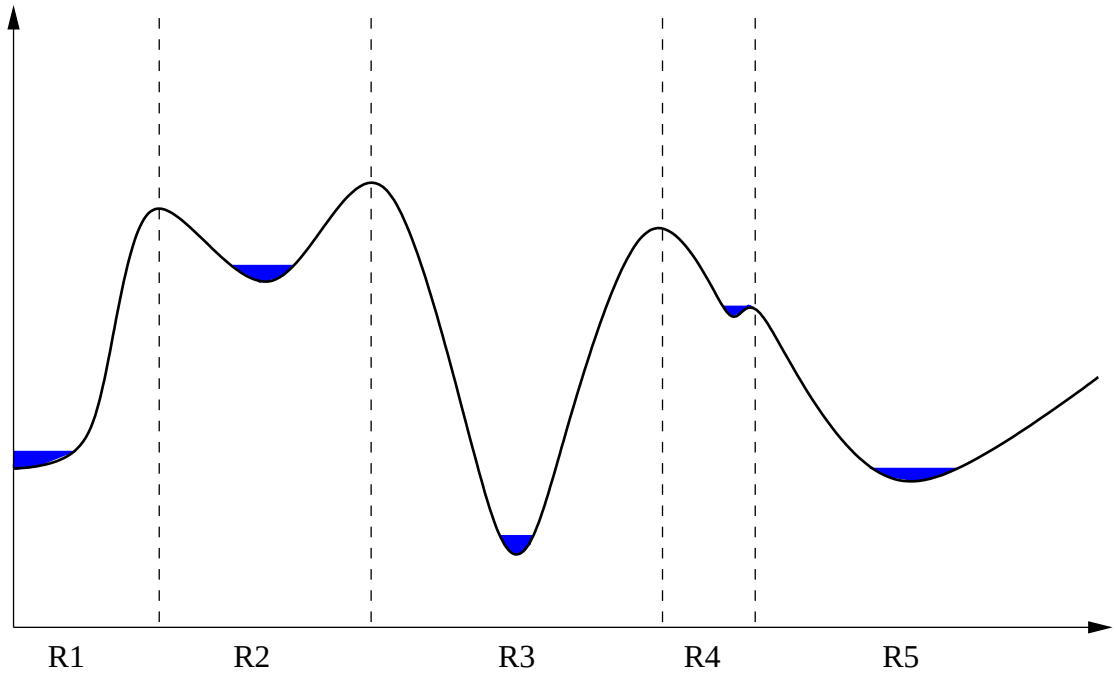


FIG. 5.7 – Minima locaux et régions générées par ligne de partage des eaux
Les frontières sont symbolisées en pointillés verticaux et les minima locaux comme des lacs au niveau des vallées du relief.

5.3.4.1 Présentation générale des algorithmes de ligne de partage des eaux

La ligne de partage des eaux est une technique classique de segmentation d'images en niveaux de gris. L'image est considérée comme un relief topographique dont les valeurs correspondent à l'altitude. La segmentation se rapporte alors à un problème hydrologique : une goutte d'eau tombant sur ce relief va s'écouler jusqu'à un minimum local. Les pixels d'une même région sont alors ceux à partir desquels une goutte d'eau s'écoule vers le même lieu ; cette zone est aussi appelée *bassin de capture*. Un exemple de relief et des régions obtenues par un tel algorithme est présentée sur la figure [5.7]. Les lignes pointillées verticales symbolisent la frontière entre les différents bassins ; elles sont appelées *ligne de partage des eaux* ou en anglais *watershed*.

Généralement, ce processus n'est pas appliqué directement sur l'image mais plutôt sur l'image d'une fonction d'énergie telle que la norme du gradient. Cette fonction doit être définie pour avoir une valeur élevée aux limites des régions et une valeur faible en leurs centres. De cette manière, les minima locaux forment l'ensemble des bassins qui s'agrandissent au cours du processus pour tendre vers les frontières où le critère est fort.

Les premiers algorithmes mis en place pour générer les différents bassins de capture étaient fondés sur le principe original et tentaient de modéliser l'écoulement le long du relief. Malheureusement, cette méthode est complexe à implanter et l'écoulement est mal défini sur des reliefs à valeurs discrètes. Pour pallier ces problèmes de coûts et de fiabilité, le principe d'immersion voit le jour dans les années 1990. L'idée générale est d'immerger progressivement le relief dans l'eau (cf. figure [5.8]). Pour qu'il puisse se remplir, il suffit de percer le relief au niveau des bassins de capture. Des puits sont donc creusés au niveau de chaque minimum local et la surface est

ensuite plongée progressivement dans l'eau. Pour séparer les différents bassins, des barrages sont construits lorsque des zones inondées se rejoignent. Une fois le relief totalement immergé, ces barrages symbolisent la ligne de partage des eaux.

Sur l'exemple proposé, le problème de la définition des minima locaux se manifeste avec la création d'un bassin B4 qu'il paraît plus naturel de regrouper avec B3. Ce phénomène est dû à des minima locaux non fondamentaux souvent dus au bruit ou représentant des structures mineures de l'image. La sur-segmentation obtenue peut être limitée par l'utilisation de marqueurs. Au sein d'une scène à analyser il est souvent possible de trouver des marqueurs à l'intérieur des objets à reconnaître. Malheureusement, leur création automatique est un problème non trivial. De cette manière, les minima locaux non significatifs peuvent être éliminés, par exemple ceux dus à une faible variation de gradient.

De nombreuses implémentations de la ligne de partage des eaux pour la segmentation ont vu le jour[VINCEN91, MEYER90]. Généralement, les pixels de l'image sont classés dans une file d'attente triée suivant la valeur de leur niveau de gris. L'immersion est alors réalisée par paliers successifs où les différentes régions sont élargies suivant les régions déjà existantes et la présence possible de nouveaux marqueurs immergés.

5.3.4.2 Notre implémentation de la ligne de partage des eaux

Généralement, les problèmes rencontrés lors de l'implémentation d'algorithme de ligne de partage des eaux sont dus à la présence de plateaux au du relief. Le fait que de nombreux pixels aient la même valeur engendre alors une ambiguïté lors du leur tri ; elle peut tout de même être contournée par la définition d'un facteur de priorité. Dans notre cas, il n'existe quasiment jamais cet effet de plateau ; il est donc beaucoup plus facile de mettre en place un système simple d'accroissement de régions.

Détection des marqueurs Nous réalisons tout d'abord une détection de marqueurs au sein même de la carte des J. Si les centres des régions sont considérés comme des zones de texture uniforme, alors ils sont représentés par des faibles valeurs de J. Par conséquent, une méthode de seuillage paraît adaptée à leur détection. Pour obtenir un bon résultat rapidement, différents seuils fonction de la distribution des J sont utilisés. Les seuils utilisés sont fonction de la moyenne μ_J et de variance σ_J des J. Pratiquement, 6 seuils différents sont pris en compte : $\mu_J + A * \sigma_J$ où $A = \{-0.6; -0.4; -0.2; 0; 0.2; 0.4\}$. Ce principe de seuillage multiple réduit les temps de calcul qu'aurait engendré un seuillage adaptatif. Il a montré expérimentalement son efficacité. De plus, comme nous le verrons plus en détail dans la présentation du système multi-échelles, ce seuillage est réalisé au sein des régions déjà détectées ce qui améliore la qualité de cette détection.

Pour chaque seuil, les pixels ayant une valeur de J inférieure au seuil sont considérés comme des vallées potentielles. Un étiquetage en composantes connexes est alors réalisé et permet de supprimer les vallées trop petites. Empiriquement, les vallées sont considérées comme trop petites si le nombre de pixels les constituant est inférieur à la taille du masque utilisé pour calculer les coefficients J. Le seuil finalement retenu est celui qui fournit le plus grand nombre de vallées (cf. figure [5.9]).

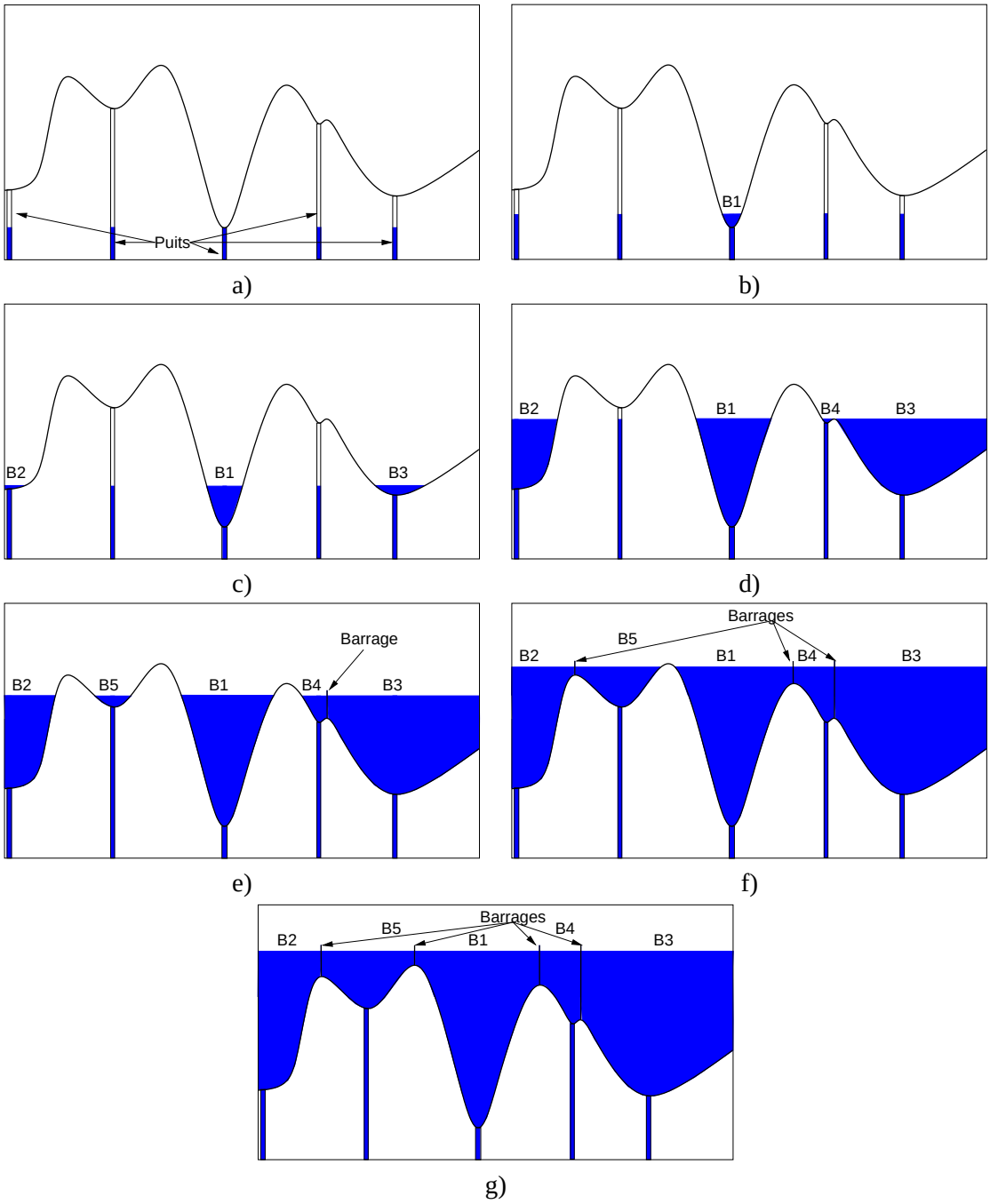
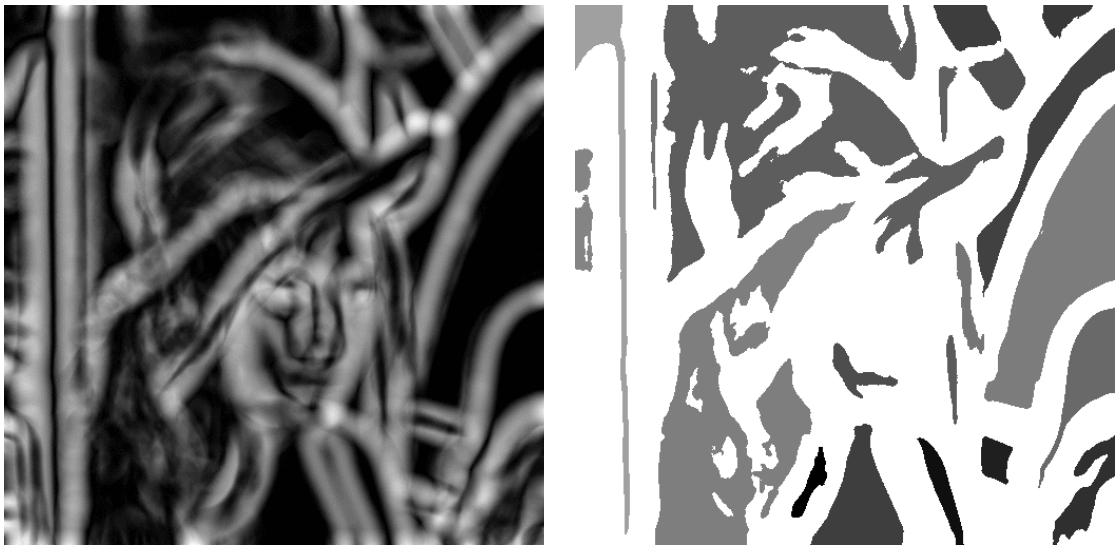


FIG. 5.8 – Principe d'immersion d'un relief

a) L'immersion commence, les puits se remplissent.
b) Un premier bassin B1 commence à se former.
c) Deux autres bassins B2 et B3 se créent.
d) Après la création du quatrième bassin B4, B3 et B4 tendent à se regrouper.
e) Le dernier bassin apparaît. Un barrage est mis en place pour séparer B3 et B4.
f) La montée des eaux continue et les barrages se font de plus en plus nombreux.
g) L'état final est atteint quand toute la surface est recouverte. Les barrages forment les lignes de partage des eaux et les bassins les régions.



Carte de J considérée Vallées détectées

FIG. 5.9 – Détection des vallées à partir d'une carte de J

Accroissement des vallées Une fois les vallées détectées, l'accroissement des régions est réalisée par une technique de ligne de partage des eaux. Les vallées sont considérées comme des marqueurs internes aux régions. De par la nature non discrète des valeurs traitées, il est possible de mettre en place un système de croissance de régions simple à partir des valeurs de la carte des J. Posons NA l'ensemble des pixels encore non affectés à une région, NA_a le sous-ensemble de NA composé des pixels adjacents à une région et NA_i celui des pixels isolés. L'égalité $NA = NA_a \bigcup NA_i$ est alors vraie. La croissance est réalisée itérativement par affectation du pixel de NA_a de plus petite amplitude. À chaque affectation, les ensembles NA_a et NA_i sont mis à jour. Le processus se termine quand tous les pixels sont affectés c'est-à-dire quand NA est vide. L'évolution d'un tel algorithme est exposé sur la figure [5.10] et montre l'expansion des vallées sur un exemple concret. Les minima locaux non considérés comme des marqueurs sont ainsi submergés progressivement. Temporairement, le niveau des différentes lignes de partage des eaux peut ne pas être constant.

Au niveau calculatoire, une optimisation possible de cette méthode consiste à ne plus effectuer une affectation point par point mais par groupe de points. Ce groupe est alors composé des pixels de plus faibles J mais sa taille peut être définie de nombreuses façons. Une première solution consiste à former l'ensemble des points à affecter comme un pourcentage de la population de NA_a. Une autre approche prend en compte la distribution des valeurs des points de NA_a. En fonction de cette distribution, un seuil déterminant les points à affecter peut être calculé. Ce type d'optimisation permet de réduire le temps de calcul mais peut produire des segmentations légèrement différentes. Les différences ne sont visibles qu'au niveau de la position finale de la ligne de partage des eaux qui peut être décalée d'un pixel mais l'ensemble des régions formées reste stable.

5.3.5 Système multi-échelles

La possibilité de calculer des cartes de J avec des voisinages plus ou moins étendus permet de détecter des ruptures de textures plus ou moins marquées au sein de l'image. Cette propriété peut

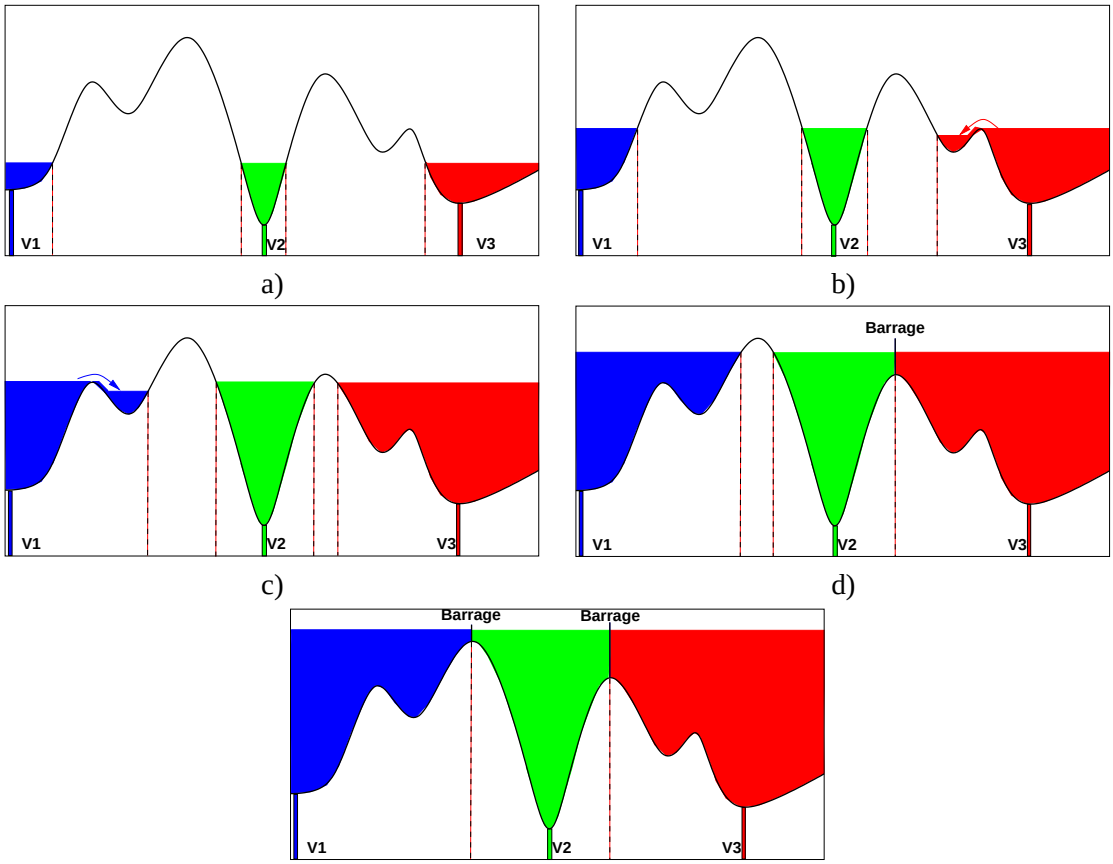


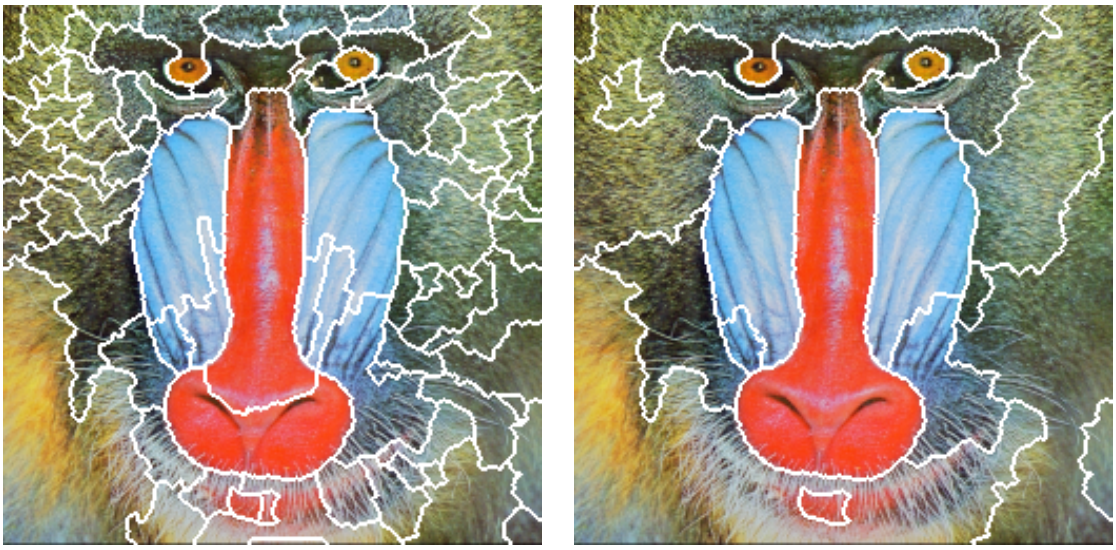
FIG. 5.10 – Principe d'accroissement des régions mis en place
a) Les vallées considérées commencent à se remplir.
b) La vallée $V3$ remplit un minimum local.
c) La vallée $V1$ remplit un minimum local.
d) Un barrage se construit entre $V2$ et $V3$.
e) État final de l'algorithme. Les différentes régions sont formées et séparées par la ligne de partage des eaux symbolisée par les barrages.

être utilisée pour mettre en place un système multi-échelles de segmentation. Les objets principaux de l'image peuvent être segmentés à partir d'une carte de J calculée avec un voisinage étendu. Leur composition est alors détectable en utilisant une carte obtenue avec un voisinage plus petit. De cette manière, en itérant le processus de découpage, les plus petites structures de l'image sont extraites progressivement.

Par ce processus, la détection des vallées est affinée. Pour cela, le seuillage est effectué sur chaque région de la segmentation déjà définie et non sur l'image entière. Les divers seuillages sont effectués en fonction des statistiques (de J) des différentes régions déjà extraites. Cette détection localisée est beaucoup plus efficace que l'étude globale car le seuillage devient adaptatif spatialement avec cette méthode. Pour rendre le traitement générique, le calcul des vallées du premier niveau est effectuée en considérant l'image entière comme une région unique ; pour utiliser la détection localisée des vallées, nous pouvons aussi envisager d'itérer cette première détection.

FIG. 5.11 – Repositionnement des niveaux supérieurs par réaffectation
De haut en bas, des niveaux supérieurs aux niveaux inférieurs.

Enfin, le placement des frontières sur les niveaux utilisant de larges voisinages n'est pas parfait alors que celui d'un niveau à voisinage réduit est très précis. Une fois les segmentations de chaque niveaux obtenues, les échelles supérieures sont alors régénérées par projection des niveaux inférieurs. Une région est repositionnée comme l'ensemble des régions du niveau inférieur se projetant majoritairement dans celle-ci. La figure [5.11] présente les segmentations initialement obtenues et les régions corrigées. Ce repositionnement des niveaux supérieurs rend aussi les différents niveaux complètement cohérents les uns avec les autres.



Segmentation initiale Après regroupement

FIG. 5.12 – Exemple de regroupement de régions en post-traitement de la segmentation

5.3.6 Regroupement de régions

Au cours de la segmentation, des régions peuvent être générées alors que visuellement elles n'ont pas lieu d'être. Un post-traitement consistant à regrouper les régions adjacentes semblables suivant un critère donné peut corriger ces erreurs. Ces fusions peuvent être basées, par exemple, sur l'histogramme couleur des différentes régions ou bien sur toute autre statistique.

Nous n'avons pas utilisé un tel regroupement dans la version de l'algorithme utilisé pour générer les diverses segmentations présentées dans ce chapitre (exceptée la figure [5.12]). De plus, les segmentations produites doivent être intégrées dans une structure de représentation hiérarchique, ces regroupements pourront donc être réalisés de manière plus efficace au sein de cette structure. Quoiqu'il en soit, il est toujours très hasardeux de réaliser ce type de regroupement sur le seul fait d'un indice statistique. En effet, même si les régions apparaissent comme semblables elles peuvent tout de même correspondre à des objets ou à des parties d'objets différents. Inversement, des regroupements peuvent ne pas être réalisés à cause de statistiques non semblables alors que visuellement le regroupement paraît naturel.

5.3.7 Résultats

La plupart des images utilisées ici sont issues de plusieurs bases d'images qui sont toutes présentées an annexe B.

Les segmentations obtenues (cf. figures [5.12], [5.15], [5.16], [5.17], [5.18] et [5.19]) sont de bonnes qualités dans l'optique d'une description des images en terme d'objets. Ces diverses segmentations sont obtenues avec une quantification préalable des images avec un nombre de couleurs déterminé automatiquement pouvant varier de 10 à 20. Aucun regroupement de régions n'a été effectué pour ne pas fausser les résultats (exceptée la figure [5.12]).

Ces résultats montrent que l'algorithme est très efficace lorsque l'image est composée de zones étendues de texture uniforme. Par contre, la présence de régions de même couleur que le fond et uniquement délimitées par un trait fin est difficilement détectée (l'épée sur la figure [5.19]). Au même niveau, de très petites structures ne sont pas détectées à cause des divers seuils (surtout la taille minimum des vallées) que nous avons volontairement choisi importants. Il peuvent bien sûr être modifiés pour que ces petits objets soient segmentés ; en contrepartie, de nombreuses régions indésirables seront générées.

Plus précisément, le problème du regroupement des régions est présenté sur la figure [5.12]. Les fusions de régions ont été réalisées à partir des histogrammes couleur des différentes régions. Si l'intersection des histogrammes est inférieure à un seuil prédéterminé, les régions sont fusionnées. La segmentation finale est alors obtenue quand tous les regroupements possibles ont été réalisés. Finalement, il apparaît que des fusions naturelles ont été effectuées mais que certaines ne l'ont pas été (par exemple les 3 petites régions restant isolées en haut à gauche). En fait, ce type de processus est assez difficile à régler. En effet, quel que soit la statistique choisie, quelle soit plus ou moins performante, la détermination du seuil de fusion n'est pas triviale et engendre toujours des décisions plus ou moins hasardeuses aux alentours du seuil.

Les résultats sur les images standards montrent que les divers objets de l'image sont bien isolés au cours du processus de segmentation (cf. figure [5.17]). Par contre, comme le montre la dernière segmentation des images *baboon* et *lena*, la sur-segmentation de l'image est assez fréquente. De tels cas peuvent être limités d'une part par la non-utilisation de tels niveaux et d'autre part par un post-traitement de fusion. La taille de l'image ou le nombre de régions détectées peuvent être des facteurs utilisés pour déterminer la cohérence potentielle ou *a posteriori* de la segmentation à un niveau donné. Pour une image 512×512 (*lena* par exemple), la taille minimum des régions détectées avec un voisinage 9×9 est de l'ordre de 0.01% de l'image. Même si les vallées croissent durant le processus, certaines ne vont pas atteindre plus de 0.5% de l'image. Cela ne paraît pas significatif et surtout ne correspond pas à un objet que l'œil humain aurait isolé. Les niveaux utilisables peuvent donc être déterminés empiriquement en fonction de la taille de l'image. Ainsi, nous utilisons toujours au minimum 2 niveaux. Ensuite le nombre de niveaux, nn, est déterminé en fonction de la taille T de l'image par $T < 2^{6+nn}$. Pour image 128×128, 2 niveaux seront utilisés, 3 si la taille de l'image est 256×256 et 4 pour 512×512.

5.4 Conclusion

L'algorithme de segmentation JSEG est donc fondé sur un critère de rupture de textures J. Sa minimisation directe au sein d'une image étant complexe, un procédé utilisant une ligne de partage des eaux est utilisé pour générer les différentes régions. De plus, un système multi-échelles amène une solution de raffinement des segmentations qui améliore la détection des objets et des parties les composants. Le schéma de principe de cet algorithme de segmentation est présenté à la figure [5.13] et les divers paramètres utilisés sur lafigure [5.14].

Les divers algorithmes détaillés des différentes parties de ce processus se trouvent en annexe A.

Les résultats obtenus, sans être parfaits, apportent une information intéressante sur la composition des images. Certains objets sont mal délimités, des petites structures ne sont pas détectées

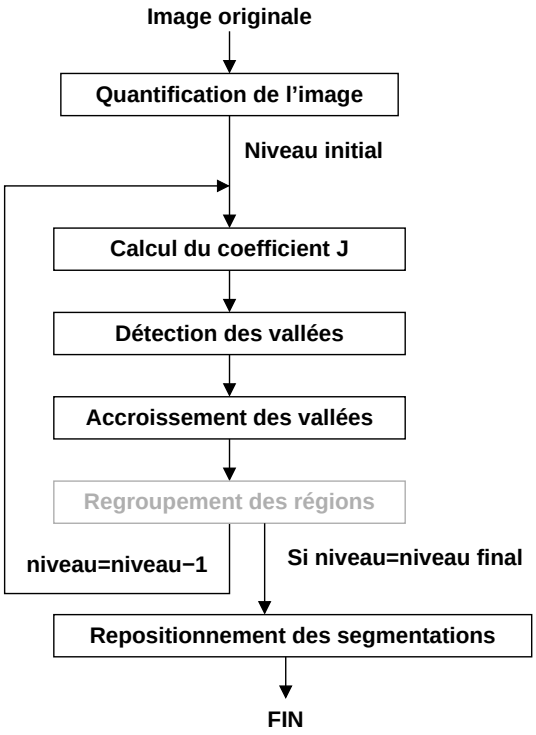


FIG. 5.13 – Schéma de principe de notre implantation de l'algorithme JSEG

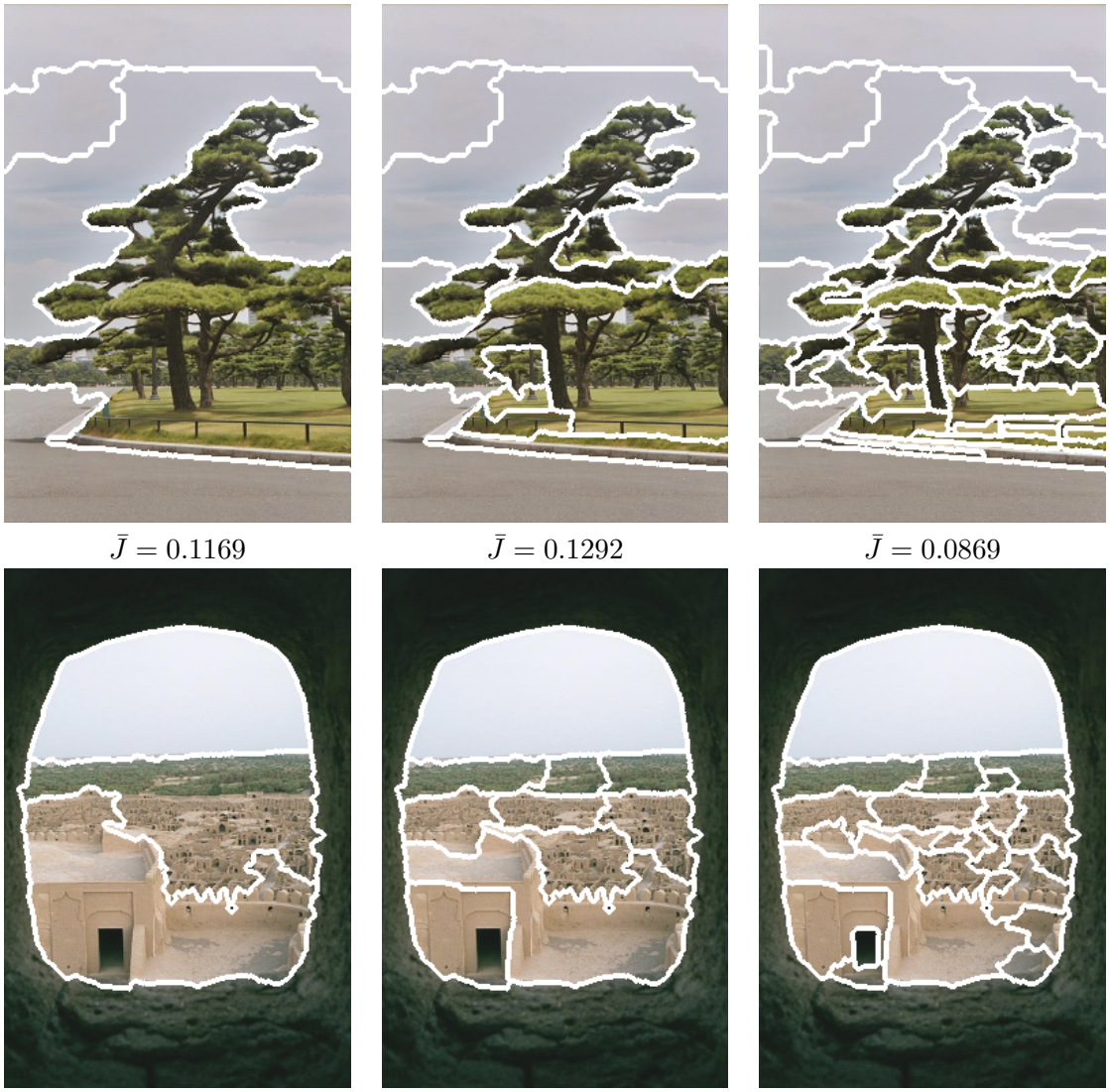
Nombre de pixels de l'image	Nombre de niveaux
<128*128	2
<256*256	2
<512*512	3
<1024*1024	4

Niveau	Fenêtre utlisée	Échantillonnage	Fenêtre de filtrage	Taille minimum des vallées
1	9×9	$1/(1 \times 1)$	pas de filtrage	32
2	17×17	$1/(2 \times 2)$	3×3	128
3	33×33	$1/(4 \times 4)$	5×5	512
4	65×65	$1/(8 \times 8)$	9×9	2048

FIG. 5.14 – Paramètres utilisés pour notre implantation de JSEG

mais les segmentations successives fournissent tout de même un ensemble de régions ayant une signification utile pour la description plus ou moins fine de l'image.

À partir de ces segmentations, nous allons décrire les images comme la pyramide des différentes régions en conservant l'information topologique de chaque niveau et celle de composition reliant les différentes segmentations. Le chapitre 6 expose les méthodes de description définies pour les différentes régions et le chapitre 7 les méthodes développées pour utiliser les informations topologiques et de composition dans le cadre de l'indexation et la recherche d'objets et d'images.

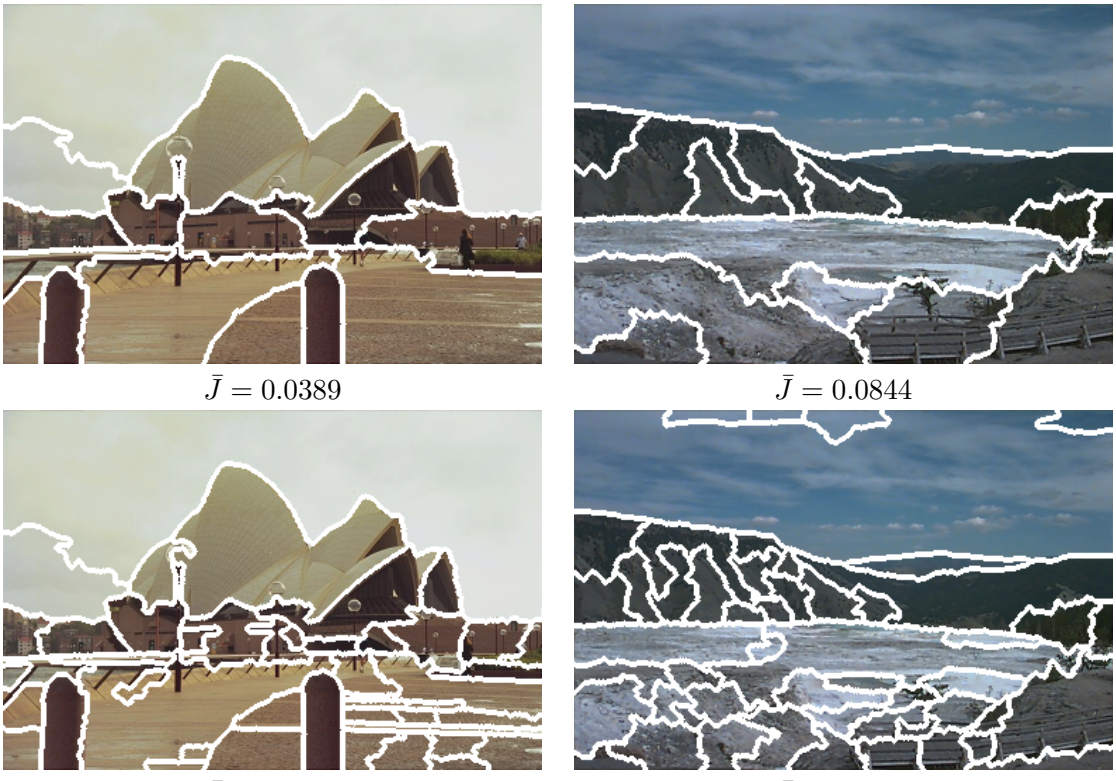


$\bar{J} = 0.1169$ $\bar{J} = 0.1292$ $\bar{J} = 0.0869$

$\bar{J} = 0.0388$ $\bar{J} = 0.0853$ $\bar{J} = 0.0530$

FIG. 5.15 – Segmentations d'images naturelles de la base de l'université de Washington (1/2)

$\bar{J} = 0.0850$ $\bar{J} = 0.0976$



$\bar{J} = 0.0389$ $\bar{J} = 0.0844$

$\bar{J} = 0.0475$ $\bar{J} = 0.0745$

FIG. 5.16 – Segmentations d'images naturelles de la base de l'université de Washington (2/2)

FIG. 5.17 – Segmentations obtenues sur des images standards

FIG. 5.18 – Segmentations obtenues sur des images artificielles issues de la base Columbia

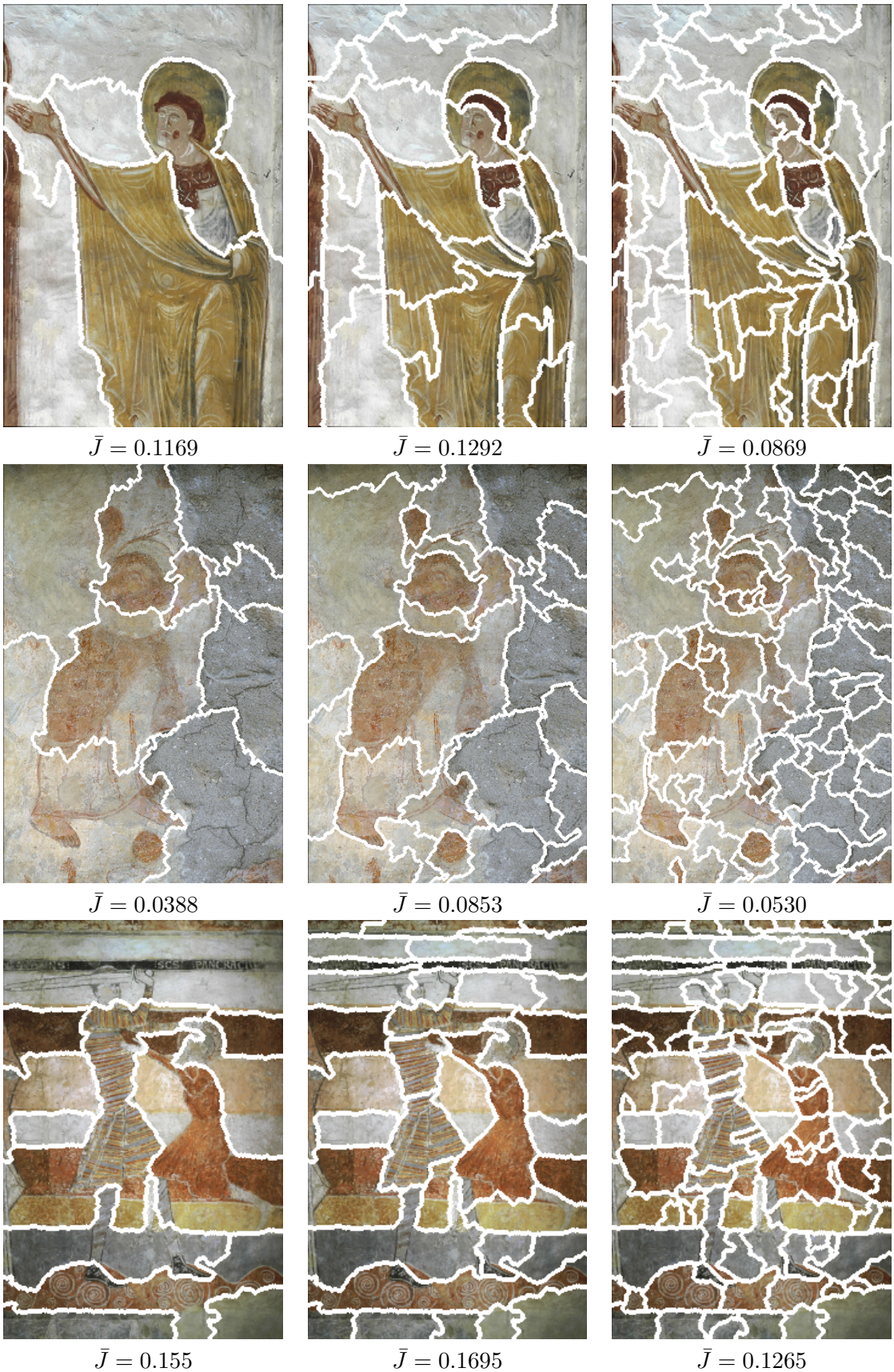


FIG. 5.19 – Segmentations obtenues sur des fresques médiévales issues de la base du CESCM

CHAPITRE 6

DÉCRIRE UNE RÉGION

Sommaire

6.1 Introduction

Généralement, diverses statistiques forment la description de l'image ; elles sont aussi appelées attributs ou descripteurs. Dans ces conditions, elles peuvent être mises sous la forme d'un vecteur, appelé vecteur d'attributs. La distance entre deux images est alors une distance dans un espace à n dimensions, où n est le nombre d'attributs. La couleur, la texture et la forme forment les trois familles de descripteurs classiquement employés.

Notre approche implique de pouvoir comparer, non plus l'image entière, mais les différentes régions la composant, i.e. toutes les régions du graphe pyramidal. De cette manière, chaque région est caractérisée par sa couleur, sa forme et sa texture. Pour chaque type de description, il existe un grand nombre de méthodes. Malheureusement, la plupart ne sont pas adaptées à notre structure de représentation. En effet, il faut qu'elles soient indépendantes de l'échelle et doivent de plus disposer de propriétés additives simples. Les descripteurs d'un niveau pourront ainsi être utilisés pour calculer facilement ceux du niveau supérieur. De même, le calcul des attributs non déduits doit être rapide, car pour une même image il sera effectué de nombreuses fois.

Après avoir fait un état de l'art des méthodes de description, nous présenterons dans ce chapitre les choix retenus au niveau des descripteurs de couleur, de forme et de texture. Des expérimentations ont donc été réalisées pour déterminer lesquels sont les plus adaptés à la caractérisation des régions. Pour cela, diverses bases d'images ont été utilisées ; elles sont présentées en détail en annexe B.

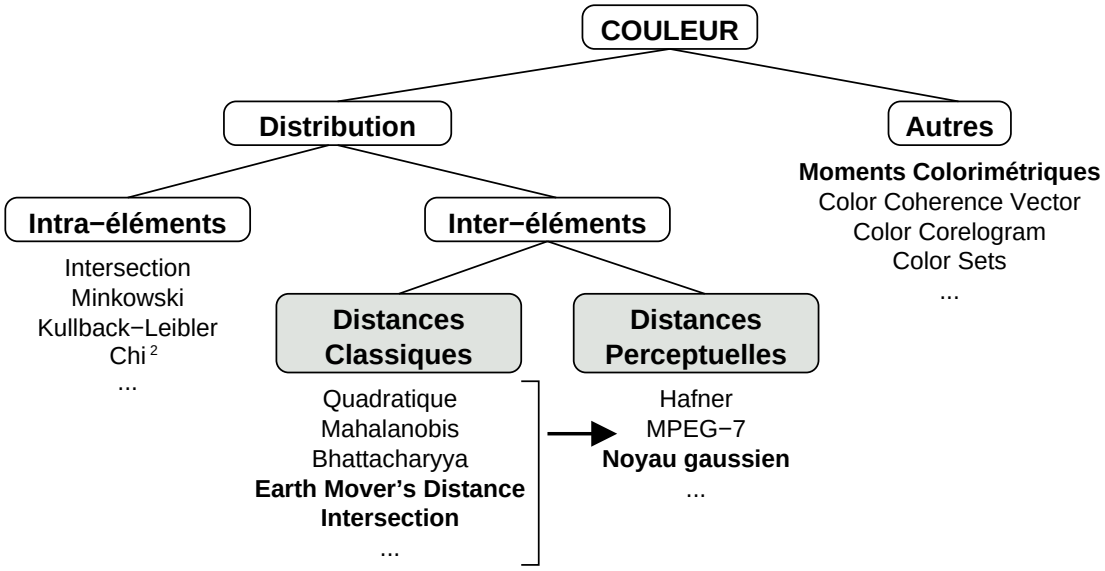


FIG. 6.1 – Organigramme des descripteurs couleur
En gras les statistiques étudiées.

6.2 Descripteurs couleur

La couleur est sûrement l'une des premières choses qui est perçue par l'homme lors de l'observation rapide d'une image. Le premier regard d'une image va permettre de déterminer la luminance et la teinte de l'image : sombre ou lumineuse, plutôt bleue, rouge ou verte. Discerner la texture ou la forme des objets demande plus d'attention.

De plus, caractériser la couleur apparaît beaucoup plus naturel que la forme ou la texture. Le vocabulaire disponible pour décrire les couleurs est très vaste ; tous les noms de couleurs peuvent être employés auxquels s'ajoutent des nuances sur la teinte et la luminosité. Des couleurs proches peuvent ainsi être distinguées : « rouge », « rouge pâle », « rose », « rouge foncé », « rouge bleuté clair »... Ce vaste espace de représentation est un avantage du point de vue de la qualité de description atteignable mais aussi un inconvénient du fait qu'elle n'est pas absolue ; deux personnes différentes vont pouvoir décrire une même couleur de deux manières différentes.
Du côté de la texture et de la forme, le vocabulaire disponible est beaucoup plus restreint ce qui explique que leur description soit beaucoup plus difficile. La texture peut être « grossière » ou « fine », et la forme plutôt « lisse » ou « brisée » mais il est difficile de donner plus de détails.

Par ces simples remarques, il est évident que la couleur est un des facteurs très importants dans la reconnaissance des images et par conséquent des objets qui les composent. D'autre part dans des applications multimédia, elle est généralement plus discriminante que la texture ou la forme qui apparaissent comme des caractéristiques complémentaires moins naturelles. Par contre, dans d'autres types d'applications tels que le contrôle qualité industriel ou pour des diagnostics dans des domaines très précis ce n'est pas toujours le cas.

Les statistiques utilisées pour décrire la couleur sont très diverses (cf. figure [6.1]). Nous exposons dans cette section le descripteur le plus couramment utilisé : l'histogramme couleur. Nous nous intéressons plus particulièrement aux techniques considérant cet attribut comme une distribu-

tion ce qui permet d'effectuer des comparaisons plus pertinentes du point de vue visuel. L'évolution des distances couleur classiques vers des modèles perceptuellement plus corrects est utilisée dans le même but. Des statistiques différentes ont également été développées. Elles sont présentées en fin de cette partie.

6.2.1 Couleur et quantification

Après acquisition, une image couleur comprend un grand nombre de couleurs différentes (jusqu'à 4 milliards en 32 bits ou 16 millions en 24 bits). Or, l'homme ne perçoit environ que 35000 nuances colorées au total et n'en différencie au grand maximum que 256 à la fois. De nombreux pixels ayant des couleurs proches ne seront donc pas différenciés. La quantification consiste alors à réduire le nombre total de couleurs de l'image en la dégradant visuellement le moins possible. C'est donc un système de compression avec pertes.

En traitement d'images, cette étape est quasiment obligatoire pour réduire la complexité des différents algorithmes. Ainsi, les diverses analyses statistiques réalisables sur les images seront facilitées si la population étudiée est plus faible. De nombreuses méthodes ont ainsi vu le jour pour répondre à ce problème de réduction de couleurs. Une des premières applications est apparue avec l'arrivée des stations graphiques codant les couleurs sur 16, 24 ou 32 bits. Les systèmes d'affichage étant alors uniquement capable de travailler avec une palette de 256 couleurs, il était nécessaire de trouver un ensemble de couleurs permettant de reproduire le plus fidèlement possible l'environnement graphique. Les couleurs réellement codées doivent ensuite être affectée à une couleur de la palette déterminée.

Quelques-unes des techniques développées sont présentées brièvement ici : quantification uniforme, *Popularity*, *Median Cut*, *Octree* et la méthode des palettes locales.

La première méthode mise en place consiste simplement à partitionner l'espace couleur de manière régulière ; c'est la quantification uniforme. L'espace colorimétrique est ainsi divisé en plusieurs zones de tailles identiques. En RVB par exemple, une quantification uniforme en 64 couleurs sera réalisée en divisant le cube des couleurs en 64 boîtes. Les couleurs d'une même boite sont alors remplacées par la couleur centrale de cette zone. Par conséquent dans cet exemple, chaque composante colorée est représentée par 4 niveaux différents ($4*4*4 = 64$ couleurs). Cette transformation est très simple à réaliser pratiquement car elle correspond simplement à une division. Mais, son inconvénient majeur est qu'elle ne tient pas compte de la distribution des couleurs au sein de l'image. De cette manière, certaines couleurs de la palette finale ne sont pas utilisées ; le rendu visuel n'est donc pas optimal. De plus, les espaces cubiques (RVB...) s'adaptent bien à ce procédé mais les espaces aux formes plus complexes ($L^*a^*b^*$, TLS...) sont difficiles à partitionner uniformément.

Une autre approche simple est *Popularity*. Elle est proposée par HECKBERT[HECKBE82] et détermine la palette couleur finale comme l'ensemble des couleurs les plus fréquentes dans l'image. Bien évidemment, les résultats obtenus avec un tel processus sont généralement assez mauvais. En effet, même s'il prend en compte la composition de l'image, de nombreuses couleurs de l'image ne disposent pas forcément de représentants perceptuellement proches. Un grand nombre de détails de l'image sont alors représentés en fausses couleurs. Malgré ces imprécisions, ce type de quantification était à l'époque suffisante dans le cadre de l'affichage pour des stations graphiques.

HECKBERT propose aussi l'idée de diviser l'espace RVB de manière à obtenir au final une équirépartition des pixels de l'image au sein de la palette couleur générée[HECKBE80]. Pour cela,

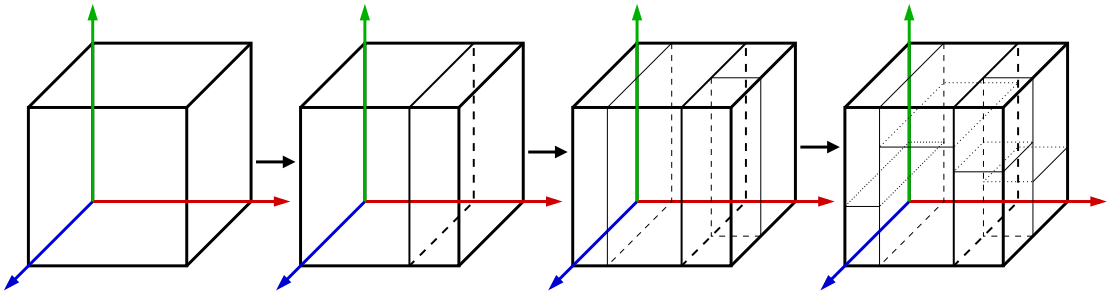


FIG. 6.2 – Divisions successives de l'espace RVB par *Median Cut*
L'espace est divisé récursivement par le plan perpendiculaire à l'axe ayant la distribution la plus large passant par la médiane de cette distribution.

le cube RVB est divisé récursivement de telle sorte à ce que les boîtes parallélépipédiques générées contiennent à peu près le même nombre de pixels. Ce découpage peut être réalisé en choisissant à chaque fois le plan perpendiculaire à l'axe ayant la distribution la plus large et passant par la médiane de la distribution, d'où le nom de *Median Cut* (cf. figure [6.2]). Bien qu'aucune information sur la perception des différentes couleurs ne soit utilisée, les résultats sont satisfaisants. HECKBERT montre d'ailleurs que le *Median Cut* fournit de meilleurs résultats que *Popularity*.

La méthode *Octree* est semblable au *Median Cut* dans son principe de division de l'espace colorimétrique. Elle a été introduit par GERVAUTZ et PURGATHOFER[GERVAU90]. L'espace couleur est tout d'abord partitionné sous forme d'un *octree* qui est similaire au principe du *quadtree* en 2 dimensions (cf. section 4.3.3.1). En RVB, le cube est divisé en 8 cubes de tailles identiques ; ce processus est alors répété sur chaque cube produit et permet de former un arbre codant les subdivisions successives. Les cubes ne contenant pas de pixels sont bien évidemment supprimés de l'arbre. De cette manière, toutes les couleurs de l'image sont représentées par les feuilles de l'*octree*. La quantification consiste alors à réduire cet arbre pour ne garder que le nombre voulu de couleurs représentatives. Différentes méthodes peuvent être utilisées pour cette étape. Le principe original consiste à supprimer récursivement les feuilles les moins peuplées jusqu'à n'avoir plus que le nombre de couleurs désiré. Les couleurs de la palette sont finalement les moyennes des différents pixels présents dans les nœuds restants.
Le choix de la méthode d'élagage apparaît très important dans une telle technique. Le principe initial semble ainsi mal adapté car il ne prend pas en compte la notion de perception des couleurs. Il est plus intéressant d'effectuer un élagage des nœuds aux endroits où celui-ci engendrera le moins de dégradations visuelles. C'est, en particulier, la stratégie employée par *ImageMagick* pour son algorithme de quantification. C'est d'ailleurs celui que nous avons choisi pour le prétraitement nécessaire à notre algorithme de segmentation (cf. section 5.3.3).

LARABI propose aussi un algorithme appelé « méthode des palettes locales »[LARABI00]. Dans celui-ci, les images sont tout d'abord partitionnées en imagettes. Pour chaque imagette, la palette caractéristique est déterminée. L'ensemble des différentes palettes forment alors la matrice des palettes locales à partir de laquelle est établie la palette finale de l'image par construction d'un arbre colorimétrique. Cet algorithme est donc fondé sur une coopération spatio-fréquentielle puisqu'il opère un partitionnement de l'image qui se traduit par l'ajout de la dimension spatiale dans l'histogramme.

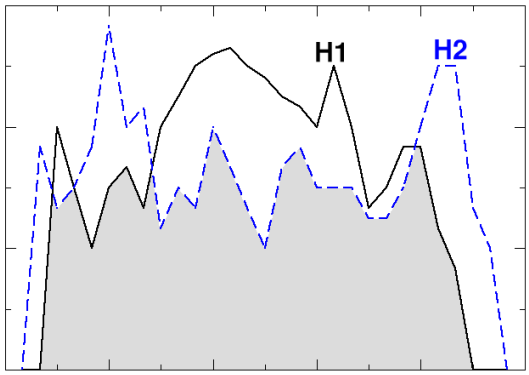


FIG. 6.3 – Principe de l'intersection d'histogrammes proposée par SWAIN et BALLARD
La valeur obtenue par l'intersection des 2 histogrammes correspond à l'aire de la zone grisée.

De nombreuses autres méthodes de quantification sont aussi disponibles :
- *Variance-Based* qui se différencie du *Median Cut* par sa technique de découpage fondée sur les variances intra-classe et inter-classes[WAN90] ;
- *Neuquant* utilise un réseau de neurone de KOHONEN. Ses résultats sont de très bonne qualité mais sa complexité est importante[DEKKER94] ;
- méthodes classiques de classification : *k-means*[VEREVK95], *fuzzy C-means*[BEZDEK81]...
- ...

Cette étape de quantification permet ainsi d'adapter le coût de la comparaison aux performances du calculateur et des besoins de discrimination.

6.2.2 L'histogramme couleur, la statistique standard

La distribution des couleurs au sein d'une image est très souvent représentée sous la forme d'un histogramme couleur. Comme nous l'avons déjà présenté à la section 3.3.1, celui-ci associe à chaque couleur sa fréquence d'apparition. Ainsi, pour une image définie par $I : x \in \Omega \rightarrow I(x) \in \Theta$ avec Ω le domaine de l'image et Θ celui des couleurs, la fonction histogramme de l'image I peut s'écrire $H_I : c \in \zeta_I \rightarrow \mathbb{R}$ où $\zeta_I = \{I(x)/x \in \Omega\}$ est l'ensemble des couleurs de l'image. $nbc_I = \text{Card}(\zeta_I)$ est par conséquent le nombre de couleurs présentes au sein de l'image I. Les valeurs de l'histogramme sont alors :

$$H_I(c) = \frac{\text{Card}(\{x \ / \ x \in \Omega,\ I(x) = c\})}{\text{Card}(\Omega)},\ \forall c \in \zeta_I.$$

Il est à noter ici qu'avec cette définition l'histogramme obtenu est normalisé, $\sum_{c \in \zeta_I} H_I(c) = 1$.

Nous appliquons ici les diverses distances ou mesures de similarité sur les histogrammes couleurs mais elles le sont aussi sur des distributions quelconques.

6.2.3 Comparaison d'histogrammes intra-éléments

SWAIN et BALLARD ont les premiers proposé une méthode d'intersection d'histogrammes permettant de comparer des images par leur distribution colorimétrique[SWAIN91]. Considérant deux images I_1 et I_2 et leurs histogrammes respectifs H_1 et H_2 définis sur le même domaine couleur ζ, la distance proposée est :

$$D_{SwainBallard}(H_1, H_2) = \frac{\sum_{c\in\zeta} min(H_1(c), H_2(c))}{\sum_{c\in\zeta} H_1(c)}$$

Cette mesure revient à trouver l'aire commune aux deux distributions (cf. figure [6.3]). Il est à noter ici que pour une utilisation correcte de cette distance, les deux histogrammes ne doivent pas être normalisés.

Cette intersection n'est réalisable que si les deux histogrammes ont même support colorimétrique ; i.e. que les ensembles de couleurs utilisés dans les deux histogrammes sont identiques. Dans cette approche, seuls les éléments correspondant aux mêmes couleurs sont comparés ; on parle alors de comparaison *intra-éléments*.

Cette distance n'est une métrique que si $\sum_{c\in\zeta} H_1(c) = \sum_{c\in\zeta} H_2(c)$ ce qui assure la propriété de symétrie. Cette condition est satisfaite quand les deux histogrammes sont normalisés.

La forme générale de la distance de MINKOWSKI est :

$$D^p_{\text{MINK}}(H_1, H_2) = \left(\sum_{c\in\zeta} |H_1(c) - H_2(c)|^p \right)^{\frac{1}{p}}$$

Trois formes de cette distance sont très répandues pour la comparaison d'histogrammes couleur :

$$\begin{cases} D^1_{\text{MINK}}(H_1, H_2) = L_1(H_1, H_2) = \sum_{c\in\zeta} |H_1(c) - H_2(c)| \\ D^2_{\text{MINK}}(H_1, H_2) = L_2(H_1, H_2) = \sqrt{\sum_{c\in\zeta} |H_1(c) - H_2(c)|^2} \\ D^\infty_{\text{MINK}}(H_1, H_2) = L_\infty(H_1, H_2) = \max_{c\in\zeta} |H_1(c) - H_2(c)| \end{cases}$$

La distance d'ordre 1 correspond à la distance L_1 aussi nommée distance de Manhattan, celle d'ordre 2 correspond à la distance à la distance L_2 ou euclidienne et celle d'ordre infini correspond à la distance L_∞ ou de CHEBYCHEV.

D'autres distances permettent de comparer deux distributions et donc deux histogrammes : la distance du χ^2 ou celle de KULLBACK-LEIBLER par exemple. Mais le gros inconvénient de ces méthodes est qu'elles réalisent une comparaison des éléments identiques aux sein des distributions. Dans le cadre des histogrammes couleur, cette approche fournit parfois des résultats perceptuellement faux dus par exemple à une variation de l'intensité lumineuse qui décale légèrement l'histogramme (cf. figure [6.4]). Ce problème peut être partiellement évité par la comparaison des éléments d'un histogramme avec l'ensemble des éléments de l'autre.

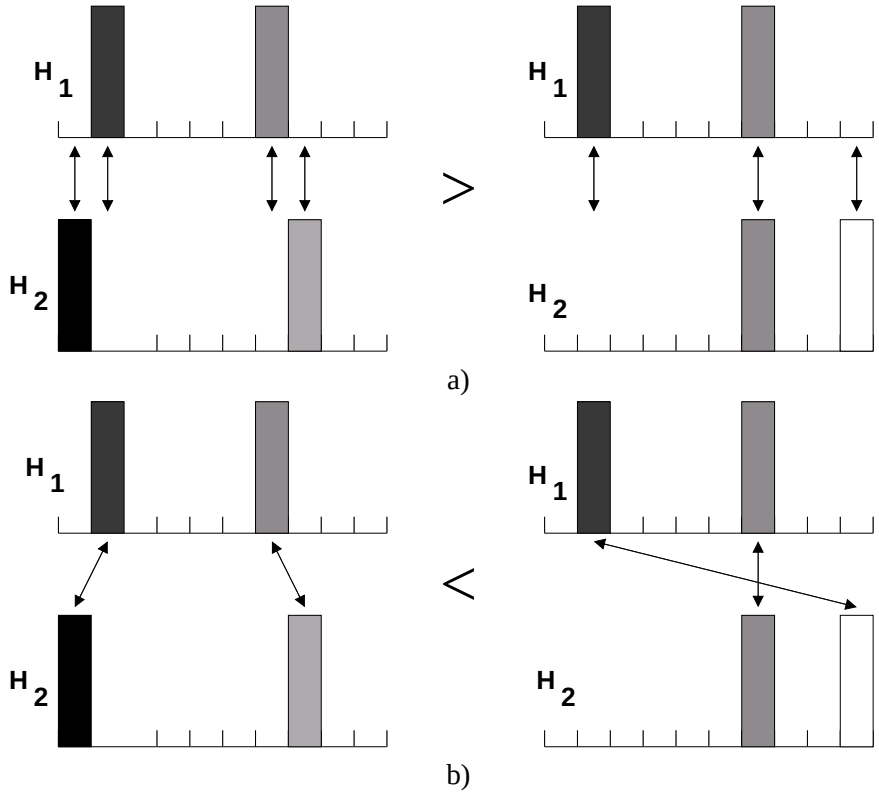


FIG. 6.4 – Limitation des approches intra-éléments pour la comparaison d'histogrammes
La distance utilisée en a) est une distance de type intra-éléments. Dans le cas de l'intersection d'histogrammes par exemple, les distances sont alors 0 à gauche et 0.5 à droite. D'après cette mesure les histogrammes de droite sont donc plus similaires que ceux de gauche ce qui est perceptuellement faux. Dans le cadre d'une distance inter-éléments présenté en b), la mesure de similarité donne alors le bon résultat.

6.2.4 Comparaison d'histogrammes inter-éléments

6.2.4.1 Distance quadratique

Pour s'affranchir des problèmes des méthodes précédentes, la comparaison peut prendre en compte les distances de chaque couleur du premier histogramme à toutes celles du deuxième ; ces méthodes font donc intervenir les relations *inter-éléments*. Certaines de ces méthodes lèvent alors la condition très limitative des palettes identiques.

La distance quadratique est la méthode la plus classique pour prendre en compte les relations inter-éléments. Pour cela, les supports colorimétriques des images sont fusionnés pour obtenir une palette commune. Matriciellement, elle s'exprime alors par :

$$D_{Quadra}(H_1, H_2) = (h_1 - h_2)^T A (h_1 - h_2)$$

où h_1 et h_2 sont les vecteurs représentants les histogrammes et A la matrice définissant la similarité entre les couleurs. En appelant a_{ij} les éléments de A caractérisant la similarité entre les couleurs i

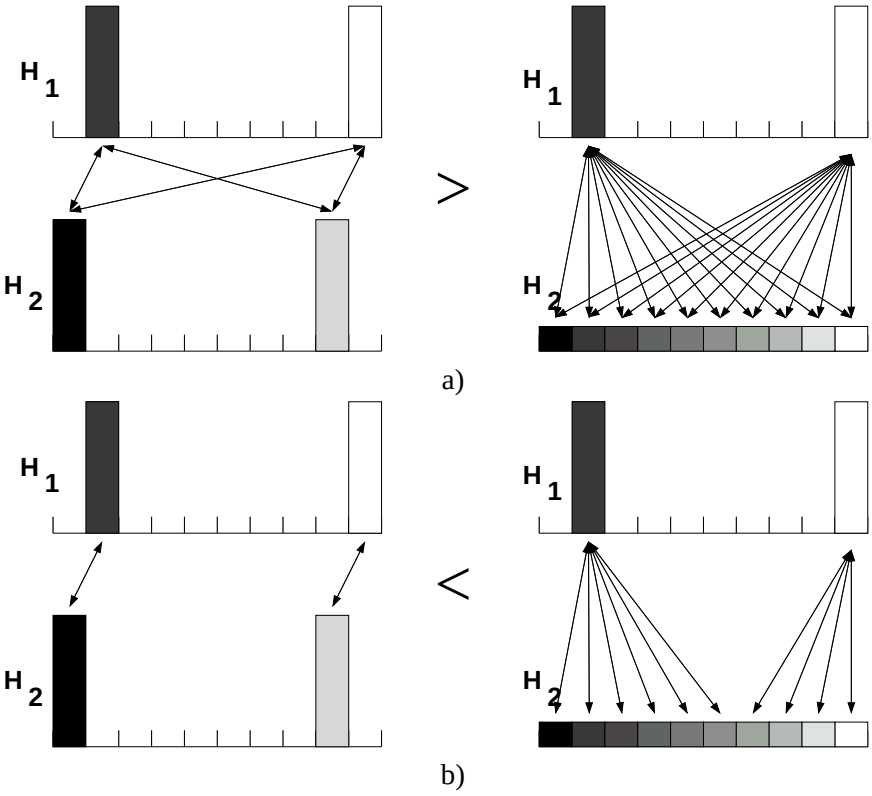


FIG. 6.5 – Limitation des approches inter-éléments pour la comparaison d'histogrammes
La distance utilisée en a) est une distance de type inter-éléments. La comparaison des éléments extrêmes de l'histogramme est peu intéressante dans ce cas et augmente considérablement la distance finale. En b), la comparaison prend en compte les différentes répartitions des histogrammes et s'adapte alors au contenu ce qui améliore le résultat.

et j et ζ le support colorimétrique de H_1 et H_2, l'équation matricielle précédente peut être développée :

$$D_{Quadra}(H_1, H_2) = \sum_{i \in \zeta} \sum_{j \in \zeta} a_{ij}(H_1(i) - H_2(i))(H_1(j) - H_2(j))$$

Cette mesure de similarité est utilisée dans le système QBIC[NIBLAC93].

À partir de la distance quadratique, de nouvelles distances ont été définies et en particulier celles de MAHALANOBIS[MAHALA36, SMITH97] et de BHATTACHARYYA[BHATTA43] qui sont toujours utilisables après fusion des palettes couleurs.

Malheureusement, dans certains cas, la prise en compte de toutes les relations entre éléments fausse le résultat final. La figure [6.5] montre un exemple de configurations où ces distances fournissent une mauvaise estimation de la similarité contrairement aux systèmes utilisant la correspondance entre les éléments. Des techniques telles que l'*Earth Mover's Distance* effectuent un *matching* adaptatif des distributions en tenant compte des distances entre les éléments. Dans ce cas, la mesure finale de similarité ne fait intervenir que certaines relations entre éléments et non leur totalité.

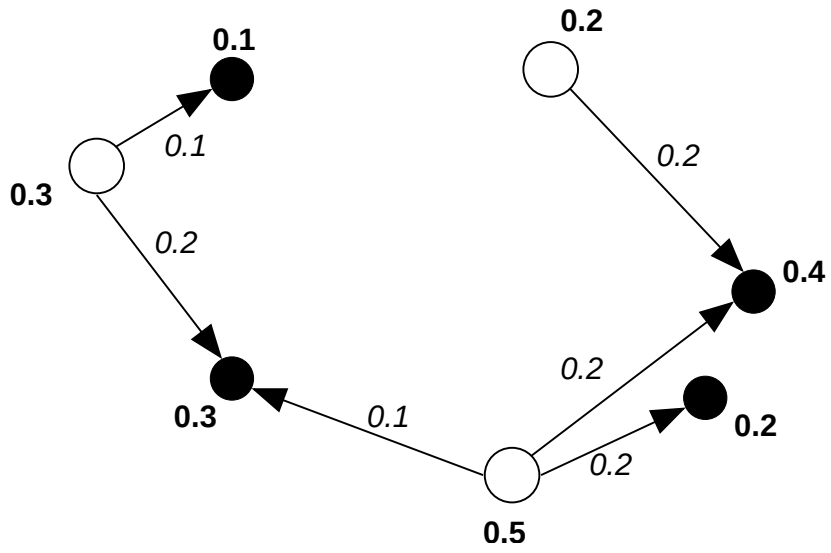


FIG. 6.6 – Exemple de transport de masses pour la distance *EMD*
Les points blancs correspondent à la première distribution et les noirs à la deuxième. Les deux distributions n'ont pas la même cardinalité.

6.2.4.2 *Earth Mover's Distance*

RUBNER propose en 1998 l'*Earth Mover's Distance* (*EMD*)[RUBNER98C]. Considérons, les éléments du premier histogramme comme des monticules de terre et ceux du second comme des trous. Les tailles des monticules et des trous sont proportionnelles aux valeurs des histogrammes. Les distances entre les différents trous et monticules correspondent quant à elles aux distances entre les couleurs. L'*EMD* calcule alors le coût minimum nécessaire pour remplir les différents trous avec la terre des monticules (cf. figure [6.6]) ; c'est la distance entre les deux distributions. Ce problème est connu depuis de nombreuses années sous le nom de *problème de transport*[RUSSEL69, DANTZI51].

Plus formellement, dans le cadre des histogrammes couleurs, nous disposons de deux histogrammes normalisés H_I et H_J calculés sur les images I et J (deux distributions associant une couleur c et sa probabilité d'apparition $H(c)$). Ils sont définis sur des supports colorimétriques ζ_I et ζ_J non nécessairement identiques de cardinalités quelconques. Soit f la fonction de transport de masses entre H_I et H_J. f est définie à valeurs positives par $f : (i,j) \in (\zeta_I, \zeta_J) \rightarrow f_{ij} \in \mathbb{R}^+$ où f_{ij} correspond à la masse transférée de la couleur i de I vers la couleur j de J. Bien évidemment, des contraintes doivent empêcher le transport de masse inexistante ou l'arrivée d'une masse dans un endroit ne pouvant la stocker :

$$\begin{cases} \sum_{i \in \zeta_I} f_{ij} \leq H_J(j),\ \forall j \in \zeta_J \\ \\ \sum_{j \in \zeta_J} f_{ij} \leq H_I(i),\ \forall i \in \zeta_I \end{cases}$$

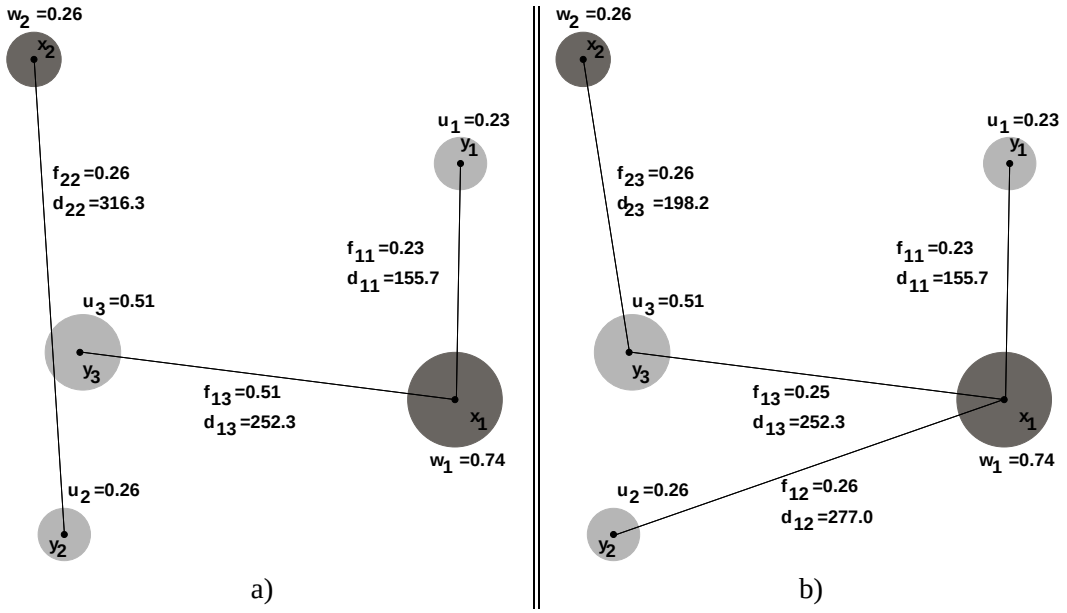


FIG. 6.7 – Exemple standard de flux optimal et non optimal entre deux distributions
Deux distributions sont présentes : x associée aux poids w et composée de 2 éléments et y de 3 éléments associée aux poids u. La surface des disques est proportionnelle aux poids affectés. La distance géométrique entre deux éléments correspond à la distance entre leurs caractéristiques. Le calcul de l'effort lors du transport de masse de a) est de $0.23*155.7+0.26*316.3+0.51*252.3 = 246.7$. *Ce flux n'est pas optimal. Le transport de b) correspond à la minimisation de l'effort à fournir pour transférer x vers y. L'effort est alors de* $0.23*155.7+0.25*252.3+0.26*198.2+0.26*277.0 = 222.4$. *Les distributions étant normalisées, l'EMD est donc égale à* $EMD(x, y) = 222.4$.

De plus, il faut garantir le transport total des masses :

$$\sum_{i\in\zeta_I}\sum_{j\in\zeta_J} f_{ij} = 1 = \min\left(\sum_{i\in\zeta_I} H_I(i), \sum_{j\in\zeta_J} H_J(j)\right)$$

Dans le cas de distributions non normalisées, la masse transportée doit être égale à la somme des poids de la distribution la plus légère.

Le coût de transport par f entre H_I et H_J est alors :

$$Coût(H_I, H_J, f) = \sum_{i\in\zeta_I}\sum_{j\in\zeta_j} f_{ij}d_{ij}$$

où d_{ij} est la distance entre les couleurs i et j dans l'espace utilisé. Par conséquent, le problème de transport consiste à déterminer f de manière à minimiser $Coût(H_I, H_J, f)$. La distance *EMD* est donc définie par :

$$EMD(H_I, H_J) = \min_f\left(Coût(H_I, H_J, f)\right) = \frac{\min_f\left(Coût(H_I, H_J, f)\right)}{\min\left(\sum_{i\in\zeta_I} H_I(i), \sum_{j\in\zeta_J} H_J(j)\right)}$$

Algorithme 1 : Algorithme de calcul de la distance *EMD* proposé par HEINRICHS si $\zeta_I = \zeta_J$

Données : H_1 et H_2 deux histogrammes définis sur la même palette couleur de n couleurs ; $\{d_{ij} \ / \ i \in [1 \dots n],\ j \in [1 \dots n]\}$ avec d_{ij} la distance entre les couleurs i et j.

Résultat : $EMD(H_1, H_2)$ par minimisation de $\sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij} f_{ij}$, où f_{ij} est la quantité transférée de la couleur i de H_1 à j de H_2.

début

  tant que $\sum_{i=1}^{n} H_1(i) \neq 0$ *ou* $\sum_{j=1}^{n} H_2(j) \neq 0$ **faire**

    1 - Calculer $w_i = \max_j d_{ij}, \forall i = 1 \dots n$ et $y_j = \max_i d_{ij}, \forall j = 1 \dots n$;

    2 - Trouver (i, j) tels que $w_i + y_j - d_{ij} = \max_{(k,l)}[(w_k + y_l - d_{kl}) > 0]$;

    3 - $f_{ij} = \min(H_1(i), H_2(j))$;

    4 - $H_1(i) = H_1(i) - f_{ij}$ et $H_2(j) = H_2(j) - f_{ij}$;

    5 - Éliminer les lignes et les colonnes correspondantes à $H_1(i) = 0$ ou $H_2(j) = 0$.

  fin

fin

Dans le cas de distributions non normalisées, il faut rapporter ce coût au volume déplacé pour obtenir le coût unitaire du transport. La figure [6.7] montre deux flux classiquement présentés pour illustrer la notion de transport optimal.

RUBNER présente aussi diverses expériences d'analyse de couleurs ou de textures montrant les propriétés intéressantes de cette distance en traitement d'images[RUBNER98B]. Ainsi, l'*EMD* est robuste au bruit et au décalage. De plus, dans le cas où les poids totaux des distributions sont égaux, RUBNER prouve que l'*EMD* est une métrique et qu'elle satisfait en particulier à l'inégalité triangulaire. Enfin, diverses mises en compétition de cette méthode et des distances standards pour des images à supports couleur identiques ont montrée que l'*EMD* est plus performante mais plus lente[RUBNER98A, PUZICH99]. En effet, sa complexité est de $\mathcal{O}(n^2)$ contrairement aux méthodes précédentes qui sont généralement calculables en $\mathcal{O}(n)$.

De nombreux algorithmes permettent de calculer cette distance. RUBNER propose ainsi d'utiliser l'algorithme du *simplex* appliqué au problème de transport[1]. HEINRICHS propose quant à lui une version optimisée pour des histogrammes à même support colorimétrique (cf. algorithme 1) [HEINRI99].

Nous verrons dans la section 6.2.8 que cette distance fournit un bon indice de ressemblance entre zones colorées.

6.2.4.3 Intersection d'histogrammes à palettes différentes

Malgré des supports colorimétriques non identiques, nous avons essayé de mettre en place une mesure de similarité semblable à l'intersection d'histogrammes. Nous devons ainsi prendre en

[1] Le code source est disponible à l'adresse http://robotics.stanford.edu/~rubner/emd/default.htm.

compte la composition des différents histogrammes mais aussi les distances entre les différentes couleurs les composant :

- H_1 est défini sur ζ_1 avec $\sum_{i \in \zeta_1} H_1(i) = 1$ et $\forall i \in \zeta_1,\ H_1(i) \neq 0$;
- H_2 est défini sur ζ_2 avec $\sum_{j \in \zeta_2} H_2(j) = 1$ et $\forall j \in \zeta_2,\ H_2(j) \neq 0$;
- Les distances entre couleurs sont $d_{ij} = d(i,j), \forall (i,j) \in (\zeta_1, \zeta_2)$ où $d(i,j)$ est la fonction distance choisie pour déterminer l'écart visuel entre les couleurs i et j.

Connaissant les dynamiques des différentes composantes, nous réalisons tout d'abord leur normalisation au sein des différents histogrammes. Les couleurs sont alors représentées par un triplet de valeurs comprises entre 0 et 1. Par conséquent, les distances classiques entre couleurs (L_1, L_2...) seront aussi comprises dans l'intervalle $[0 \ldots 1]$. À partir de ces données, nous proposons la mesure de similarité suivante :

$$\begin{aligned} s_{inter}(H_1, H_2) &= \frac{\sum_{i \in \zeta_1} \left[\min_{j \in \zeta_2} \{(k_1 + |H_1(i) - H_2(j)|) * (k_2 + d_{ij}) - k_1 k_2\} \right]}{(1 + k_1 + k_2)} \\ &= \frac{\sum_{i \in \zeta_1} \left[\min_{j \in \zeta_2} \{k_2 |H_1(i) - H_2(j)| + k_1 d_{ij} + d_{ij} |H_1(i) - H_2(j)|\} \right]}{(1 + k_1 + k_2)} \end{aligned}$$

où k_1 et k_2 sont des constantes à déterminer. Empiriquement le choix de $k_1 = k2 = 1$ paraît intéressant car il donne le même poids à la distance entre les éléments et leur différence fréquentielle. Expérimentalement, les résultats sont les meilleurs dans ce cas.

Pour chaque couleur de H_1, la couleur minimisant l'aire du rectangle de côtés $(k_1 + |H_1(i) - H_2(j)|)$ et $(k_2 + d_{ij})$ est retenue (cf. figure [6.8]). La distance totale est alors égale à un facteur près à la somme de toutes les aires des rectangles formés.

Cette mesure n'étant pas symétrique ($s_{inter}(H_1, H_2) \neq s_{inter}(H_2, H_1)$), elle mesure la ressemblance d'un histogramme par rapport à l'autre mais de manière non inversible. Pour rendre cette relation symétrique il suffit d'écrire :

$$S_{inter}(H_1, H_2) = \frac{s_{inter}(H_1, H_2) + s_{inter}(H_2, H_1)}{2}$$

De plus, la non symétrie de la première mesure peut, pour certaines configurations, fournir une similarité faible alors qu'elle n'a pas lieu d'être. En effet, si les couleurs d'une image sont toutes très proches elles se projetteront majoritairement sur la même couleur de l'autre histogramme. Cette technique comparera alors un histogramme avec un élément unique du second. L'inversion du processus permet de gommer partiellement ce problème. En fait, grâce au processus de quantification des images généralement réalisé avant leur comparaison, ce type de configuration n'est quasiment jamais présent.

Cette mesure tient compte de la différence de volume entre les éléments des histogrammes et des distances couleurs entre les différents éléments respectivement grâce aux facteurs $|H_1(i) - H_2(j)|$ et d_{ij}. L'utilisation des éléments du premier ordre $|H_1(i) - H_2(j)|$ et d_{ij} est nécessaire

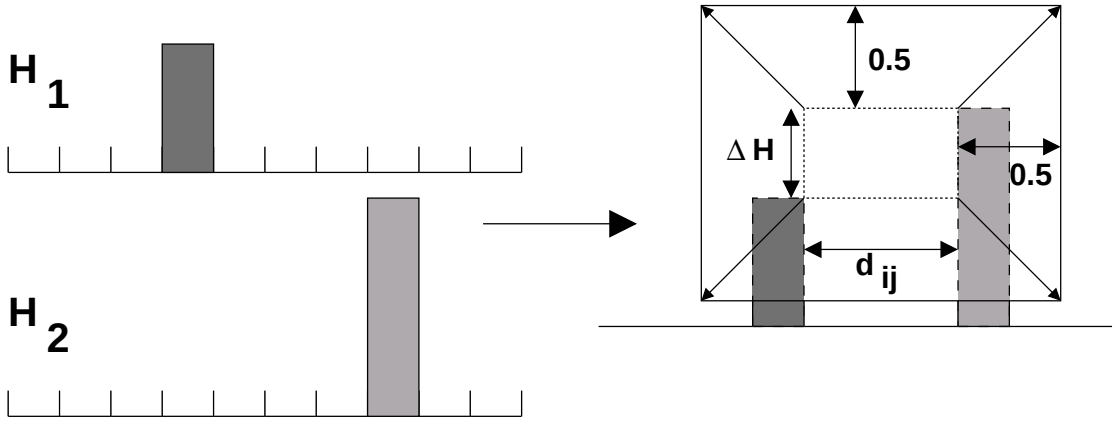


FIG. 6.8 – Comparaison de deux éléments pour l'intersection d'histogrammes proposée
Un rectangle peut être formé entre les deux éléments. Ses côtés correspondent à la distance entre couleurs et à la différence de volume entre les éléments. Les longueurs de ses côtés sont comprises entre 0 et 1 de par la normalisation de l'histogramme et celle des valeurs des couleurs. Dans le cas où $k_1 = k_2 = 1$, chaque côté est alors agrandi d'une unité de valeur pour former le rectangle final symbolisant la distance entre les éléments.

pour éviter une distance nulle dans le cas où la répartition des couleurs au sein des histogrammes est identique mais les couleurs différentes. Sans ces deux termes, la mesure de similarité est aussi faussée quand les couleurs présentes au sein des deux histogrammes sont identiques mais les répartitions fréquentielles différentes.

Nous verrons dans la section 6.2.8 que cette mesure quoique biaisée est très efficace dans la comparaison d'histogrammes couleur. De plus, sa complexité est inférieure à l'*EMD* puisqu'en $\mathcal{O}(n)$.

6.2.5 Comparaison adaptative d'histogrammes

Définir une distance perceptuellement acceptable dans un espace couleur est très difficile surtout pour des couleurs très éloignées. En effet, l'écart entre un bleu et un vert est-il plus important qu'entre un rouge et un jaune ? Il est facile de dire que les couleurs sont différentes mais quantifier l'écart entre des couleurs très éloignées est difficile. Pour les méthodes classiques telle que la distance quadratique, il a ainsi été montré que l'utilisation de distances classiques entre couleurs a tendance à trouver similaires des images peu semblables. HAFNER propose ainsi d'utiliser une forme gaussienne des distances classiques pour la distance quadratique[HAFNER95]. Récemment, dans la norme MPEG-7, a été introduite une métrique perceptuellement pertinente composée d'une combinaison linéaire de cosinus et de sinus appliquée à une distance classique.

À partir de la même idée, nous définissons une mesure basée sur une distance classique au sein d'un espace couleur tout en considérant qu'au delà d'un seuil la distance entre les couleurs est constante. Nous mettons ainsi en place une zone d'influence dans laquelle une couleur peut être comparée avec les autres. Au delà de cette zone, les couleurs sont considérées comme *différentes* de la couleur de référence mais sans notion de distance. Cette idée peut facilement être mise en

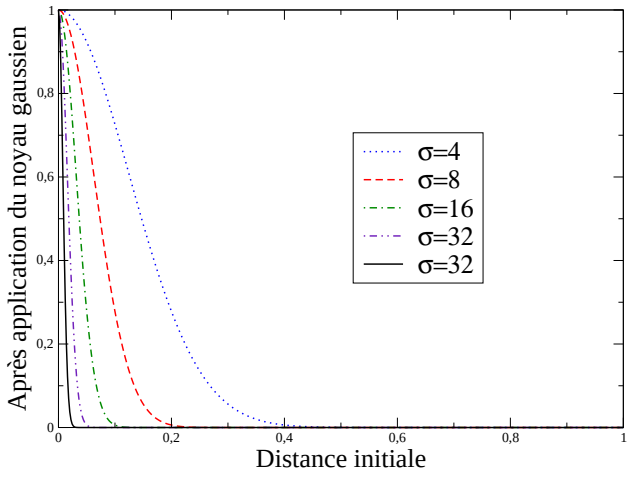


FIG. 6.9 – Évolution du noyau gaussien en fonction de σ (pour $A = 1$)

place en appliquant un noyau gaussien à l'échelle des distances. Dans un espace perceptuellement uniforme et linéaire, la distance $d(i,j)$ entre une couleur i et j peut alors évoluer vers :

$$p(i,j) = \frac{1}{\sigma\sqrt{2\pi}} * e^{-\frac{d(i,j)^2}{2\sigma^2}} = A * e^{-\frac{d(i,j)^2}{2\sigma^2}}$$

où σ définit la taille du noyau gaussien (cf. figure [6.9]) et A est une constante.

Cette nouvelle mesure de similarité peut être utilisée dans un cadre standard tel que l'intersection d'histogrammes définie précédemment. La distance utilisée est alors $1 - p(i,j)$. Pour l'*EMD*, elle peut aussi être vue comme la probabilité qu'un transport de masse soit réalisé entre les deux couleurs.

Nous avons également proposé un nouvel algorithme de comparaison d'histogrammes, la projection d'histogrammes, qui est une version simplifiée de l'*EMD*. De même que pour l'*EMD*, nous disposons de deux histogrammes et nous désirons transférer la masse de l'un vers l'autre. Sans vouloir complètement minimiser l'énergie nécessaire à cette opération, nous considérons que les couleurs les plus proches ne produiront pas visuellement de différences perceptibles. Elles doivent donc par tous les moyens être associées. Au final, les couleurs n'aillant pas trouvé de correspondants fourniront un bon indicateur sur la différence visuelle entre les histogrammes.

De manière imagée, tout transport de la masse est réalisée au moindre coût. Considérons un cobaye humain et plaçons le devant le problème de remplissage des trous à partir des monticules de terres présenté avec l'*EMD*. Il remplira tout d'abord (par simplicité) les trous pour lesquels il dispose d'une source de terre proche. Puis, il finira le travail en allant chercher la terre de plus en plus loin.

L'algorithme 2 présente en détail les différentes étapes de ce processus. Il est à noter que cette mesure de similarité est symétrique. Nous verrons dans la section suivante qu'elle fournit de très bons résultats dans le cadre de recherche de régions. La valeur de σ devant être adaptée à l'espace couleur et à la distance employés.

Algorithme 2 : Algorithme de mesure de similarité par projection d'histogrammes

Données : H_I et H_J deux histogrammes définis sur des palettes couleurs quelconques, respectivement ζ_I et ζ_J ;
$P = \{(ms(i,j), i, j)/i \in \zeta_I, j \in \zeta_J\}$ où $ms(i,j)$ est une mesure de similarité entre la couleur i de H_I et j de H_J (par exemple p définie précédemment).

Résultat : Coût de la projection entre les histogrammes.

début

$D_{proj} = 0$;

tant que P *non vide* **faire**

1 - Déterminer la similarité maximale : $q(i_q, j_q) = \max_{ms(i,j)} P$;

2 - Déterminer la masse transportée : $m = \min(H_1(i_q), H_2(j_q))$;

3 - Mettre à jour les histogrammes : $H_1(i_q) = H_1(i_q) - m$ et $H_2(j_q) = H_2(j_q) - m$;

4 - Mettre à jour P, suppression des éléments vides :

si $m = H_1(i_q)$ **alors**

$P = P - \{(ms(i,j), i, j) \ / \ i = i_q, \ j \in \zeta_J\}$

fin

si $m = H_2(j_q)$ **alors**

$P = P - \{(ms(i,j), i, j) \ / \ i \in \zeta_I, \ j = j_q\}$

fin

5 - Ajouter le coût du transport à la distance : $D_{proj} = D_{proj} + m * (1 - q(i_q, j_q))$;

fin

fin

6.2.6 Moments colorimétriques

Les moments colorimétriques (ou *Color Sets* en anglais) sont utilisés avec succès dans plusieurs systèmes d'indexation d'images tel que QBIC[2][NIBLAC93]. Ils sont particulièrement efficaces quand l'image contient un objet unique. Le premier ordre (la moyenne), le second (la variance) et le troisième (asymétrie) ont démontré leur cohérence et leur efficacité dans la représentation des distributions colorimétriques des images[STRICK95].

Mathématiquement, les trois premiers moments sont définis par :

$$\mu_i = \frac{1}{N} \sum_{j=1}^{N} c_{ij}, \ \forall i = 1 \ldots 3$$

$$\sigma_i = \sqrt{\frac{1}{N} \sum_{j=1}^{N} (c_{ij} - \mu_i)^2}, \ \forall i = 1 \ldots 3$$

$$s_i = \sqrt[3]{\frac{1}{N} \sum_{j=1}^{N} (c_{ij} - \mu_i)^3}, \ \forall i = 1 \ldots 3$$

[2]Le site Web correspondant au projet QBIC est http://wwwqbic.almaden.ibm.com/.

où c_{ij} est la valeur de la i^{e} composante couleur du j^{e} pixel de l'image et N le nombre total de pixels de l'image. La similarité entre images est alors définie par :

$$D_{MomCol}(I,J) = \sum_{i=1}^{3} \omega_{i1}|\mu_i^I - \mu_i^J| + \omega_{i2}|\sigma_i^I - \sigma_i^J| + \omega_{i3}|s_i^I - s_i^J|$$

où les ω_{i*} représentent les poids attribués par l'utilisateur aux différents ordres.

Généralement, les moments colorimétriques sont plus performants s'ils sont définis sur $L^*u^*v^*$ et $L^*a^*b^*$. Le troisième ordre améliore la performance des recherches comparé à l'utilisation des deux premiers ordres seulement. Cependant, le troisième ordre rend la description plus sensible aux variations de prise de vue et peut alors faire décroître la précision.

Puisque seulement 9 nombres sont utilisés pour représenter le contenu colorimétrique des images, les moments colorimétriques sont plus compacts que les autres attributs couleur mais ils ont aussi un pouvoir discriminant moins fort. Ils sont souvent destinés à réaliser une première passe du traitement avant l'emploi de descripteurs couleur plus sophistiqués.

6.2.7 Les autres statistiques

Les histogrammes couleurs n'introduisant aucune information spatiale dans la description, PASS propose le vecteur de cohérence couleur (ou *Color Coherence Vector* en anglais)[PASS96]. Les couleurs de l'histogramme sont divisées en deux types : cohérentes ou incohérentes selon qu'elles appartiennent ou non à une région étendue de couleur uniforme. Posons α_i le nombre de pixels cohérents de la i^{e} couleur et β_i celui des incohérents. Pour une image composée de N couleurs, le vecteur de cohérence couleur est alors $< (\alpha_1, \beta_1), (\alpha_2, \beta_2), \ldots, (\alpha_N, \beta_N) >$. Notons que $< \alpha_1 + \beta_1, \alpha_2 + \beta_2, \ldots, \alpha_N + \beta_N >$ correspond à l'histogramme couleur non normalisé. Du fait de l'information spatiale additionnelle, il a été montré qu'il produit de meilleurs résultats que les histogrammes couleur au niveau de la recherche d'images, et en particulier dans le cas d'images comportant des régions uniformes en terme de couleur ou de texture.

Les corrélogrammes couleurs (ou *Color Correlogram* en anglais) proposent quant à eux de caractériser la distribution couleur mais aussi la corrélation spatiale entre les couples de couleurs[HUANG97]. Les deux premières dimensions de l'histogramme à 3 dimensions sont composées des couleurs de chaque paire de pixels de l'image et la troisième code leur distance spatiale. Un corrélogramme couleur est une table indexée par les paires de couleurs où la k^{e} entrée à la position (i,j) spécifie la probabilité de trouver un pixel de couleur j à une distance k d'un pixel de couleur i. Il peut aussi être vu comme l'extension d'une matrice de cooccurrences généralement utilisée pour caractériser une texture (cf. section 6.4.1.1). Si I est l'image représentant l'ensemble de tous les pixels (avec N couleurs différentes) et I_i l'ensemble des pixels de couleur i. Le corrélogramme couleur est alors défini par :

$$\gamma_{(i,j)}^{(k)} = \frac{\text{Card}\,(p_2 \in I_j \;/\; p_1 \in I_i,\; D(p_1,p_2) = k)}{\text{Card}\,(p \in I \;/\; p_1 \in I_i,\; D(p_1,p) = k)},\; \forall (i,j) \in \zeta_I^2,\; k \in [1 \ldots K]$$

où $D(p_1,p_2)$ est la distance géométrique entre les pixels p_1 et p_2 et K la distance limite de calcul du corrélogramme. Si toutes les combinaisons possibles de couleurs sont considérées la taille de ce descripteur est très grande (dN^2). Une version simplifiée est donc généralement utilisée :

l'autocorrélogramme couleur qui ne considère que les paires de couleurs identiques dont la taille est réduite à dN. Comparés aux histogrammes couleurs et aux vecteurs de cohérences couleur, les autocorrélogrammes couleur donnent de meilleurs résultats en indexation d'images mais sont beaucoup plus coûteux à générer et à comparer à cause de leur dimension importante.

Ces deux premiers attributs sont plutôt adaptés à la description d'images globales et non à celle des régions. De plus, ils apportent une information sur la texture. Or, dans notre approche, nous désirons séparer les différentes notions utilisées (couleur, texture et forme) pour pouvoir agir très simplement par la suite sur le mélange de ces informations.

Nous pouvons également citer ici les ensembles couleurs (ou *Color Sets* en anglais), développés par SMITH et CHANG, qui sont utiles pour les problèmes d'indexation des descripteurs eux-mêmes[SMITH95]. Ils sont construits de manière à faciliter et à accélérer la recherche d'attributs semblables.

6.2.8 Comparaison des descripteurs

Nous avons mis en compétition les algorithmes les plus prometteurs présentés dans cette section. Disposant d'images très diverses, la définition d'une palette de couleur standard est impossible. Travaillant sur des supports colorimétriques variables, les méthodes testées sont : l'*Earth Mover's Distance*, les moments colorimétriques et les deux méthodes que nous avons définies (l'intersection d'histogrammes à palettes différentes (avec $k_1 = k_2 = 1$) et la projection d'histogrammes). L'*EMD* et l'intersection d'histogrammes ont également été testées avec la mesure de similarité entre couleurs définie à partir d'un noyau gaussien.

La première base d'images utilisée est la base Columbia restreinte à 17 vues de 60 objets (cf. annexe B.3). Les images entières sont utilisées pour la recherche. Des variations de couleur interviennent donc au sein d'une même catégorie. Les objets peuvent avoir des formes et des compositions colorimétriques irrégulières. La rotation des objets peut alors entraîner des variations de leur surface visible et des couleurs les composant. Une base locale a fourni une deuxième source d'information (cf. annexe B.7). Elle a été générée à partir de 12 textures, 6 manipulations de couleurs, 2 rotations et 6 formes différentes (cf. figure [6.10]). Seul l'objet central est alors pris en compte. Ces deux bases sont très complémentaires :

- la base Columbia est très générique avec des couleurs variées au sein d'une même image. Au sein d'une même catégorie, les couleurs présentes changent peu mais leur distribution peut varier fortement ;
- la base locale est très spécifique, au sein des catégories il existe des variations sur la distribution couleur tant du point de vue des couleurs présentes que de leur volume mais elles sont moins importantes que pour la base Columbia.

Sur les différents graphes précision/rappel présentés par la suite, les chiffres apparaissant entre parenthèses dans la légende sont les valeurs de $\widetilde{Rang}$ (cf. section 3.5).

Nous cherchons tout d'abord à déterminer le nombre de couleurs minimum à utiliser pour avoir une précision suffisante. La figure [6.11] présente donc les résultats obtenus pour deux méthodes différentes, l'*EMD* et l'intersection d'histogrammes utilisant toutes les deux la distance euclidienne au sein de l'espace couleur $L^*a^*b^*$. L'*EMD* étant trop complexe à calculer avec 128 couleurs, les résultats correspondants n'apparaissent pas sur le graphique. Sur la base locale, il apparaît tout

FIG. 6.10 – Images de la base locale générée
Première ligne : images d'une même catégorie ; le changement de forme engendre une légère variation de la distribution couleur ;
Deuxième ligne : différentes manipulations couleur sur une même image ; les images sont donc dans des catégories différentes ;
Troisième ligne : exemples d'objets de catégories différentes mais de couleurs semblables.

de suite que les différences sont mineures et fonction de la méthode utilisée. Par contre pour la base Columbia qui est composée d'images plus complexes en terme de couleurs, l'intersection d'histogrammes est améliorée par l'utilisation d'un plus grand nombre de couleurs. Sur cette base, nous pouvons également noter que l'*EMD* n'est pas favorisée par une telle augmentation. Pour décrire une région seules 32 couleurs semblent donc suffisantes ce qui implique que les images globales doivent être quantifiées en 64 voire 128 couleurs comme le montre les résultats obtenues sur la base Columbia.

Le choix de l'espace couleur le mieux adapté paraît également important (cf. figures [6.12] et [6.13]). Sur les bases d'images utilisées, les différences entre les espaces RVB et $L^*a^*b^*$ sont très faibles. Il apparaît même que l'espace RVB se comporte mieux sur la base Columbia que $L^*a^*b^*$ (jusqu'à 7% supérieur). En fait, $L^*a^*b^*$ est particulièrement adapté aux bases d'images acquises rigoureusement, pour lesquelles l'illuminant utilisé est constant. Or, dans le cas des bases multimédia, seules des hypothèses peuvent être réalisées quant à l'illuminant et à son invariance ce qui entraîne des résultats sensiblement identiques en RVB comme en $L^*a^*b^*$. Un soin très particulier est apportée à la prise de vue des images de fresques médiévales ; $L^*a^*b^*$ apporte alors des résultats nettement supérieurs à RVB. Sur les deux bases étudiées, les résultats obtenus sont

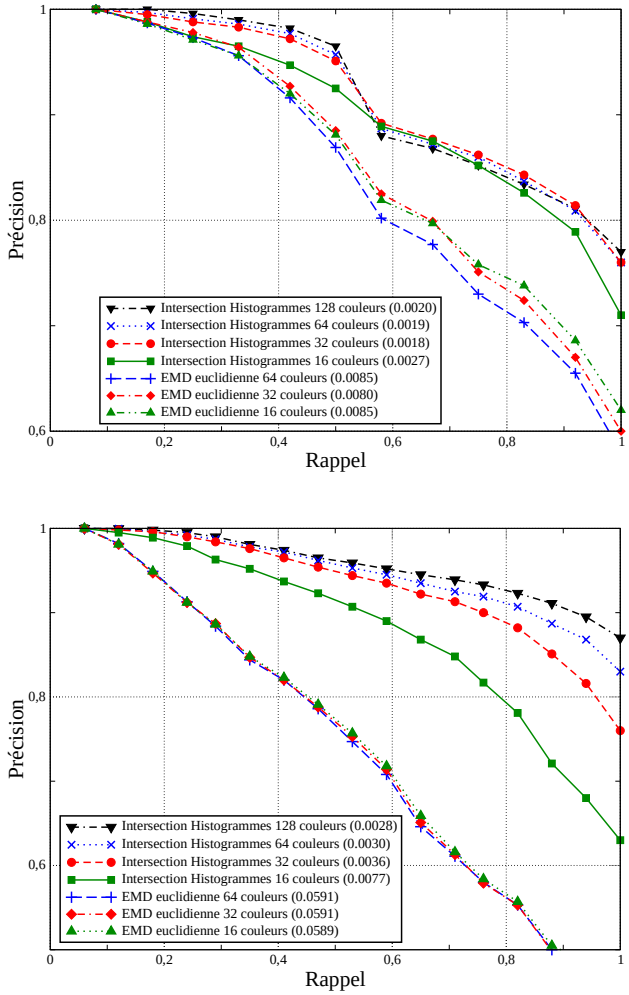


FIG. 6.11 – Détermination du nombre optimal de couleurs
En haut : base locale ; en bas : base Columbia.

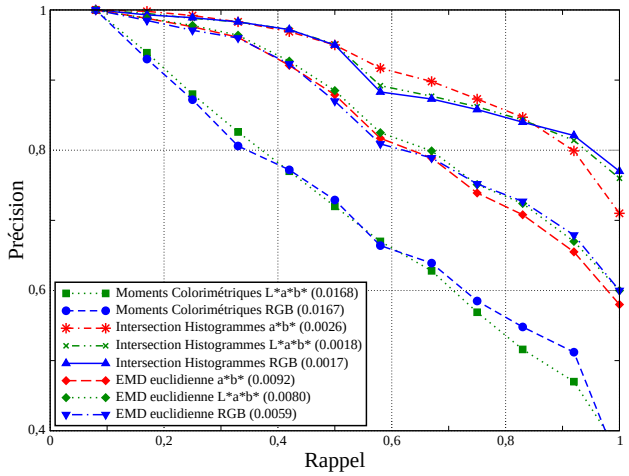


FIG. 6.12 – Détermination de l'espace couleur le mieux adapté pour la base locale

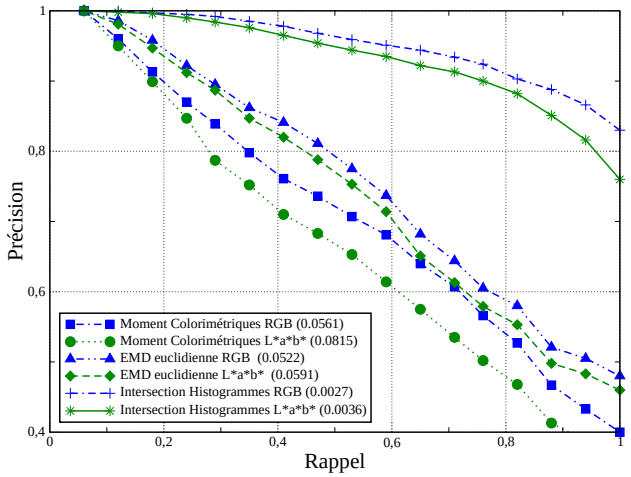


FIG. 6.13 – Détermination de l'espace couleur le mieux adapté pour la base Columbia

Nb. couleurs	Project. Histo.	*EMD*	Rapport (*EMD*/Projection)
16	0.9±0.7	21±9	23
32	4.5±3	98±10	22
64	22±5	659±39	30
128	197±22	3377±96	17
196	600±21	17086±619	28

FIG. 6.14 – Comparaison des temps de calcul de l'*EMD* et de la projection d'histogrammes
Temps de calcul en millisecondes obtenus sur un Athlon 1500+ avec 512Mo de mémoire vive. Chaque temps correspond à la comparaison de 48 histogrammes. Les résultats sont obtenus par moyennage de 10 expériences.

très comparables. Le choix se portera plutôt vers $L^*a^*b^*$ pour ses qualités d'uniformité. Nous avons également essayé de n'utiliser que les composantes chromatiques pour les recherches. Les résultats sont inférieurs mais très proches de ceux prenant en compte L^*. Cette manipulation peut être intéressante dans certains cas pour obtenir une invariance partielle aux variations de luminance (cf. section 6.2.9).

Les différents tests réalisés pour comparer l'*EMD* et notre proposition de projection d'histogrammes pour des valeurs différentes de σ et dans des espaces différents (RVB et $L^*a^*b^*$) n'ont montré que d'infimes différences dans les résultats (pas plus d'un centième de pourcent dans les courbes précision/rappel). Or, notre implémentation rend possible des comparaisons d'images avec un nombre important de couleurs (128 par exemple) alors que l'*EMD* est trop complexe pour être mise en place dans ces conditions. Le tableau [6.14] présente les temps de calcul des deux méthodes et montre que notre méthode est 25 fois plus rapide en moyenne que l'*EMD* pour un résultat identique.

Pour l'utilisation de l'évolution des distances par application d'un noyau gaussien, il faut tout d'abord rechercher la valeur de σ la plus adaptée à l'espace couleur utilisé et à la mesure de distance employée (cf. figure [6.15]). Ainsi, l'utilisation d'une distance euclidienne pour l'*EMD* ou la projection d'histogrammes en $L^*a^*b^*$ conduira à utiliser $\sigma = 1/32$ et $\sigma = 1/16$ en RVB pour ob-

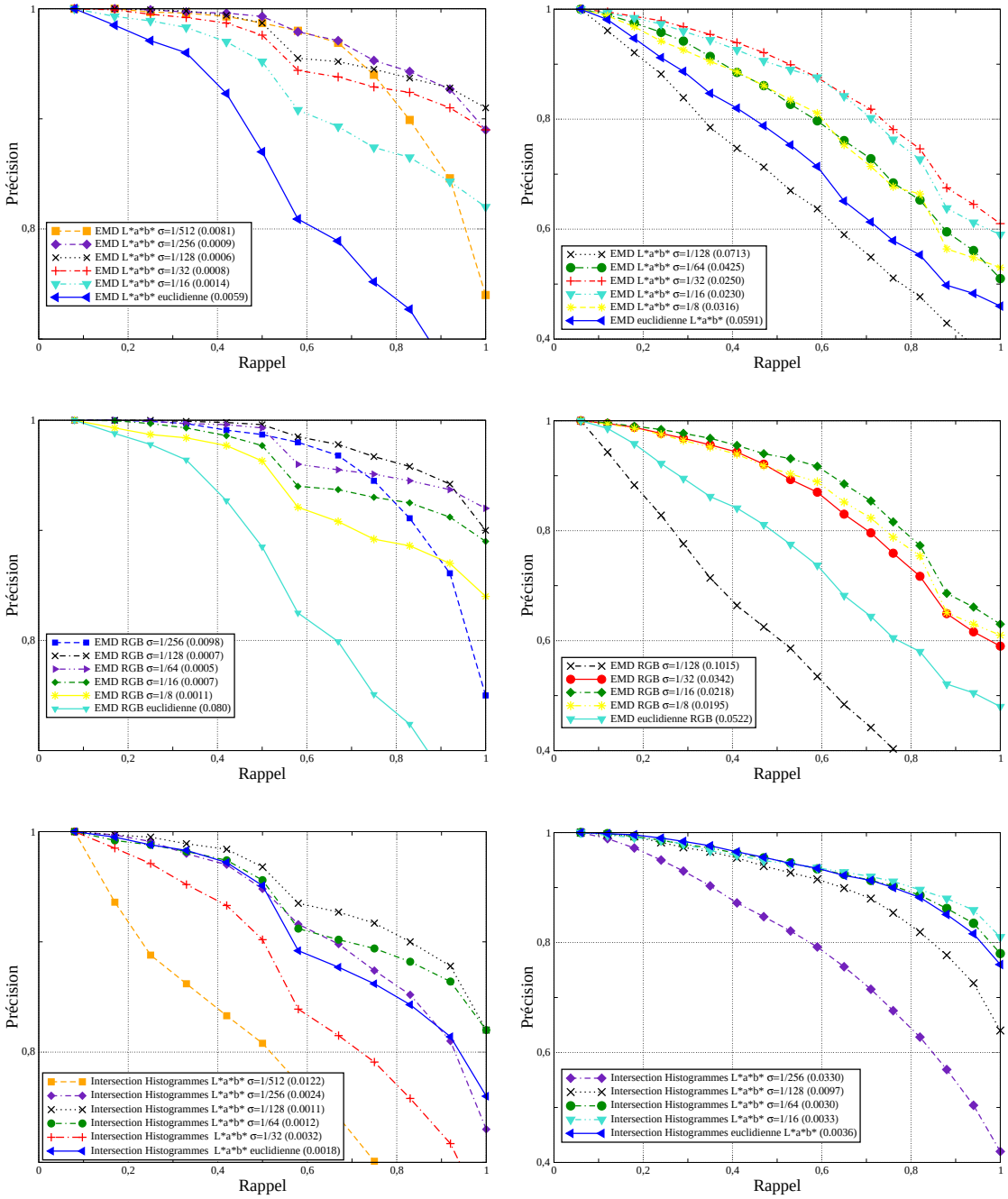


FIG. 6.15 – Influence de σ sur la qualité des mesures de similarité pour la base locale
Les résultats ont été obtenus après quantification en 32 couleurs des images.
Colonne de gauche : base locale ; colonne de droite : base Columbia.
Haut : EMD en $L^*a^*b^*$; milieu : EMD en RVB ; bas : Intersection d'histogrammes en $L^*a^*b^*$.

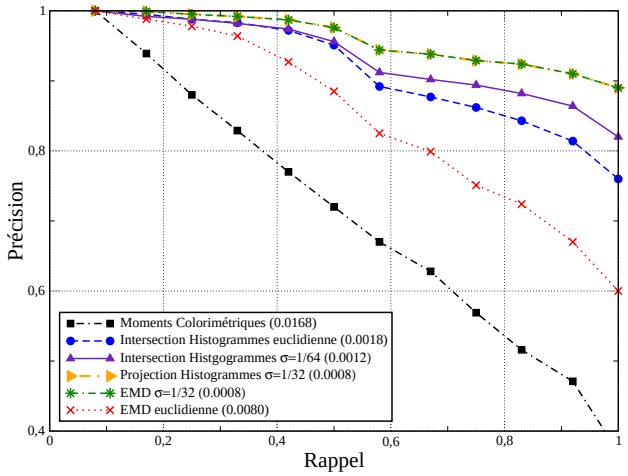


FIG. 6.16 – Performance des différents descripteurs couleur pour la base locale
Les résultats ont été obtenus après quantification en 32 couleurs des images en $L^*a^*b^*$.

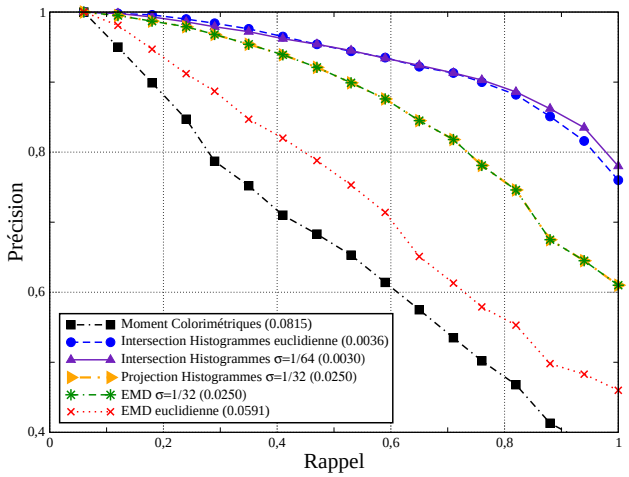


FIG. 6.17 – Performance des différents descripteurs couleur pour la base Columbia
Les résultats ont été obtenus après quantification en 32 couleurs des images en $L^*a^*b^*$.

tenir des résultats moyens optimaux. Une augmentation de σ pour la base locale augmente encore la précision. Cela est dû à la faible variabilité de la base qui permet d'être plus sélectif sans perdre en précision. Par rapport à l'utilisation d'une distance euclidienne simple, les résultats obtenus sont de bien meilleure qualité pour un grand nombre de valeurs de σ. Pour un σ non sélectif, les résultats sont proches de ceux de la distance euclidienne simple. L'augmentation progressive de la sélectivité améliore les résultats jusqu'au moment où la sélectivité est trop importante. À ce moment, les résultats deviennent aberrants dans certains cas. Il peut donc être intéressant pour certaines applications spécifiques d'effectuer un raffinement de σ afin d'obtenir une performance optimale. Pour l'intersection d'histogrammes en $L^*a^*b^*$ la valeur la plus adaptée est $\sigma = 1/64$ mais le gain en performance est moindre que dans les autres techniques. Un mauvais choix de σ pour cette méthode diminue fortement la précision. Il est sûrement plus prudent avec cette technique de ne pas utiliser l'évolution de la distance par un noyau gaussien.

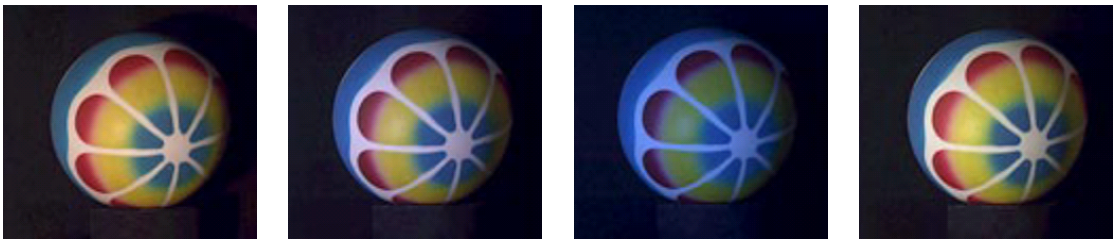


FIG. 6.18 – Un objet vu sous différentes sources lumineuses

En résumé, les graphes Précision/Rappel [6.16] et [6.17] regroupent les divers résultats des techniques testées. Globalement, il apparaît que l'intersection d'histogrammes utilisé simplement avec une distance euclidienne en $L^*a^*b^*$ fournit une mesure de similarité plus performante que l'*Earth Mover's Distance*. L'extension de la distance par un noyau gaussien semble très intéressant avec *Earth Mover's Distance* ou la projection d'histogrammes mais semble amener peu d'information supplémentaire pour l'intersection d'histogrammes. Pour des bases spécialisées telles que la base locale, cette association permet alors d'obtenir de très bonnes performances. La possibilité de faire varier σ rend possible l'adaptation de la distance aux images traitées en déterminant la sélectivité souhaitée au niveau colorimétrique.

6.2.9 Problème de la constance des couleurs ou de l'invariance à l'illuminant

Nous n'avons pas du tout parlé jusqu'à présent du problème de l'invariance des descripteurs par rapport à l'illumination. Or, ce phénomène est très important dans de nombreuses applications. En effet, les conditions sont rarement constantes durant l'acquisition d'un ensemble d'images. L'exemple le plus flagrant est la construction d'une base d'images à partir du Web. La provenance des images est la plupart du temps inconnue sans parler des conditions de prises de vue et en particulier de l'éclairage employé. Malgré cela, l'homme arrive encore à reconnaître les couleurs similaires entre deux scènes (cf. figure [6.18]). EDWIN H. LAND, inventeur du Polaroïd, fut le premier à mettre en avant ce phénomène grâce à deux expériences très significatives réalisées sur des tableaux de PIET MONDRIAN (cf. figure [6.19]) :

1. Mesurez le spectre réfléchi par un patch vert. Calibrez l'éclairage pour qu'un patch bleu réfléchisse exactement le spectre mesuré précédemment. Le patch bleu paraît toujours bleu !
2. Choisissez un patch vert positionné au sein de nombreux autres patchs de couleurs différentes. Mesurez sa réflectance et réglez votre éclairage pour qu'il émette la réflectance mesurée. Couvrez tous les autres patchs, le patch vert apparaît gris. Découvrez les patchs, il redevient vert !

Ces deux expériences, dites de MONDRIAN, montrent que l'apparence colorée dépend du spectre réfléchi mais aussi de son entourage. En fait, la réflectance relative est plus importante que la réflectance absolue.

Des tentatives de reproduction artificielle de ce phénomène ont alors été tentées. SWAIN et BALLARD notifient déjà le problème de variation d'amplitude de l'illumination qui réduit fortement la qualité des recherches pour leur système d'intersection d'histogrammes[SWAIN91]. De nombreux travaux ont alors eu pour but de régler cette question. Des descripteurs ont ainsi été dé-

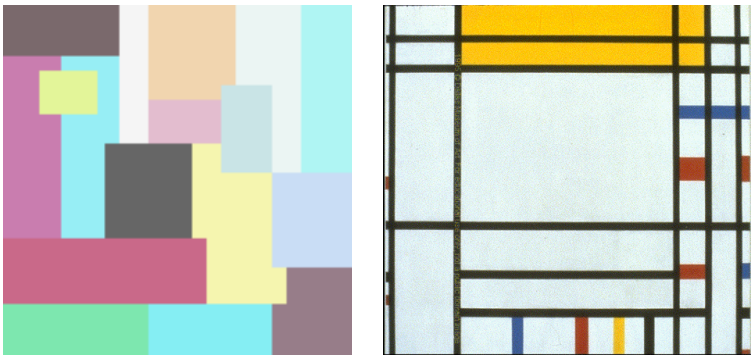


FIG. 6.19 – Tableaux de MONDRIAN utilisés par EDWIN H. LAND pour ses expériences

veloppés en tenant compte de ce phénomène. Des méthodes de normalisation colorimétrique des images ont également été recherchées.

6.2.9.1 Descripteurs invariants

Un des premiers descripteurs couleur indépendant de l'illumination a été développé par FUNT et FINLAYSON : le *color constant color indexing*[FUNT95]. Posant l'hypothèse que l'illumination est constante localement, ils indexent les images par la dérivée de son logarithme ce qui correspond en fait au rapport entre les couleurs voisines. Cette méthode est moins bonne que l'intersection d'histogramme dans le cadre d'une illumination constante mais est sensiblement meilleure quand l'illumination varie spectralement et spatialement.

Dans une publication ultérieure[FINLAY96], ils proposent le *color angular indexing* pour déterminer la similarité entre images. Elles sont alors représentées par 3 angles couleurs auxquels les auteurs ajoutent 3 angles décrivant la texture. Cette représentation très compacte offre une réduction de la complexité des comparaisons et permet d'obtenir de meilleurs résultats que leur première approche dans le cas de changements d'illumination.

Différents modèles couleurs ont également été développés par GEVERS et SMEULDERS de manière à les rendre indépendants de l'illumination et en particulier des ombres et des spots lumineux[GEVERS96, GEVERS99A]. Un des premiers modèles suppose la réflexion dichromatique et un éclairage blanc ; il est indépendant par rapport au point de vue, à l'orientation de la surface, à la direction d'éclairage, à son intensité et aux ombres. La transformation appliquée s'écrit :

$$I_1(R,G,B) = \frac{(R-G)^2}{(R-G)^2+(R-B)^2+(G-B)^2}$$

$$I_2(R,G,B) = \frac{(R-B)^2}{(R-G)^2+(R-B)^2+(G-B)^2}$$

$$I_3(R,G,B) = \frac{(G-B)^2}{(R-G)^2+(R-B)^2+(G-B)^2}$$

où R, G et B sont les coordonnées de la couleur dans l'espace RVB. Un autre modèle invariant par rapport à la couleur de l'illuminant peut être défini par :

$$m_1 = \frac{R^{x_1}G^{x_2}}{R^{x_2}G^{x_1}}, m_2 = \frac{R^{x_1}B^{x_2}}{R^{x_2}B^{x_1}}, m_2 = \frac{G^{x_1}B^{x_2}}{G^{x_2}B^{x_1}}$$

où x_1 et x_2 sont les positions dans l'image de deux pixels voisins.

MANDAL définit quant à lui des moments invariants[MANDAL96]. En supposant le changement d'illumination au sein de l'image constant et ne produisant pas d'effets non linéaires, l'histogramme de l'image après variation peut être approximé par une version translatée et mise à l'échelle de l'original. Une transformation des histogrammes permet alors de rendre les moments invariants.

6.2.9.2 Invariance par prétraitement des images

Une autre approche pour rendre la description des images invariante par rapport à l'éclairage est de les prétraiter. Ces méthodes estiment l'illuminant de l'image puis la transforme vers un illuminant de référence. FUNT et BARNARD présente ainsi une comparaison de cinq prétraitements de ce type[FUNT98] :

- L'algorithme *White Patch Retinex* compare les composantes RVB de l'illumination de référence avec le maximum de chaque canal de l'image[BRAINA86]. La correction couleur est alors réalisée par des facteurs de type $R_{réf}/R_{max}$ pour R, G et B. Il est à noter que de nombreuses versions de RETINEX sont disponibles ;
- De la même façon *GreyWorld* compare la moyenne des composantes RVB de l'image avec le gris moyen de référence. Les facteurs de normalisation sont alors du type R_{gris}/R_{moyen} ;
- Le *2D Gamut-constraint* travaille avec un gamut de référence qui représente l'enveloppe convexe des couleurs représentables sous l'illuminant de référence[FUNT96]. Le gamut des images est alors recalé sur celui de référence et détermine la transformation à effectuer pour obtenir l'image corrigée. Cette opération est réalisée dans le plan chromatique d'où le terme $2D$;
- Pour le *3D Gamut-constraint*, le principe est identique à la version $2D$ excepté le fait que le recalage est réalisé dans l'espace couleur à 3 dimensions (RVB par exemple) ;
- Un réseau de neurones peut aussi être utilisé pour ce traitement. FUNT et CARDEI utilisent ainsi un perceptron[FUNT96].

Les tests réalisés sur ces 5 méthodes montrent que malgré une amélioration de la précision de l'indexation dans le cas de changements d'illumination, ils sont fortement moins bons que le cas où l'illuminant est connu (perte de 25% de précision).

FINLAYSON propose aussi une normalisation complète des images (la *comprehensive image normalization*) qui élimine la dépendance à la géométrie et à la couleur de l'éclairage[FINLAY98]. Pour cela, des normalisations successives de l'intensité et de la couleur d'illumination sont effectuées jusqu'à stabilisation de l'image.

RETINEX a été introduit initialement par LAND et MCCANN dans le but d'étudier et de reproduire la perception humaine des couleurs[LAND71]. Dans les 30 dernières années, de nombreuses versions ont été développées. CIOCCA a récemment étudié[CIOCCA01] la performance de diverses méthodes de recherche d'images basée sur la couleur (histogrammes, vecteur de cohérence couleur...) couplées avec un préfiltrage basées sur l'algorithme *Brownian Look Up Table*

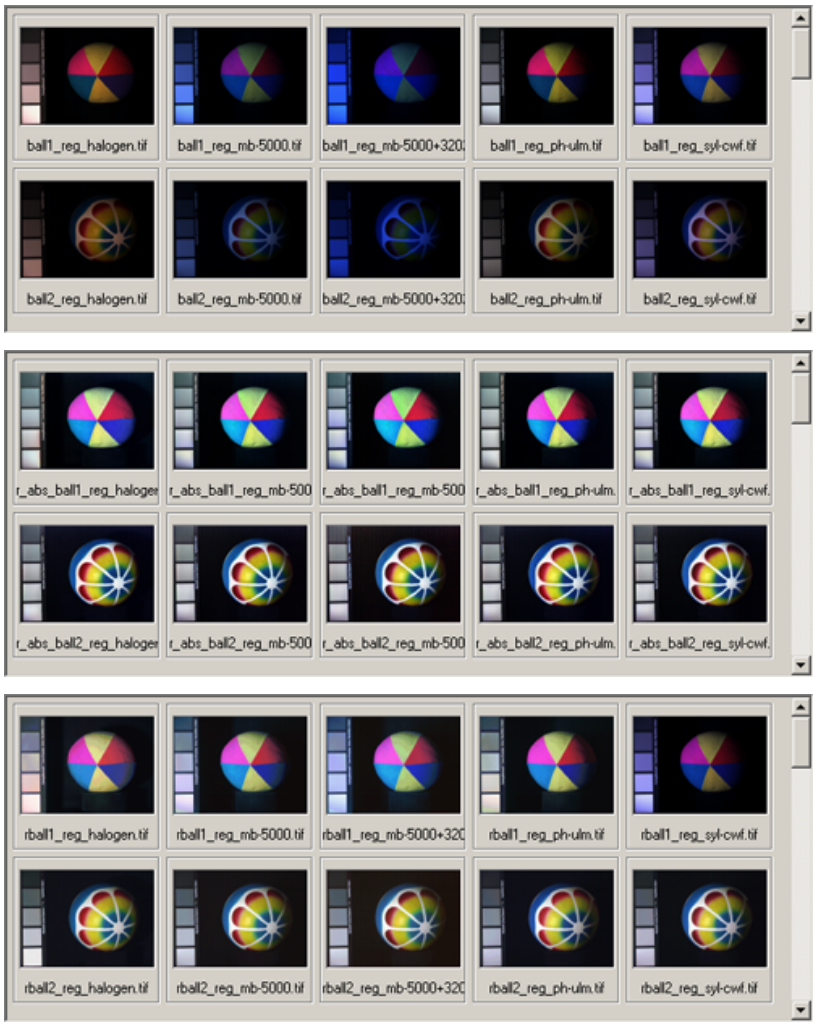


FIG. 6.20 – Normalisations d'images par RETINEX (tirées de l'étude de CIOCCA[CIOCCA01])
En haut : les images originales ;
Au milieu et en bas : normalisations par deux systèmes Retinex différents.

Retinex[MARINI00]. La conclusion de ces travaux est que RETINEX améliore la qualité de l'indexation de toutes les méthodes. La figure [6.20] présente les résultats de normalisations d'images par deux systèmes RETINEX différents (tirées de l'étude de CIOCCA).

Il paraît donc intéressant d'effectuer un tel prétraitement dès que l'ensemble des images de la base étudiée comporte des variations fortes d'illumination.

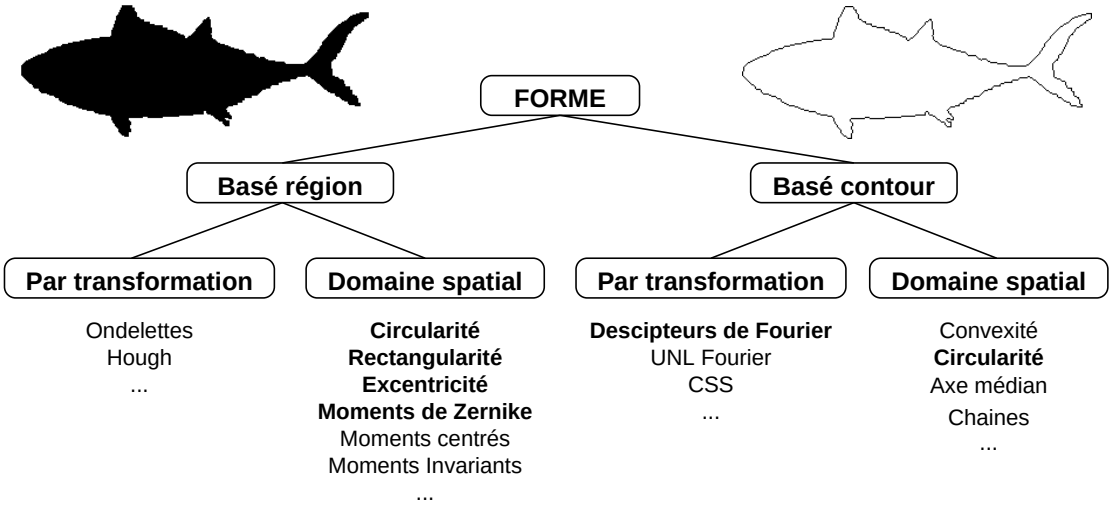


FIG. 6.21 – Organigramme des descripteurs de forme
En gras les statistiques étudiées.

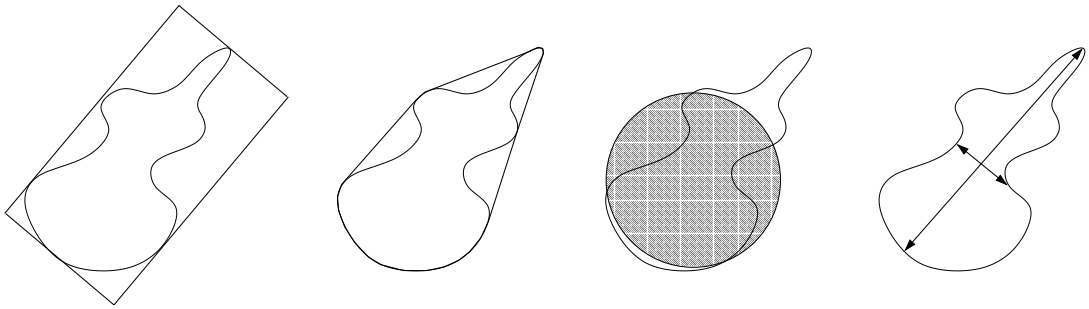
6.3 Descripteurs de forme

La forme des objets est une caractéristique importante dans la recherche de similarité entre images et surtout entre objets. Elle est très utilisée, par exemple, au niveau des tris de pièces dans l'industrie ou pour la reconnaissance de caractères. Un très grand nombre de techniques ont été proposées pour décrire la forme d'un objet. Elles sont divisées en deux grandes catégories : celles basées sur la frontière et celles basées sur la forme elle-même (cf. figure [6.21]). Ces deux approches sont présentées dans cette partie ainsi qu'une comparaison des différentes techniques retenues.

Au sein de notre graphe pyramidal, la forme est un critère important pour comparer deux régions. Cet attribut doit bien sûr être invariant par transformations affines. Malheureusement, comme nous le verrons dans ce chapitre, il n'existe pas de statistiques utilisables efficacement dans notre structure, contrairement à la couleur où les histogrammes sont sommables simplement. La concaténation de deux régions produit une nouvelle région pour laquelle la caractérisation de la forme ne peut pas se déduire directement des deux précédentes. Pour les approches basées sur la frontière, il apparaît facilement que la somme des frontières de deux régions adjacentes ne donne pas celle de l'ensemble car la partie correspondante à l'adjacence est commune aux deux régions. De même, les statistiques basées sur la forme sont généralement calculées par rapport au centre de gravité de la forme ; le centre de gravité de l'ensemble étant différent de ceux des régions, il n'y a aucun lien entre les descripteurs. Actuellement, nous n'avons trouvé aucune statistique possédant des propriétés additives satisfaisantes ; pour chaque région l'attribut de forme devra donc être recalculé entièrement.

6.3.1 Définitions

La forme d'une région est décrite sous la forme d'une image I de taille $[l, h]$. Cette image définit alors la boîte englobante de l'objet. Les points de l'image I sont donc tous les couples (x, y) où $x \in [0, l-1]$ et $y \in [0, h-1]$.



Rectangularité Convexité Compacité Excentricité

FIG. 6.22 – Caractéristiques géométriques simples de forme

Le masque définissant la forme est alors la fonction :
$f(x,y) : (x,y) \in I \rightarrow \mathbb{B}$ où $f(x,y) = vrai$ si le point (x,y) fait partie de la forme.
De même, la frontière de la région est décrit par la fonction :
$g(x,y) : (x,y) \in I \rightarrow \mathbb{B}$ où $g(x,y) = vrai$ si le point (x,y) fait partie de la frontière.

6.3.2 Caractéristiques géométriques simples

Dans la littérature, un grand nombre de statistiques simples ont été développées pour caractériser la forme d'un objet. Celles-ci sont presque toutes invariantes par transformations affines. Quelques unes d'entre elles sont présentées de manière graphique sur la figure [6.22].

Rectangularité : elle représente la place prise par l'objet dans sa boîte englobante. Elle est définie par

$$\textit{Rectangularité} = \frac{\mathcal{A}}{\mathcal{A}_{RectEnglo}}$$

où $\mathcal{A}$ est l'aire de la forme et $\mathcal{A}_{RectEnglo}$ l'aire de sa boîte englobante rectangulaire.

Convexité : elle indique la régularité de la forme. Sa valeur correspond au rapport du périmètre de l'enveloppe convexe $\mathcal{P}_{EnvConv}$ sur celui de la forme $\mathcal{P}$:

$$\textit{Convexité} = \frac{\mathcal{P}_{EnvConv}}{\mathcal{P}}$$

La détermination de l'enveloppe convexe d'une forme peut être réalisée brutalement en $\mathcal{O}(n^2)$. Des algorithmes linéaires existent tout de même pour résoudre ce problème[3].

Compacité ou circularité : elle apporte des informations sur la complexité de la forme. La compacité est souvent définie de la manière suivante :

$$\textit{Compacité} = \frac{\mathcal{P}^2}{\mathcal{A}}$$

[3]L'histoire de la recherche de la linéarité dans ce domaine est présentée au travers différentes méthodes correctes ou incorrectes à l'adresse http://cgm.cs.mcgill.ca/~athens/cs601/. Elle est proposée par GREG ALOUPIS et s'intitule « *A History of Linear-time Convex Hull Algorithms for Simple Polygons* ».

où $\mathcal{A}$ est l'aire de la forme et $\mathcal{P}$ son périmètre. Sa valeur sera grande pour un objet allongé et bruité mais faible pour un objet rond. Sa valeur minimum, 4π, est d'ailleurs obtenue pour un cercle. Une autre mesure possible est égale au rapport du périmètre d'un cercle de même aire que la forme $\mathcal{P}_{Cercle}$ à celui de la forme $\mathcal{P}$:

$$\textit{Compacité} = \frac{\mathcal{P}_{Cercle}}{\mathcal{P}} = \frac{2\sqrt{\mathcal{A}\pi}}{\mathcal{P}} \text{ car } \mathcal{A}_{Cercle} = \mathcal{A}.$$

Avec cette définition, la valeur minimale est toujours obtenue pour un cercle mais vaut 1.

Moments centrés : Pour l'image I les moments d'ordre $(p+q)$ sont définis par :

$$m_{pq} = \sum_{(x,y)\in I} x^p y^q f(x,y),\ \forall (p,q) \in \mathbb{N}^2$$

Les moments centrés s'écrivent alors :

$$\mu_{pq} = \sum_{(x,y)\in I} (x-\bar{x})^p (y-\bar{y})^q f(x,y), \text{ avec } \bar{x} = \frac{m_{10}}{m_{00}} \text{ et } \bar{y} = \frac{m_{01}}{m_{00}}.$$

$\bar{x}$ et $\bar{y}$ sont donc les coordonnées du centre de gravité de la forme. Ces moments sont invariants par translation et peuvent aussi le devenir par changement d'échelles après normalisation de l'aire de l'objet. Par contre, ils ne le sont pas par rotation ; les moments de HU (cf. section 6.3.3.1) sont une combinaison de ceux-ci et fournissent 7 combinaisons de moments qui sont invariants par rotation.

Excentricité ou rapport des axes principaux : C'est le rapport entre l'axe principal et l'axe secondaire de la forme qui mesure l'allongement de la forme. Elle peut être calculée en utilisant les moments centrés d'ordre 2 :

$$\epsilon = \frac{(\mu_{20}-\mu_{02})^2 + 4\mu_{11}^2}{(\mu_{20}+\mu_{02})^2}$$

Elle est comprise entre 0 et 1 et vaut 1 pour un cercle.

Descripteurs basés sur les trous : Dans le cadre de l'analyse de pièces manufacturées, des valeurs comme le rapport entre l'aire des trous et celle de l'objet ou encore celui entre la surface de contact trous/matière et la surface de la matière sont utilisées. Elles sont adaptées à des applications précises où les trous des objets sont caractéristiques de la forme mais elles sont complètement inutilisables de manière générale.

Ces premiers descripteurs utilisés seuls n'apportent que très peu d'informations sur la forme. Seuls les moments centrés peuvent fournir une statistique assez robuste, quoique non invariante par rotation, s'ils sont utilisés en nombre suffisant.

6.3.3 Descripteurs basés sur la région

6.3.3.1 Moments de HU

Ces moments sont dérivés des moments centrés normalisés. Ils sont invariants par transformations affines et par réflexion[HU62]. Ils sont très utilisés de par leurs propriétés d'invariance pour de la recherche d'images par le contenu[OZER99] ou d'identification d'objets et même de gestes[BOBICK96].
HU montre tout d'abord que les moments produits par la normalisation suivante[4] sont invariants par changement d'échelles :

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\left(1+\frac{p+q}{2}\right)}}, \text{ pour } p+q \geq 2 \text{ et } (p,q) \in \mathbb{N}^2.$$

L'invariance par rotation est alors obtenue par combinaison des η_{pq}. HU présente ainsi les 7 moments d'ordres 2 et 3 suivants :

$$\begin{aligned}
\phi_1 &= \eta_{20} + \eta_{02} \\
\phi_2 &= (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \\
\phi_3 &= (\eta_{30} - 3\eta_{12})^2 + 3(\eta_{21} - \eta_{03})^2 \\
\phi_4 &= (\eta_{30} + \eta_{12})^2 + (\eta_{21} - \eta_{03})^2 \\
\phi_5 &= (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})\left[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2\right] + \\
&\quad (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})\left[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2\right] \\
\phi_6 &= (\eta_{20} - \eta_{02})((\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2) + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \\
\phi_7 &= (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})\left[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2\right] + \\
&\quad (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})\left[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2\right]
\end{aligned}$$

LI énumère 52 moments de type HU d'ordres 2 à 9 invariants par translation, rotation et changement d'échelles[LI92]. BELKASIM en liste quant à lui 32 d'ordres 2 à 7 ; il en identifie d'ailleurs moins que LI aux ordres 2 à 7[BELKAS91]. YANG propose une méthode de calcul rapide de ses moments[YANG94].

D'autres moments invariants ont été définis par FLUSSER[FLUSSE93] et TAUBIN[TAUBIN92]. PARK montre l'invariance de ces différents moments par rapport à la rotation et au changement d'échelles dans [PARK99A].

6.3.3.2 Moments de ZERNIKE

ZERNIKE définit en 1934 une famille de polynômes complexes, ZP, formant une base orthogonale dans le cercle unité ($x^2 + y^2 \leq 1$) :

$$ZP = \{V_{nm}(x,y) \ / \ x^2 + y^2 \leq 1\}$$

où les polynômes complexes V_{nm} d'ordre n et de répétition m sont définis sous les conditions $n \in \mathbb{N}^+$, $m \in \mathbb{N}$, $n - |m|$ pair et $|m| \leq n$:

$$V_{nm}(x,y) = R_{nm}(r)e^{jm\theta}$$

[4] Notez que dans le papier original de HU, cette normalisation est écrite avec une erreur de typographie.

en posant $r = \sqrt{x^2 + y^2}$, $\theta = \arctan(\frac{y}{x})$ et la partie radiale R_{nm} :

$$R_{nm}(r) = \sum_{s=0}^{\frac{(n-|m|)}{2}} (-1)^s \frac{(n-s)!}{s! \left(\frac{n+|m|}{2} - s\right)! \left(\frac{n-|m|}{2} - s\right)} r^{n-2s}$$

R_{nm} est donc un polynôme composé de termes en r^q où $q \in [n, n-2, ..., |m|]$.

KHOTANZAD et HONG définissent alors les moments de ZERNIKE $|A_{nm}|$ comme l'amplitude de la projection de la forme $f(x, y)$ sur la base orthogonale des fonctions $V_{nm}(x, y)$[KHOTAN90]. A_{nm} est alors le nombre complexe correspondant à la projection :

$$A_{nm} = \frac{n+1}{\pi} \sum_{(x,y) \in I} V_{nm}(x, y)^* f(x, y)$$

avec * le symbole du conjugué complexe et $x^2 + y^2 \leq 1$. Cette dernière condition implique que la forme doit préalablement être normalisée au sein du disque unité pour être entièrement caractérisée.

Les moments ainsi définis sont invariants par rotation, translation et changement d'échelles (après normalisation de la taille de la forme). De par l'utilisation d'une base orthogonale, l'information portée est nettement moins redondante que pour les autres types de moments.

Cette représentation étant inversible, l'image peut être reconstruite à partir des moments calculés :

$$\hat{f}(x, y) = \lim_{N \to \infty} \sum_{n=0}^{N} \sum_{m} A_{nm} V_{nm}(x, y)$$

où m prend toutes le valeurs possibles telles que $|m| < n$ et $n - |m|$ pair. La limitation de cette somme apporte une reconstruction partielle de l'image originale et permet d'appréhender l'information portée par les différents moments. La figure [6.23] propose des exemples de reconstruction de formes à partir des moments de ZERNIKE. Il apparaît que peu de moments permettent de décrire correctement des formes simples (le cœur) mais pour les autres, plus complexes (l'éléphant), la reconstruction n'est pas parfaite même en allant jusqu'au 45^{e} ordre mais semble suffisante dans un cadre de mesure de similarité.

CHONG a présenté dernièrement une comparaison des divers algorithmes d'optimisation de calcul des moments de ZERNIKE[CHONG03].

6.3.3.3 Axe médian

La transformée d'axe médian est sûrement la plus populaire et la plus étudiée des techniques de description de forme basées sur la région. Elle a été proposée initialement par BLUM[BLUM67]. L'idée de cette approche est de représenter la forme en utilisant un graphe et en espérant que les caractéristiques importantes seront préservées dans ce graphe. Les termes squelettes, graphes de chocs (*shock graphs* en anglais), transformée d'axe symétrique ou *prairie fire transform* font référence à la même approche.

BLUM avançait l'idée que le processus de formation d'une image sur la rétine est similaire à la propagation d'un feu dans une prairie. Le feu ne peut revenir en arrière. En fait, s'il est allumé à la

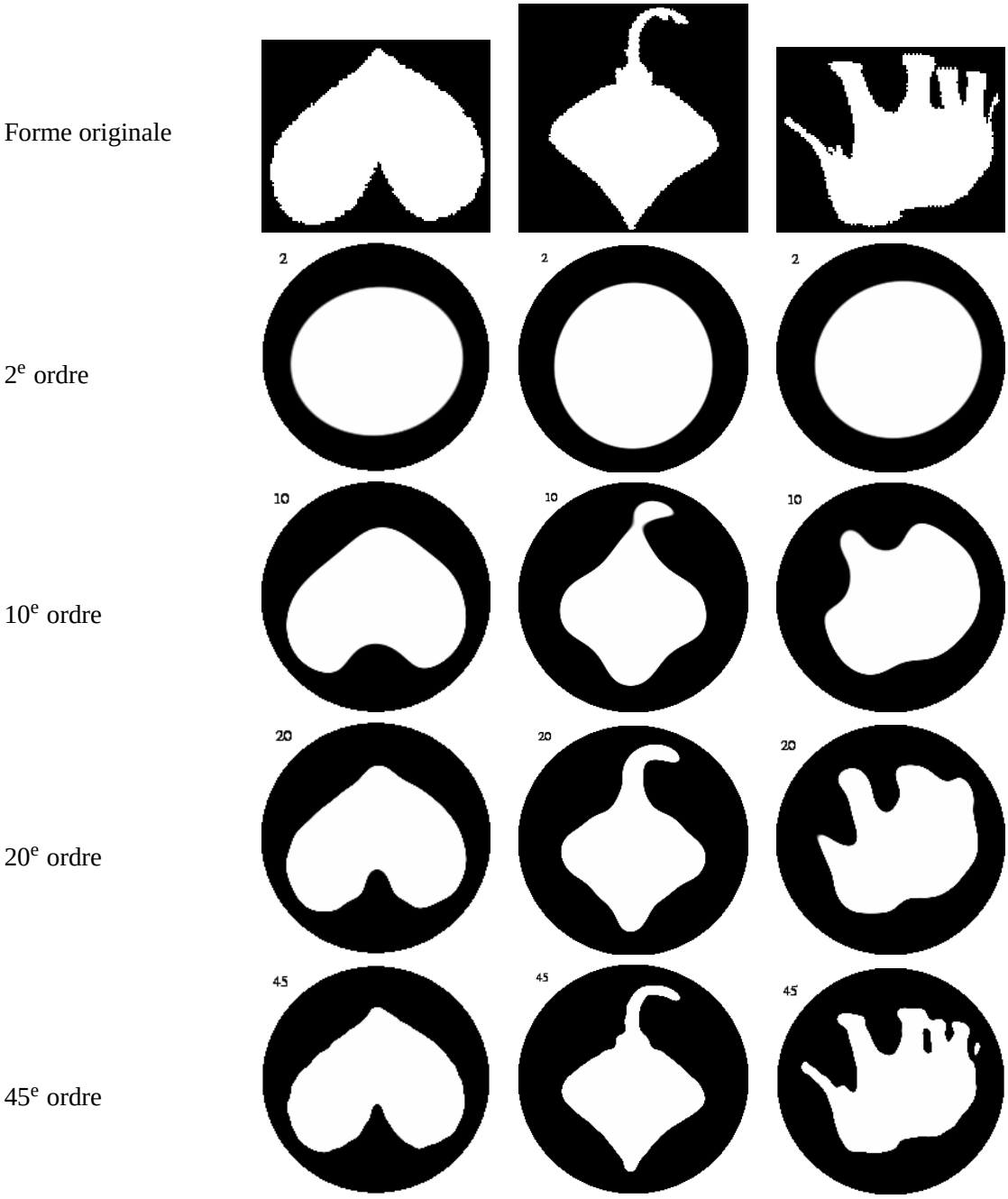
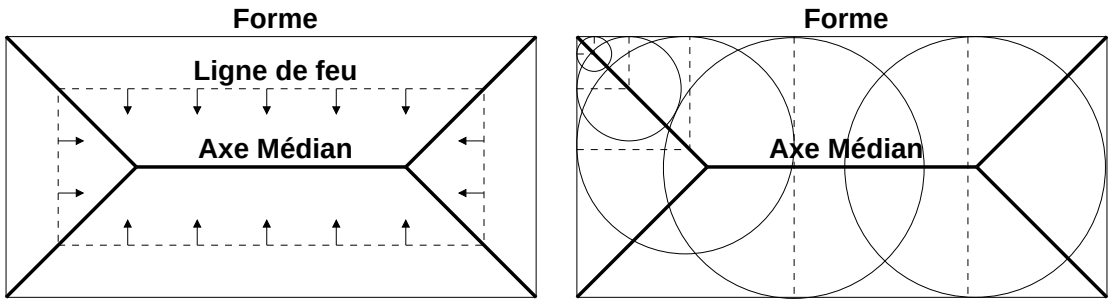


FIG. 6.23 – Reconstruction de formes à partir des moments de ZERNIKE



a) b)

FIG. 6.24 – Axe médian d'un rectangle
a) méthode originale b) généralisation de la méthode.

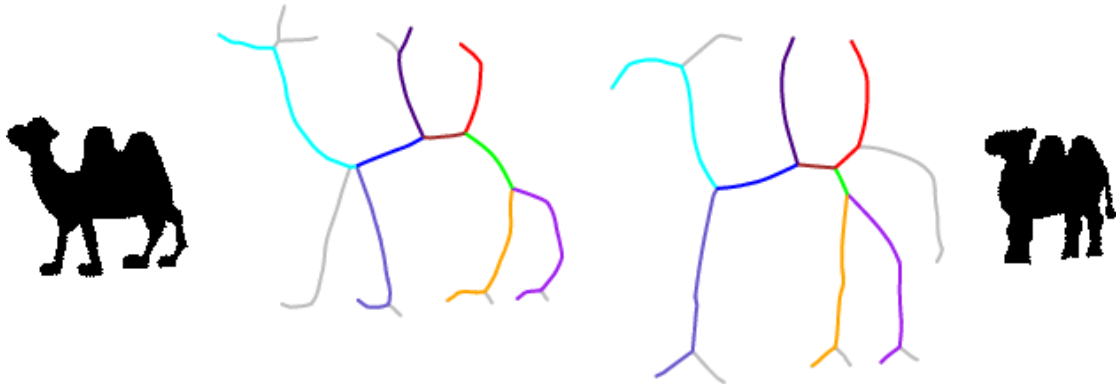


FIG. 6.25 – Association de graphes d'axe médian (tirée de [SEBAST01])

frontière de l'objet il se propage de manière régulière vers son centre. L'axe médian est ainsi défini comme l'ensemble des points où deux fronts de feu se rejoignent (cf. figure [6.24].a). Cela revient à extraire un squelette de la forme. Malheureusement ce mode de construction est très sensible au bruit, c'est pourquoi BLUM et NAGEL proposent la méthode d'axe médian généralisé[BLUM78]. Les points de l'axe médian sont alors les centres des cercles intérieurs et tangents deux fois à la forme de rayon supérieur à un seuil (cf. figure [6.24].b).

Une fois l'axe médian détecté, de nombreuses méthodes basées sur la recherche de similarité entre graphes sont possibles. La méthode d'assignement gradué[GOLD96A] (cf. section 7.2.2.5) est ainsi utilisée par SHARVIT pour déterminer la ressemblance entre les *shock graphs*[SHARVI98]. Les derniers travaux de SEBASTIAN déterminent les déformations simples à réaliser pour passer d'une configuration à une autre[SEBAST01]. Chaque opération ayant un coût, la mesure de similarité entre graphes est obtenue une fois trouvé l'ensemble des manipulations fournissant le coût total minimum (cf. figure [6.25]).

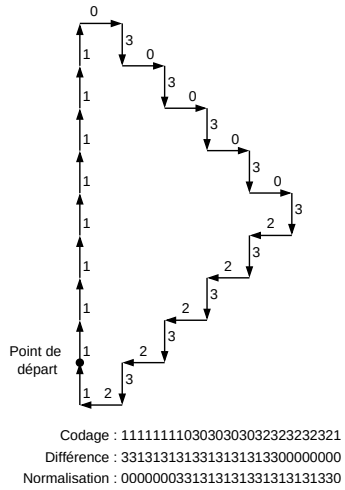


FIG. 6.26 – Codage et normalisation de la chaîne en 4-connexité

6.3.4 Descripteurs basés sur la frontière

6.3.4.1 Chaînage du contour

À partir des deux systèmes de codage des contours présentés à la section 4.2.2, des techniques ont été développées pour comparer deux formes. Les formes standards des codes de FREEMAN et des *crack-codes* sont invariantes par translation. L'invariance par changement d'échelles peut être atteinte par normalisation de la taille de la forme.

Pour rendre cette représentation insensible à la rotation, une transformation de la chaîne est effectuée en codant les changements de direction plutôt que les directions. Pour cela, la différence (dans le sens des aiguilles d'une montre) entre chaque direction et la précédente donne une chaîne invariante par rotation. En 4-connexité par exemple, le passage de 0 à 3 est codé 3 et celui de 1 à 3 donne 2. De plus, cette chaîne est généralement transformée par rotation de l'ensemble des symboles pour obtenir la chaîne maximale (au sens numérique du terme) (cf. figure [6.26]).

La comparaison de deux contours peut être réalisée en considérant les codes comme des chaînes de caractères et en appliquant des algorithmes de recherche de sous-chaînes. Il faut bien évidemment dans cette partie faire attention au fait que le point de départ de la chaîne a été choisi arbitrairement. La chaîne doit donc être considérée comme cyclique. La ressemblance entre deux formes est alors obtenue par extraction de la sous-chaîne commune maximale.

La notion d'histogramme de chaînes est aussi avancée par IIVARINEN[IIVARI97]. L'histogramme des différentes directions prises par le contour peut permettre de caractériser statistiquement une forme. Cette méthode est peu complexe et apporte des résultats satisfaisants quand les différentes classes d'objets sont bien distinctes ce qui limite fortement son utilisation.

6.3.4.2 Représentation *CSS*

La représentation *Curvature Scale Space* (*CSS*) est une organisation multi-échelles des points de passage par zéro de la courbure proposée par MOKHTARIAN et MACKWORTH[MOKHTA88, MOKHTA95]. C'est d'ailleurs la méthode de description de formes intégrée dans MPEG-7.

La courbure est définie comme la dérivée de l'angle tangent à la courbe. Soit $\overrightarrow{r}$ une définition vectorielle paramétrique d'une forme :

$$\overrightarrow{r}(u) = (x(u), y(u))$$

où u est un paramètre arbitraire. La fonction de courbure est alors :

$$\kappa(u) = \frac{\dot{x}(u)\ddot{y}(u) - \ddot{x}(u)\dot{y}(u)}{(\dot{x}^2(u) + \dot{y}^2(u))^{\frac{3}{2}}} \tag{6.1}$$

où $\dot{x}$ et $\ddot{x}$ sont les dérivées premières et secondes de x par rapport à u (idem pour y).

Si Γ est une courbe fermée plane, le paramètre u peut être l'abscisse curviligne normalisée ce qui induit :

$$\Gamma = \{(x(u), y(u)) \ / \ u \in [0, 1]\}$$

De nombreuses approches existent pour calculer la courbure digitale de la courbe. MOKHTARIAN et MACKWORTH utilisent le principe d'évolution de la courbe qui étudie les propriétés de la courbe tout en la déformant dans le temps. Une évolution est réalisable par lissage gaussien permettant de calculer la courbure à différents niveaux de détails. Soit $g(u, \sigma)$ un noyau gaussien $1D$ de largeur σ. $X(u, \sigma)$ et $Y(u, \sigma)$ représentent alors les composantes de la courbe « évoluante » :

$$X(u, \sigma) = x(u) \otimes g(u, \sigma) \qquad Y(u, \sigma) = y(u) \otimes g(u, \sigma)$$

Les propriétés de l'opération de convolution ($\otimes$) permettent de déterminer les dérivées des composantes facilement :

$$X_u(u, \sigma) = x(u) \otimes g_u(u, \sigma) \qquad X_{uu}(u, \sigma) = x(u) \otimes g_{uu}(u, \sigma)$$

Des relations similaires sont obtenues pour $Y_u(u, \sigma)$ et $Y_{uu}(u, \sigma)$. Les formes exactes de $g_u(u, \sigma)$ et $g_{uu}(u, \sigma)$ étant connues, la courbure de la forme digitale évoluante est calculable facilement :

$$\kappa(u, \sigma) = \frac{X_u(u, \sigma)Y_{uu}(u, \sigma) - X_{uu}(u, \sigma)Y_u(u, \sigma)}{(X_u(u, \sigma)^2 + Y_u(u, \sigma)^2)^{\frac{3}{2}}}$$

σ augmentant, la forme Γ_σ évolue. Ce processus de génération d'une séquence ordonnée de courbes correspond à l'évolution de Γ (cf. figure [6.27]). La courbe se rétrécit et devient de plus en plus lisse jusqu'à ne plus avoir de point de passage de la courbure par zéro quand σ est suffisamment grand.

Les points de passages de la courbure par zéro pouvant être déterminés à chaque niveau de Γ_σ, ils sont reportés sur le plan (u, σ) où u est l'abscisse curviligne normalisée et σ la largeur du

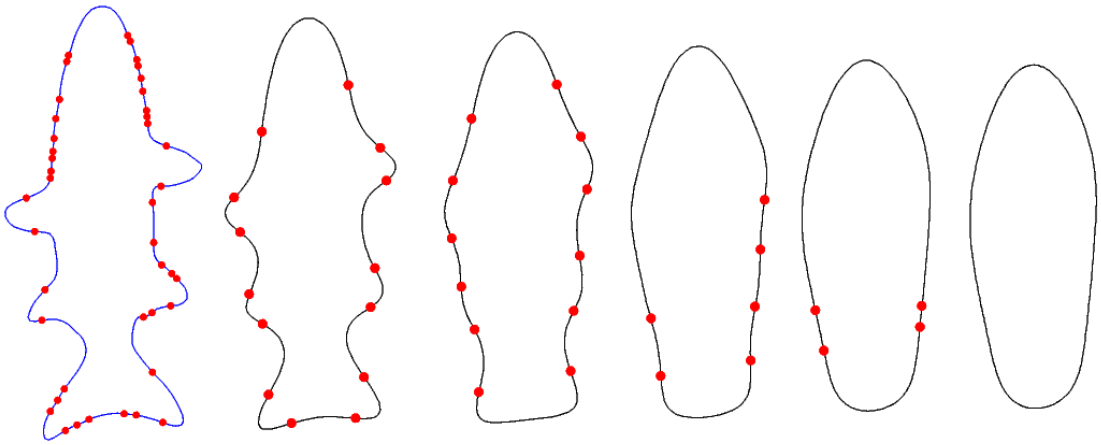


FIG. 6.27 – Rétrécissement et lissage de la forme au cours du processus *CSS*[MOKHTA96]
Le nombre de points de passage de la courbure par zéro diminue au cours de l'évolution (symbolisés par des points). De gauche à droite : $\sigma = 1, 4, 7, 10, 12, 14$.

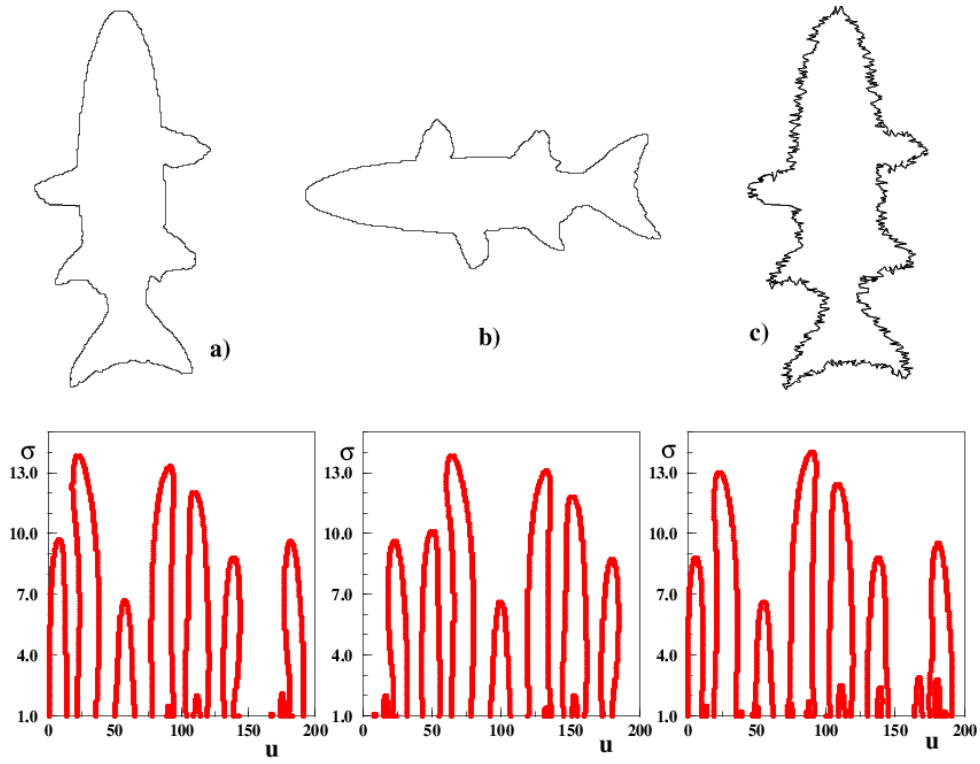


FIG. 6.28 – Représentation *CSS* d'une forme, de sa rotation et d'une version bruitée[MOKHTA96]
La normalisation des formes est réalisée en rapportant le nombre de points du contour à 200.
a) forme originale et son image CSS ;
b) rotation de la forme. L'image CSS subit un décalage circulaire ;
c) Le bruit engendre uniquement la création de petits contours dans l'image CSS.

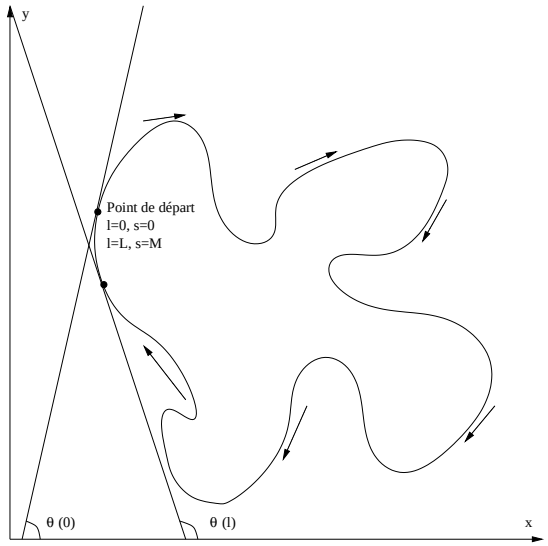


FIG. 6.29 – Direction angulaire pour les descripteurs de FOURIER

noyau gaussien. Cette représentation sous forme binaire est appelée image *CSS* de la courbe[5] (cf. figure [6.28].a). Les intersections d'une ligne horizontale avec les contours de cette image indiquent les positions des points de passage par zéro de la courbure sur la forme Γ_σ correspondante.

De par la normalisation du contour, cette représentation est évidemment invariante par changement d'échelles. La rotation de la forme engendre uniquement un décalage de l'image *CSS* ; cet effet est exactement le même que celui dû à un changement de l'origine de l'abscisse curviligne (cf. figure [6.28].b). Le bruit génère uniquement de petits contours qui ne sont pas significatifs par rapport au reste de la représentation (cf. figure [6.28].c). Ils ne seront d'ailleurs pas pris en compte dans les divers algorithmes de comparaison de représentation *CSS*.

Le *matching* est ensuite réalisé à partir des maxima significatifs des images. Ceux correspondant au bruit ne sont donc pas pris en compte. Les différents nuages de points peuvent ensuite être mis en correspondance pour déterminer la similarité entre les formes. Il faut à ce niveau prendre garde au fait que la description peut être décalée à cause d'un changement de point de départ ou une rotation.

6.3.4.3 Descripteurs de FOURIER

Cette technique permet la caractérisation d'un contour fermé Γ[PERSOO77]. Supposons Γ orienté dans le sens des aiguilles d'une montre (cf. figure [6.29]) ; sa représentation paramétrique est alors fonction de l'abscisse curviligne s : $Z(s) = (x(s), y(s))$ avec s variant de 0 à L la longueur du contour. Trois types de représentation sont possibles : la courbure, la distance au centre et la fonction de coordonnées complexes.

[5] Le système SQUID (*Shape Queries Using Image Databases*) propose une démonstration animée. Elle peut être trouvée à l'adresse http://www.ee.surrey.ac.uk/Research/VSSP/imagedb/demo.html.

Notons $\theta(s)$ la direction angulaire de Γ au point d'abscisse curviligne s, la courbure de la forme est définie par :

$$K(s) = \frac{d}{ds}\theta(s)$$

La distance au centre est la distance des points de la forme au centre de gravité (x_c, y_c) de celle-ci :

$$R(s) = \sqrt{(x_s - x_c)^2 + (y_s - y_c)^2}$$

Les coordonnées complexes de la forme sont également définies à partir du centre de gravité :

$$Z(s) = (x_s - x_c) + j(y_s - y_c)$$

La transformée de FOURIER de ces trois types de représentations donnent trois ensembles de coefficients complexes représentants la forme de l'objet dans le domaine fréquentiel. Les basses fréquences apportent de l'information sur l'aspect global de la forme alors que les hautes fréquences précisent les détails. De manière à être invariant par rotation, seules les amplitudes des coefficients sont utilisées. L'invariance par changement d'échelles est obtenue par division des éléments par la composante continue où le premier coefficient non nul. L'indépendance par rapport aux translations est intrinsèque à ces représentations.

Les descripteurs de FOURIER basés sur la courbure f_K et ceux basés sur la distance au centre f_R s'écrivent alors :

$$f_K = \left[|F_1|, |F_2|, \ldots, |F_{M/2}|\right]$$

$$f_R = \left[\frac{|F_1|}{|F_0|}, \frac{|F_2|}{|F_0|}, \ldots, \frac{|F_{M/2}|}{|F_0|}\right]$$

où F_i est le i^{e} coefficient de la transformée de FOURIER. Dans ces deux cas, seuls les coefficients positifs sont utilisés car les deux fonctions étant à valeurs réelles leur transformée de FOURIER est symétrique : $|F_i| = |F_{-i}|$.

Pour les coordonnées complexes, les descripteurs de FOURIER sont :

$$f_Z = \left[\frac{|F_{-(M/2-1)}|}{|F_1|}, \ldots, \frac{|F_{-1}|}{|F_1|}, \frac{|F_2|}{|F_1|}, \ldots, \frac{|F_{M/2}|}{|F_1|}\right]$$

où F_1 est la première composante fréquentielle non nulle utilisée pour normaliser les autres coefficients. Les valeurs correspondantes aux fréquences négatives et positives sont considérées dans ce cas. La composante continue est dépendante de la position de la forme ; elle n'est donc pas utilisée.

Pour obtenir des descriptions de mêmes tailles, les différentes formes sont rééchantillonnées à M points avant d'appliquer la transformée de FOURIER. ZHANG montre que les trois types de descriptions ne sont pas équivalentes et que la distance au centre donne de meilleurs résultats[ZHANG01]. Dans [ZHANG02], il compare également les descripteurs de FOURIER et la représentation *CSS*. Les résultats obtenus sont fortement en faveur des descripteurs de FOURIER tant du point de vue des performances pour l'indexation que pour la complexité de génération et de comparaison des descriptions.

Cette description n'est valable que pour des contours fermés. RAUBER propose ainsi une évolution permettant de décrire des contours quelconques, la représentation UNL (pour *Universidade*

Descripteur	Invariance	Stockage	Complexité	Indexable
Compacité	TRS	$\mathcal{O}(1)$	$\mathcal{O}(LH)$	Oui
Convexité	TRS	$\mathcal{O}(1)$	$\mathcal{O}(l)$	Oui
Rectangularité	TRS	$\mathcal{O}(1)$	$\mathcal{O}(LH)$	Oui
Moments centrés	TS˜	$\mathcal{O}(1)$ pour chaque ordre	$\mathcal{O}(LH)$	Oui
Excentricité	TRS	$\mathcal{O}(1)$	$\mathcal{O}(LH)$	Oui
Moments de HU	TRS	$\mathcal{O}(1)$	$\mathcal{O}(LH)$	Oui
Axe médian	TRS	$\mathcal{O}$(segments du squelette)	$\mathcal{O}(LH)$	Non
Moments de ZERNIKE	TRS	$\mathcal{O}(1)$ pour chaque ordre	$\mathcal{O}(LH)$	Oui
Représentation *CSS*	TRS	$\mathcal{O}$(nombre maxima)	$\mathcal{O}(cg\sigma_c)$	Non
Chaînage du contour	TR˜S˜	$\mathcal{O}(l)$	$\mathcal{O}(l)$	Oui
Descripteurs de FOURIER	TRS	$\mathcal{O}$(fréq. échantillonnage)	$\mathcal{O}(l)$	Oui

FIG. 6.30 – Comparaison des caractéristiques des descripteurs de formes étudiés
Invariances :
T : translation, R : rotation, S : changement d'échelles et ˜ : oui, après normalisation de la forme.
L : largeur du masque de l'objet.
H : hauteur du masque de l'objet.
l : longueur de la frontière l'objet.
Les points faibles apparaissent en grisés.

Nova de Lisboa)[RAUBER94]. Cette extension rend possible la caractérisation de contours ouverts. Au sein du projet MARS (*Multimedia Analysis Retrieval System*[6]), une version modifiée des descripteurs de FOURIER est aussi proposée[RUI96A]. Elle est plus compacte et rapide à calculer.

6.3.5 Comparaison des attributs

6.3.5.1 Caractéristiques globales

Le tableau [6.30] présente les différentes caractéristiques des descripteurs présentés. Il apparaît très rapidement que vu les critères que nous nous sommes fixés (invariances et utilisation pour l'indexation) très peu d'entre eux sont adaptés. Il faut aussi savoir que nous avons rejeté dès le départ certaines techniques très coûteuses au niveau de la comparaison des descriptions : axes médians, représentation *CSS*... Ces approches sont intéressantes dans le cadre de l'indexation de forme unique ; mais dans notre cas, la comparaison entre attributs pourra être effectuée entre toutes les régions des images ce qui engendrera une somme de calculs importante.

6.3.5.2 Performances

Les résultats présentés ici ont été obtenus sur la base de formes SIID[SEBAST01] (*Shape Indexing of Image Databases*) (cf. section B.5). Ils fournissent les courbes précision/rappel moyenne

[6]La page Web du projet est http://www-db.ics.uci.edu/pages/research/mars/index.shtml.

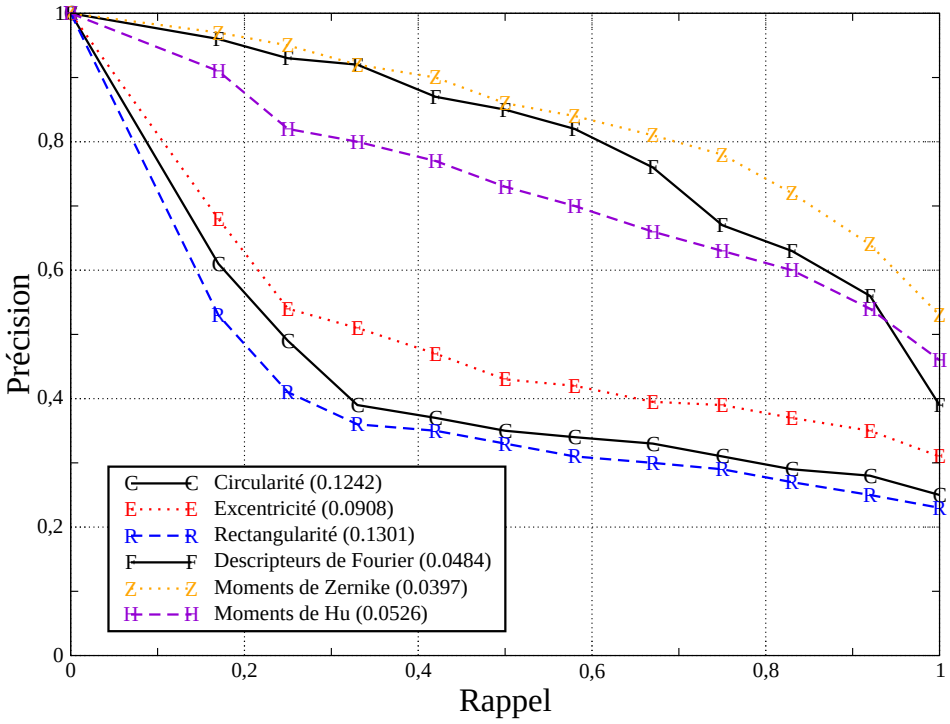


FIG. 6.31 – Performance des différents descripteurs de formes mis en place

calculées sur l'ensemble des 216 formes de la base. Au sein des légendes les valeurs de $\widetilde{Rang}$ sont indiquées entre parenthèses. Des expériences préliminaires nous ont amené à échantillonner les courbes en 64 points pour les descripteurs de FOURIER et à prendre les 15 premiers ordres des moments de ZERNIKE. D'après ces premiers tests, ces choix sont optimaux dans notre cas du point de vue du rapport précision/temps de calcul.

En terme de complexité et de stockage les statistiques simples (circularité, rectangularité ou l'excentricité) paraissent plus intéressantes. Utilisées seules (cf. figure [6.31]) leurs performances ne sont pas satisfaisantes du fait de leur irréversibilité et de leur simplicité. Par contre combinées (cf. figure [6.32]), elles peuvent fournir un descripteur de forme rapide et efficace. Les descripteurs plus évolués (descripteurs de FOURIER, moments de HU et de ZERNIKE) fournissent à eux seuls de bonnes statistiques.

Une fois normalisées les distributions des différentes statistiques, la combinaison des descripteurs évolués améliore la représentation. Il apparaît d'ailleurs que la combinaison d'une approche contour et d'une autre basée sur la région apporte de meilleurs résultats. Cela paraît logique de par la complémentarité des informations utilisées. Ainsi, l'association des moments de ZERNIKE est beaucoup plus intéressante avec des descripteurs de FOURIER qu'avec les moments de HU. C'est d'ailleurs une des conclusions de l'étude de PARK[PARK99A]. Pour la suite, l'association descripteurs de FOURIER/moments de ZERNIKE sera conservée étant donné la fiabilité moindre des moments de HU.

Il serait ainsi possible de combiner autant de descripteurs que voulu pour essayer d'améliorer la performance du système. Mais au delà du problème de complexité croissante que cela engendrerait, il serait alors très difficile d'obtenir un descripteur général utilisable au sein de n'importe quelle

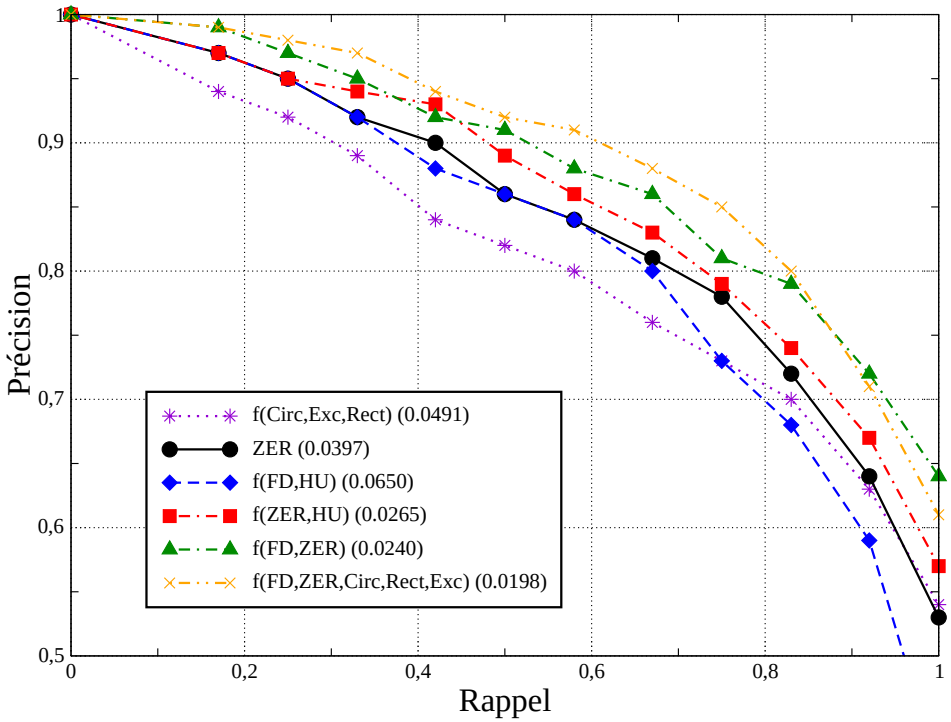


FIG. 6.32 – Combinaison des descripteurs de forme

base de donnée (et c'est ce que nous voulons !). De plus, comme nous l'avons vu l'association de descripteurs n'est réellement efficace que s'ils sont complémentaires. L'ensemble (excentricité, circularité, rectangularité) décrit grossièrement la forme et n'entre pas vraiment dans un schéma de description de la forme ou de la frontière. Il a donc été ajouté au descripteur précédent, mais pondéré faiblement, et apporte encore une amélioration de performance.

Le système choisi de description de la forme d'une région correspond donc à l'association des descripteurs FOURIER et des moments de ZERNIKE à laquelle est ajoutée la contribution pondérée de l'ensemble (excentricité, circularité, rectangularité).

6.3.5.3 Résistance à l'occultation

Pour tester les différents descripteurs au problème de l'occultation partielle des formes, nous avons simulé le passage d'un objet devant les différentes formes de la base SIID (cf. figure [6.33]). Pour chaque niveau d'occultation, nous avons comme précédemment calculé les courbes précision/rappel moyenne sur l'ensemble des images de la base (ici composée de 432 images : les 216 originales et les 216 tronquées).

La première appréciation qui peut être faite est que les différents descripteurs réagissent de manière satisfaisante si l'occultation n'est pas trop importante (cf. figure [6.34]). Par contre, à partir de 33% environ, une fracture franche apparaît pour un rappel de 50%. Elle correspond au fait que chaque catégorie est composée des versions originale et tronquée des formes. Elle symbolise donc la sensibilité à l'occultation. Les moments de ZERNIKE semblent ainsi beaucoup plus sensibles à l'occultation que les descripteurs de FOURIER ou les statistiques simples. La précision décroît

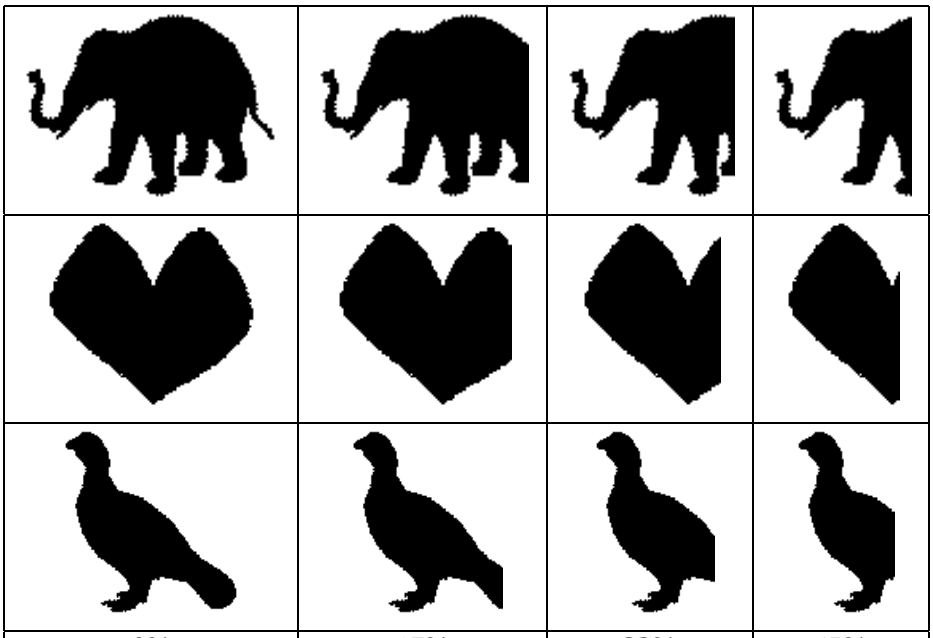


FIG. 6.33 – occultation progressive de différentes formes

fortement même pour des occultations faibles et la fracture à 50% est nette. Cette sensibilité est due au fait que ces moments sont calculés sur le disque unité au sein duquel la forme étudiée est centrée. Le fait de tronquer les formes les décalent progressivement au sein du disque ce qui engendre des modifications importantes des moments calculés. Cet effet est particulièrement visible sur la figure [6.35] qui présente une comparaison des descripteurs pour des occultations de 33%. Ce graphique montre d'ailleurs l'intérêt de combiner des descripteurs complémentaires ; les résultats du descripteurs choisi sont ainsi meilleurs que toutes les autres méthodes.

L'étude de la résistance globale du descripteur retenu est satisfaisante (cf. figure [6.36]). Jusqu'à 33% d'occultation, la précision de la recherche est satisfaisante.

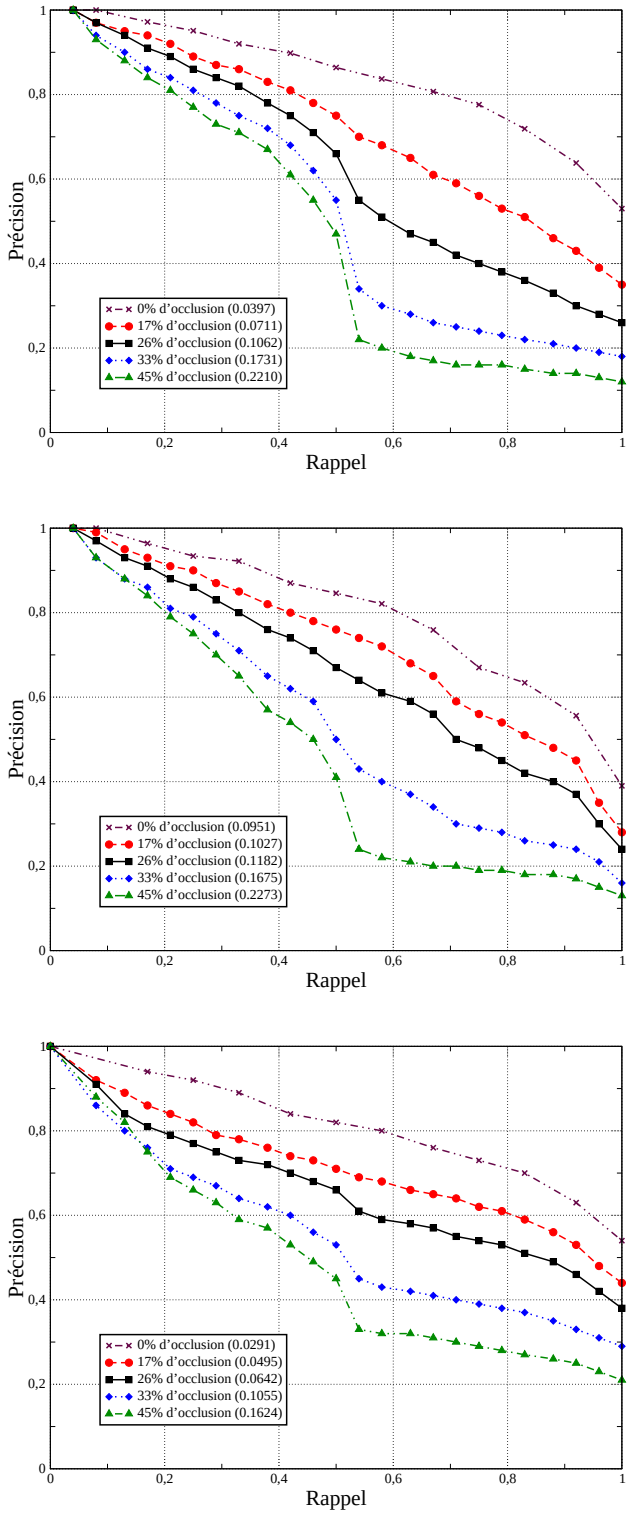


FIG. 6.34 – Résistance des trois descripteurs choisis aux occultations
En haut : moments de Zernike ;
Au milieu : descripteurs de Fourier ;
En bas : statistiques simples.

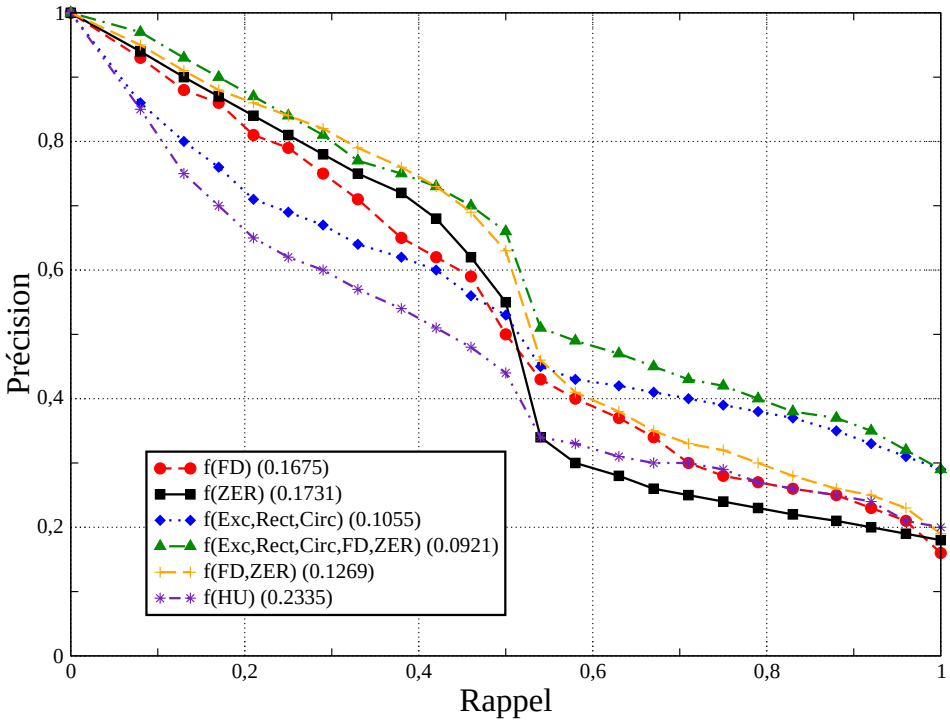


FIG. 6.35 – Comparaison des descripteurs pour une occultation de 33%

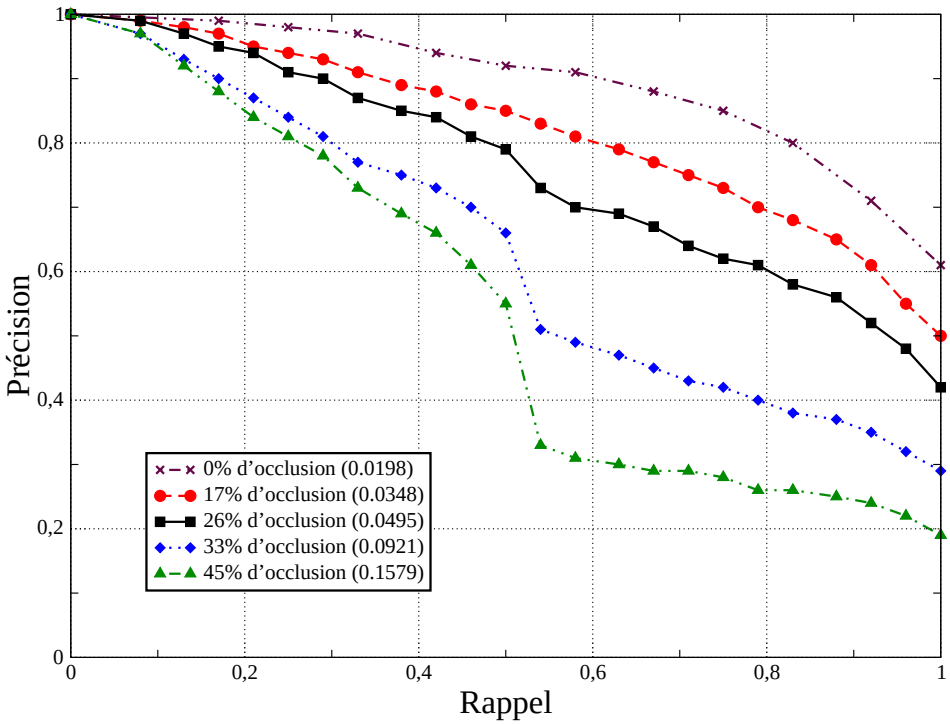


FIG. 6.36 – Résistance du descripteur de forme retenu aux occultations

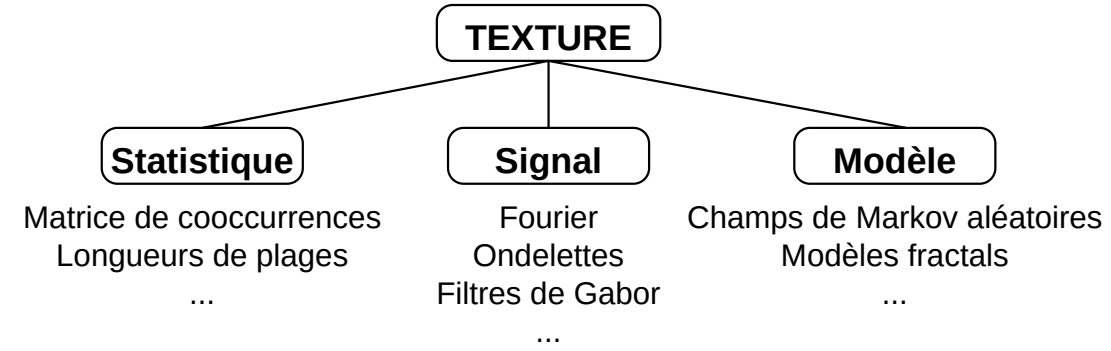


FIG. 6.37 – Organigramme des descripteurs de texture

6.4 Descripteurs de texture

Notre étude de la représentation de la texture au sein des images est beaucoup moins poussée que celles sur la couleur ou la forme. En effet, l'usure des différentes peintures engendre du bruit au sein des images de fresques médiévales. C'est alors ce bruit qui génère des textures au sein des images. L'aspect texture est donc peu pertinent pour cette application. Par contre, dans d'autres applications et en particulier dans un système de recherche d'images général, cette caractéristique devra être prise en considération de manière plus importante.

La plupart des attributs textures ont été développés pour caractériser des images en niveaux de gris. Nous les présenterons d'abord dans ce cadre puis verrons comment ils peuvent être étendus à la couleur. Nous les avons divisés en trois catégories (cf. figure [6.37]) :

- les méthodes statistiques caractérisant la distribution spatiale des niveaux de gris ;
- celles décrivant la texture par un modèle ;
- les techniques issues du traitement des signaux travaillant dans le domaine fréquentiel.

6.4.1 Méthodes statistiques

6.4.1.1 Matrice de cooccurrences

Comme nous l'avons déjà vu à la section 3.3.3, la fonction de cooccurrences au sein d'une image I est définie telle que $MC_I : \zeta_I^2 \rightarrow \mathbb{R}$ où ζ_I est le support colorimétrique de I. À un couple de couleurs (c_1, c_2) est associée la probabilité qu'elles soient voisines. Il faut donc définir une fonction de voisinage : $V : \Omega^2 \rightarrow \mathbb{B}$ qui définit si deux pixels de l'image $(p_1, p_2) \in \Omega^2$ sont voisins suivant le critère choisi : 4-connexité, 8-connexité ou autre. Ainsi, $\forall (c_1, c_2) \in \zeta_I^2$:

$$MC_I(c_1, c_2) = \frac{\text{Card}(\{(p_1, p_2) \ / \ (p_1, p_2) \in \Omega^2,\ I(p_1) = c_1,\ I(p_2) = c_2,\ V(p_1, p_2) = vrai\})}{\text{Card}(\{(p_1, p_2) \ / \ (p_1, p_2) \in \Omega^2,\ V(p_1, p_2) = vrai\}}$$

La matrice de cooccurrences est égale à la matrice carrée de taille $nbc_I = \text{Card}(\zeta_I)$:

$$[MC_I] = [MC_I(c_1, c_2)]$$

En 4-connexité par exemple, l'image suivante contenant 3 niveaux de gris différents a pour matrice de cooccurrences :

$$I = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 2 & 2 \\ 0 & 0 & 2 & 2 \end{bmatrix} \Bigg\| MC_I = \begin{bmatrix} \frac{1}{3} & \frac{1}{13} & \frac{1}{12} \\ \frac{1}{12} & \frac{1}{6} & 0 \\ \frac{1}{6} & 0 & \frac{1}{12} \end{bmatrix}$$

À partir de la matrice de cooccurrences en 8-connexité d'images en niveaux de gris, HARALICK [HARALI73] définit 14 mesures caractérisant une texture telle que le contraste, la corrélation, la variance... Voici la définition de quelques-unes d'entre elles pour une image I comportant m couleurs :

- Énergie $= \sum_{s=0}^{m-1} \sum_{t=0}^{m-1} MC_I(s,t)^2$;
- Entropie $= \sum_{s=0}^{m-1} \sum_{t=0}^{m-1} MC_I(s,t) \log MC_I(s,t)$;
- Corrélation $= \sum_{s=0}^{m-1} \sum_{t=0}^{m-1} \frac{(s-\mu_s)(t-\mu_t)MC_I(s,t)}{\sigma_s \sigma_t}$

  avec $\mu_s = \sum_{s=0}^{m-1} s \sum_{t=0}^{m-1} MC_I(s,t)$, $\mu_t = \sum_{t=0}^{m-1} t \sum_{s=0}^{m-1} MC_I(s,t)$,

  $\sigma_s = \sum_{s=0}^{m-1} (s-\mu_s)^2 \sum_{t=0}^{m-1} MC_I(s,t)$ et $\sigma_t = \sum_{t=0}^{m-1} (t-\mu_t)^2 \sum_{s=0}^{m-1} MC_I(s,t)$;
- Contraste $= \sum_{s=0}^{m-1} \sum_{t=0}^{m-1} (s-t)^2 MC_I(s,t)$;
- Homogénéité ou Moment des Différences Inverses $= \sum_{s=0}^{m-1} \sum_{t=0}^{m-1} \frac{MC_I(s,t)}{1+(s-t)^2}$;

GOTLIEB montre expérimentalement que le triplet contraste, entropie et moment des différences inverses est le plus intéressant pour la mise en place d'un descripteur général de texture [GOTLIE90]. Malheureusement, le calcul de tels attributs est assez complexe. Néanmoins, une quantification de l'image permet de réduire les temps de calcul.

6.4.1.2 Matrice de longueurs de plages

La matrice de cooccurrences prend en compte le voisinage des pixels deux à deux. La matrice de longueurs de plages quant à elle considère les plages de pixels consécutifs d'un même niveau de gris[GALLOW75]. Une plage est un ensemble de pixels adjacents de même niveau de gris dans une direction donnée. L'ensemble des informations des plages est généralement stocké dans une matrice dite *de Longueurs De Plages* (MLDP). Pour une direction donnée, elle est de taille $N \times L$ où N est le nombre de niveaux de gris de l'image et L la taille maximale de l'image dans la direction étudiée. Pour chaque couleur, le nombre de plages de différentes longueurs est alors stocké. Une texture « fine » sera donc caractérisée par une MLDP contenant peu de longues plages.

Une quantification préalable est généralement effectuée avant le calcul de cette matrice. Elle permet de réduire la dynamique de l'image sans détériorer la texture.

Pour la direction $\theta = 0°$, l'image suivante contenant 3 niveaux de gris différents a pour matrice de longueurs de plages :

$$I = \begin{bmatrix} 0 & 1 & 2 & 3 \\ 0 & 2 & 3 & 3 \\ 2 & 1 & 1 & 1 \\ 3 & 0 & 3 & 0 \end{bmatrix} \quad MLDP_I = \begin{array}{c|cccc} & 1 & 2 & 3 & 4 \\ \hline 0 & 4 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 2 & 3 & 0 & 0 & 0 \\ 3 & 3 & 1 & 0 & 0 \end{array}$$

À partir de cette définition, divers paramètres peuvent être calculés[CONNER80, CHU90]. En voici quelques exemples pour une direction donnée avec N le nombre de niveaux de gris de l'image et L la taille maximale de l'image dans la direction étudiée :

- Nombre de plages $= NP = \sum_{i=1}^{N}\sum_{j=1}^{L} MLDP(i,j)$;
- Poids des plages courtes $= \frac{1}{NP}\sum_{i=1}^{N}\sum_{j=1}^{L} \frac{MLDP(i,j)}{j^2}$;
- Poids des plages longues $= \frac{1}{NP}\sum_{i=1}^{N}\sum_{j=1}^{L} j^2 MLDP(i,j)$;
- Distribution des niveaux de gris $= \frac{1}{NP}\sum_{i=1}^{N}\left[\sum_{j=1}^{L} j^2 MLDP(i,j)\right]^2$ qui mesure l'uniformité de la distribution des plages ;
- Distribution des longueurs de plages $= \frac{1}{NP}\sum_{j=1}^{L}\left[\sum_{i=1}^{N} j^2 MLDP(i,j)\right]^2$;
- ...

Suivant l'application visée, l'utilisation de telle ou telle statistique est pertinente. Une combinaison de celles-ci permet d'obtenir un descripteur efficace.

6.4.2 Méthodes basées sur un modèle

Ces méthodes d'analyse de texture sont basées sur la construction d'un modèle pouvant être utilisé pour synthétiser des textures mais aussi pour caractériser celle d'une image.

6.4.2.1 Champs de MARKOV aléatoires

Les champs de MARKOV aléatoires sont très populaires dans le cadre de la modélisation d'images. Ils sont capables de saisir l'information locale (spatiale) au sein d'une image. Ces modèles considèrent que l'intensité des pixels n'est dépendante que de celles de ses voisins. Ils ont

été utilisés dans de nombreux domaines : synthèse et classification de texture, segmentation, restauration et compression. CHELLAPPA et CHATTERJEE proposent ainsi un champ de MARKOV aléatoire gaussien pour caractériser une texture[CHELLA85]. Les attributs de texture sont alors les paramètres définissant le modèle.

La *Wold decomposition* est également basée sur un champ aléatoire décomposé en trois composantes dites « harmoniques », « évanescentes » et « indéterminées » qui correspondent respectivement à la périodicité, le directionalité et au caractère aléatoire[FRANCO93]. Les textures périodiques ont une harmonique forte, celles qui sont fortement orientées ont une évanescence forte et les textures les moins structurées ont une indétermination importante. Dans le domaine spatial, les trois composantes orthogonales peuvent être obtenues par une estimation du maximum de vraisemblance qui implique un ajustement d'un modèle auto-régressif par minimisation d'une fonction de coût et résolution d'un ensemble d'équations linéaires. Dans le domaine fréquentiel, la *Wold decomposition* peut être obtenue par un seuillage global du spectre de FOURIER de l'image. Dans [LIU96], une méthode utilisant l'extraction des pics harmoniques et une modélisation par un modèle auto-régressif multi-résolutions simultané (*MRSAR* pour *Multi-Resolution Simultaneous Auto-Regressive model* en anglais) sans décomposition directe de l'image est présentée.

Le modèle auto-régressif simultané (*SAR* pour *simultaneous Auto-Regressive* en anglais) est un type particulier de champ de MARKOV aléatoire. Il a montré tout son intérêt dans la modélisation de textures. Comparé aux autres champs de MARKOV aléatoires, il utilise moins de paramètres. Le *MRSAR* est alors une extension multi-résolutions du *SAR* basée sur la construction de la pyramide gaussienne de l'image au sein de laquelle le modèle *SAR* est appliqué sur les différents niveaux[MAO92].

6.4.2.2 Dimension Fractale

L'utilisation de la dimension fractale (DF) pour la classification ou la segmentation de textures a été proposé initialement par PENTLAND[PENTLA84]. La propriété d'auto-similarité des fractales implique que cette description d'une image est indépendante de l'échelle. De nombreuses méthodes existent pour estimer la dimension fractale d'une image : généralisation des méthodes originales de MANDELBROT[PELEG84], utilisation de la transformée de FOURIER[PENTLA84] ou différentes techniques de *box-counting*[CHAUDH93]. Le principe d'auto-similarité peut s'énoncer ainsi : « si un ensemble borné A est composé sans recouvrement de N_r copies d'un ensemble similaire à A mais réduit d'un facteur r, alors A est auto-similaire ». À partir de cette définition, la dimension fractale est :

$$DF = \frac{\log(N_r)}{\log(r^{-1})}$$

La détermination de DF est très complexe à réaliser pratiquement. DF peut être approximée en estimant N_r pour différentes valeurs de r puis en déterminant la pente de $log(N_r)/log(r^{-1})$ par la méthode des moindres carrés (cf. figure [6.38]). La méthode proposée par CHAUDHURI peut être utilisée pour mettre en place ce système[CHAUDH93]. Il propose alors quatre attributs normalisés entre 0 et 1 :

- DF de l'image originale I ;
- DF d'une image des niveaux de gris supérieurs ;
- DF d'une image des niveaux de gris inférieurs ;
- une valeur multi-fractales de DF.

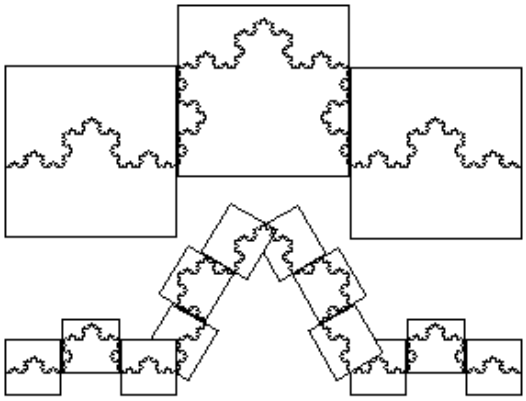


FIG. 6.38 – Placage de boîtes de tailles différentes sur une courbe

Les image de niveaux de gris inférieurs et supérieurs correspondent à l'image seuillée pour laquelle seuls les pixels de niveaux hauts ou bas sont conservés.

6.4.3 Méthodes du traitement du signal

Des recherches sur la psychologie de la vision ont mis en évidence que le cerveau humain effectue une analyse fréquentielle des images[CAMPBE68]. La texture rentre spécialement dans ce cadre de par ses propriétés de redondance.

6.4.3.1 Ondelettes

L'analyse fréquentielle est bien évidemment réalisée plus efficacement dans le domaine de FOURIER. Comme les résultats des études psychologiques le montrent, le système visuel humain analyse les images texturées en décomposant leurs composantes fréquentielles et d'orientation[CAMPBE68]. La transformée de FOURIER caractérise le contenu global du signal, alors que certaines applications requièrent une analyse localisée dans le domaine spatial. Pour cela, les ondelettes apportent une solution très intéressante en séparant les différentes sous-bandes spectrales. Les énergies des sous-bandes sont par la suite généralement utilisées pour caractériser une texture. Cette technique très rapide a été mise en place par SMITH et CHANG qui emploient la moyenne et la variance des différentes sous-bandes comme descripteurs[SMITH94].

6.4.3.2 Filtres de GABOR

La technique la plus souvent utilisée pour caractériser des textures au niveau fréquentiel est sans aucun doute les filtres de GABOR[DAUGMA88]. Cette méthode calcule les énergies de GABOR qui mesurent la similarité entre les voisinages de l'image et les masques de GABOR. Ces derniers sont en fait des oscillateurs harmoniques caractérisés par une gaussienne modulée par une sinusoïde. Elles peuvent être générées à partir de 4 paramètres : λ la longueur d'ondes, θ l'orientation, ϕ la

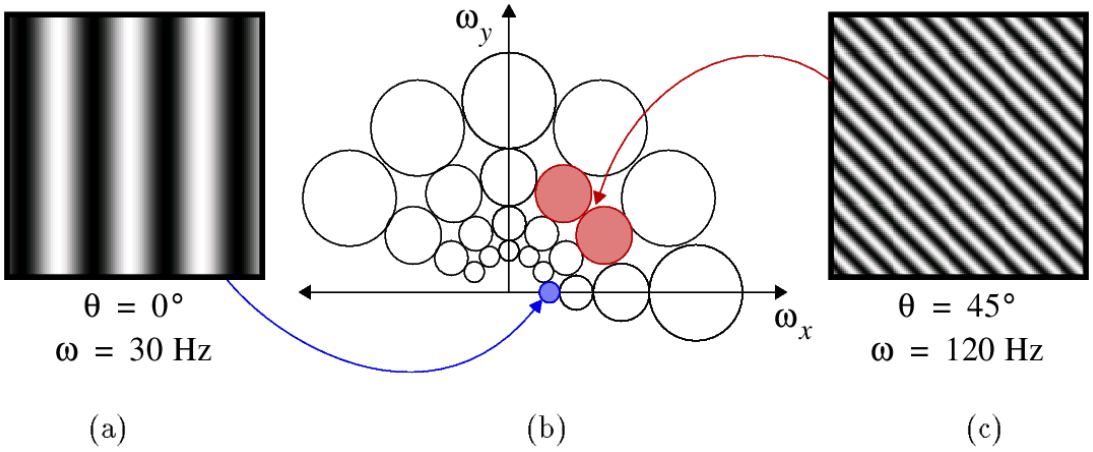


FIG. 6.39 – Répartition au sein du spectre des filtres de GABOR (tirée de [SMITH97])
Avec $\omega = \frac{2\pi}{\lambda}$, les différents cercles symbolisent la localisation fréquentielle des différents filtres.

phase de l'onde et σ l'écart-type de la gaussienne. Pour un masque de taille $l \times L$, la fonction de GABOR est alors :

$$G(x, y|\lambda, \theta, \phi) = e^{-\frac{(x - \frac{l-1}{2})^2 + (y - \frac{L-1}{2})^2}{2\sigma^2}} J(x, y|\lambda, \theta, \phi),$$

où J est l'oscillateur harmonique suivant défini pour $x \in [0 \dots l-1]$ et $y \in [0 \dots L-1]$:

$$J(x, y|\lambda, \theta, \phi) = \sin(\frac{2\pi}{\lambda}(x \cos\theta - y \sin\theta) + \phi)$$

Les attributs de texture sont généralement les énergies de la convolution de l'image avec le filtre de GABOR sommées sur les phases pour toutes les combinaisons de longueurs d'ondes et d'orientations :

$$S^2(\lambda, \theta) = \sum_{\phi} \left[\sum_{x,y} G(x, y|\lambda, \theta, \phi) I(x, y) \right]^2$$

Les différents filtres de GABOR étant peu étendus dans le domaine fréquentiel (cf. figure [6.39]), il est beaucoup plus rapide de calculer l'énergie résultante de la convolution dans ce domaine.

Le filtrage de GABOR est très intéressant dans le cadre de l'analyse de textures car il mêle deux objectifs opposés : la représentation de la texture et sa localisation. Ce filtrage apporte de très bons résultats en classification de textures et en segmentation[BOVIK90]. De plus, cette approche se justifie du point de vue du système de vision humaine. BECK montre ainsi que la discrimination des textures par l'homme est réalisée à partir d'informations spatio-fréquentielles[BECK87]. Toutefois, les fonctions de GABOR ne sont pas orthogonales et sont complexes à mettre en œuvre.

MANJUNATH et MA mettent en compétition les filtres de GABOR et les ondelettes dans le cadre de la classification de textures[MANJUN96]. Les filtres de GABOR apparaissent plus précis mais plus lents. Ils proposent donc une méthode permettant de rendre équivalents les temps de calculs des deux méthodes.

6.4.4 Extension à une image multi-composantes

Les méthodes présentées jusqu'ici pour caractériser une texture ont toutes été initialement proposées pour des images en niveaux de gris. Avec l'évolution des technologies et des applications étudiées, il était nécessaire de les étendre aux images multi-composantes et en particulier aux images couleur. Suivants les méthodes quatre voies différentes peuvent être employées.

La première qui paraît la plus naturelle est l'extension de la notion de distance entre niveaux de gris simple $|s-t|$ à une distance couleur. Celle-ci doit bien entendu être la plus perceptuellement acceptable. Cette technique peut ainsi être employée pour les matrices de cooccurrences [CHANG96]. Pour les matrices de longueurs de plages, l'extension est différente. FERNANDEZ propose de considérer une plage couleur comme une suite de pixels dont les différences perceptuelles de couleurs sont inférieures à un seuil[FERNAN01]. De nouveaux paramètres sont également présentés pour s'adapter au problème couleur.

La deuxième solution utilise la tridimensionalité de la couleur sans réellement en prendre en compte ses propriétés. Les trois plans couleurs peuvent être analysés indépendemment comme des images en niveaux de gris, c'est l'approche marginale. Cette approche est bien sûr fortement biaisée car elle n'utilise pas la notion de couleur en tant que telle ce qui engendre généralement des résultats peu pertinents.

JAIN et HEALEY proposent une troisième approche inspiré du système visuel humain[JAIN98]. Ils définissent les filtres de GABOR sur l'image de luminance et sur deux images symbolisant l'antagonisme des couleurs.

La dernière approche, sûrement la plus prometteuse, est l'extension de la transformée de FOURIER à la couleur. PALM propose la transformée de FOURIER chromatique complexe[PALM02]. À partir d'une représentation complexe de la couleur obtenue à partir des coordonnées dans l'espace HSV, PALM met en place des filtres de GABOR complexes et montre qu'ils fournissent une meilleure représentation que les filtres de GABOR marginaux. Il présente aussi la représentation de la couleur par les quaternions qui sont des nombres complexes à trois parties imaginaires :

$$q = q_0 + i_1 q_1 + i_2 q_2 + i_3 q_3$$

$$\text{avec } i_1^2 = i_2^2 = i_3^2 = i_1 i_2 i_3 = -1,\ i_x * i_{(x+1)\%3} = i_{(x+2)\%3} \text{ et } i_x * i_{(x-1)\%3} = i_{(x-2)\%3}$$

où $\%$ est l'opérateur modulo.

Une image couleur est alors interprétée comme une image de quaternions où la partie réelle est nulle et les trois parties imaginaires correspondent aux trois composantes couleur. La transformée de FOURIER quaternion introduite par SANGWINE peut ensuite être utilisée[SANGWI79]. SANGWINE a déjà appliqué cette méthode avec succès pour calculer l'auto-corrélation entre images couleur[SANGWI00]. Elle sera sûrement employée dans un futur proche pour de la classification de textures.

6.4.5 Comparaison des descripteurs

Comme nous l'avons exposé en introduction de cette section, la texture est un critère peu discriminant au sein des images de fresques médiévales. Nous n'avons donc pas réalisé d'étude comparative poussée sur les différents descripteurs mais nous nous sommes principalement basés sur

les différentes mises en compétition effectuées au sein du projet MEASTEX[7]. Nous sommes également allés au plus simple dans la mise en place d'un descripteur de textures. C'est ainsi la matrice de cooccurrences qui a été retenue pour sa simplicité même si elle n'apporte pas les meilleurs résultats comme nous le verrons par la suite. Ce choix est également motivé par les très bons résultats de cette méthode pour la classification de textures de peintures qui se rapprochent fortement des fresques médiévales. Enfin, son extension à la couleur est très simple contrairement aux autres techniques. Il suffit pour cela de remplacer la différence des niveaux de gris par une distance perceptuellement correcte (L_2 dans $L^*a^*b^*$ par exemple). L'utilisation des filtres de GABOR, des champs de MARKOV aléatoires gaussiens ou encore des ondelettes serait plutôt recommandée pour obtenir une meilleure précision dans un cadre général.

La mesure de performance utilisée pour l'étude réalisée par MEASTEX est un score calculé par :

$$\text{Score} = \sum_{c \in C} \frac{1}{N_c} \sum_{i \in I_c} \delta_{c_i, c'_i}$$

où I_c est l'ensemble de toutes les images de la catégorie c, C est l'ensemble de toutes les catégories, N_c le nombre d'images de la catégorie c, c_i la classe déterminée de l'image i par le processus de classification alors que c'_i est la classe réelle de cette même image. δ_{c_i, c'_i} vaut 1 si $c_i = c'_i$ et 0 dans le cas contraire. C'est donc le pourcentage de bonnes classifications réalisées grâce au descripteur étudié.

Les attributs comparés sont l'énergie de GABOR, les champs de MARKOV aléatoires gaussiens et les statistiques de la matrice de cooccurrences tous les trois définis pour des images en niveaux de gris. Les résultats obtenus sont alors les suivants :

Base d'images	Mat. Cooc.	GABOR	MARKOV
bomb	0.8406	0.846	0.9446
bombRot	0.6829	0.9248	0.9603
brodatz	0.9239	0.9451	0.9713
grass	0.9162	0.89	0.9483
material	0.965	0.9678	0.9797
VisTex	0.8523	0.9066	0.9355
lattice	0.6953	0.8919	0.7396
latticeRot	0.6643	1	0.9647
mortar	0.715	0.8758	0.7551
mortarRot	0.6055	0.9921	0.9626
mortarRotS	0.6248	1	0.9763

Ils montrent que la matrice de cooccurrences est globalement la moins bonne des descriptions même si ses performances sont tout à fait acceptables. Les champs de MARKOV aléatoires gaussiens sont plus performants sur les micro-textures alors que l'énergie de GABOR est plus précise avec des macro-textures.

En 1992, OHANIAN et DUBES ont mis en compétition ces trois descripteurs avec la dimension fractale[OHANIA92]. Leurs conclusions sont complètement contredites par celles de l'étude réalisée au sein de MEASTEX sur les mêmes bases d'images mais avec leurs implémentations propres

[7] Disponible à l'adresse http://www.cssip.uq.edu.au/staff/meastex/meastex.html.

et donc différentes des descripteurs. Sur des textures fractales, ils annoncent que la dimension fractale est la plus performante contrairement à MEASTEX au sein duquel les énergies de GABOR et les champs de MARKOV aléatoires gaussiens sont significativement meilleurs. Pour les textures peintes, leurs résultats montrent que la matrice de cooccurrences est la plus adaptée suivi de près par la dimension fractale alors que pour MEASTEX c'est toujours les deux autres statistiques qui sont légèrement supérieures. Il faut donc faire très attention aux divers résultats publiés. Une implémentation différente peut engendrer des conclusions complètement différentes. Ainsi, OHANIAN et DUBES obtiennent globalement le classement suivant : matrice de cooccurrences, dimension fractale, champs de MARKOV aléatoires gaussiens et enfin énergies de GABOR. Les tests réalisés au sein de MEASTEX montrent quant à eux le classement suivant : champs de MARKOV aléatoires gaussiens, énergies de GABOR, matrice de cooccurrences et enfin dimension fractale.

Une autre comparaison est présentée par PORTER en 1997[PORTER97]. Il trouve alors que les ondelettes sont les plus précises et qu'elles sont les plus résistantes au bruit tout en étant les moins complexes à calculer. D'autres expérimentations mises en place par VALKEALATHI concluent qu'un histogramme multidimensionnel de cooccurrences est encore plus précis que les ondelettes[VALKEA98].

6.5 Mélange des descripteurs

Une fois défini les trois types de descripteurs : couleur, forme et texture, il est nécessaire de les fusionner en fonction de l'application visée. Une seule caractéristique peut être utilisée dans le cadre d'applications spécifiques. Pour un système de tri de pièces manufacturées par exemple, la forme seule sera généralement utilisée. Mais si la différence entre les pièces est uniquement la couleur alors c'est uniquement cet attribut qui sera employé. Pour d'autres applications, une des trois caractéristiques pourra ne pas être pertinente ; elle devra donc être écartée du processus de reconnaissance. Le choix du système de mélange des statistiques est donc fortement dépendant de l'application.

En considérant les trois statistiques au sein d'un vecteur d'attributs, la recherche de similarité entre régions peut être réalisée dans un espace à 1, 2 ou 3 dimensions en fonction du nombre de caractéristiques prises en compte. En prenant soin de rendre similaire par normalisation les distributions des divers éléments du vecteur d'attributs, la distance entre régions peut alors s'exprimer de différentes manières :

$$
\begin{aligned}
D_1(R_1,R_2) &= \omega_{coul} d_{coul}(R_1,R_2) + \omega_{forme} d_{forme}(R_1,R_2) + \omega_{text} d_{text}(R_1,R_2) \\
D_2(R_1,R_2) &= \sqrt{\omega_{coul} d_{coul}(R_1,R_2)^2 + \omega_{forme} d_{forme}(R_1,R_2)^2 + \omega_{text} d_{text}(R_1,R_2)^2} \\
D_\infty(R_1,R_2) &= \max\left(\omega_{coul} d_{coul}(R_1,R_2), \omega_{forme} d_{forme}(R_1,R_2), \omega_{text} d_{text}(R_1,R_2)^2\right)
\end{aligned}
$$

où les ω_* sont les poids alloués à chaque composante du vecteur d'attributs et les d_* sont les diverses distances ou mesures de similarité retenues pour chaque type de descripteurs.

Cette méthode est généralement celle utilisée dans les systèmes d'indexation par le contenu. L'utilisateur définit les poids alloués aux différentes caractéristiques disponibles. Les recherches d'images ressemblantes sont alors réalisées en utilisant la pondération définie. Cela est très intéressant dans une phase de tests des différentes statistiques mais est souvent trop complexe ensuite. En effet, les utilisateurs de tels systèmes n'ont généralement pas les connaissances suffisantes pour réaliser les choix pertinents. De plus, les pondérations optimales varient suivant la base d'image traitée mais n'ont pas la plupart du temps à être modifiées au sein d'une même base. Il est donc important de trouver les divers poids optimaux. Dans cette étude, cette détermination sera réalisée de manière empirique mais il faut noter que des techniques de détermination automatique existent. LARABI propose en particulier un système de coopération d'attributs[LARABI02]. À partir de diverses expériences réalisées par des experts, les poids optimaux sont calculés en prenant en compte :

- les divers résultats obtenus pour l'ensemble des experts ;
- la confiance associée aux différents experts en fonction de la pertinence de leurs résultats.

Un autre choix possible est la mise en place d'un arbre de décision. La mesure de similarité est alors déterminée en prenant en compte certaines conditions. Par exemple, si la mesure couleur est considérée comme prédominante alors elle pourra correspondre à une première estimation de ressemblance. Si la ressemblance est forte alors cette mesure sera conservée sinon la texture et/ou la forme pourront être utilisées pour voir s'il n'existe tout de même pas une similarité entre les régions (cf. figure [6.40]).

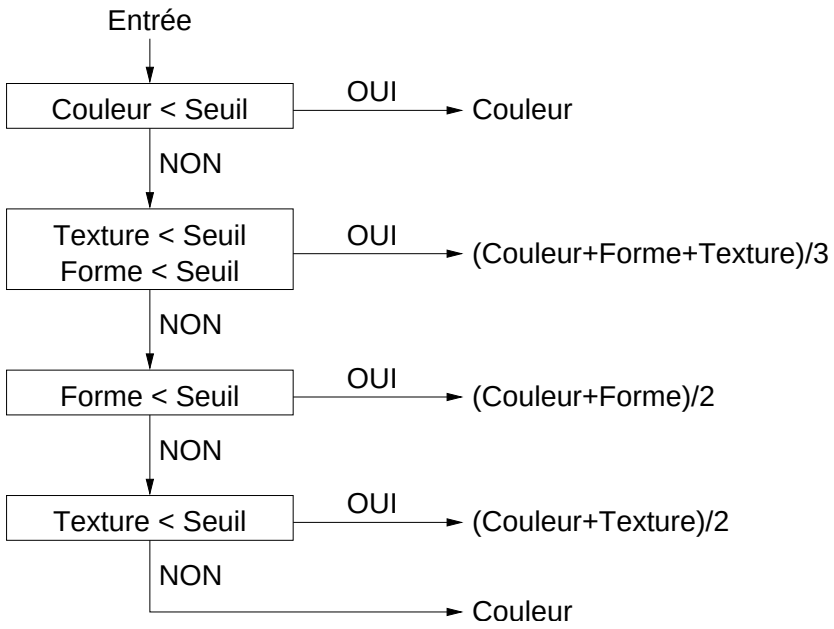


FIG. 6.40 – Exemple d'arbre de décision utilisable pour mélanger les descripteurs
Dans cet arbre de décision la couleur est considérée comme facteur prédominant. Si la distance couleur est inférieure à un seuil alors elle fournit à elle seule la mesure de similarité entre régions. Dans le cas contraire, le forme et la texture sont étudiées pour voir s'il existe tout de même une ressemblance. Si une seule ou les deux statistiques secondaires sont inférieures à un seuil alors elles sont prises en compte pour réajuster la mesure de similarité dans laquelle la couleur intervient toujours.

6.6 Conclusion

Ce chapitre a donc présenté les trois types de descripteurs que nous avons mis en place. Pour chacun d'eux un état de l'art des principales méthodes existantes a été réalisé pour aboutir à un ensemble de tests qui nous ont permis de définir l'attribut le mieux adapté à notre structure de représentation.

Pour la couleur, nous avons proposé deux nouvelles techniques de comparaison d'histogrammes utilisables même si les palettes couleurs ne sont pas identiques. L'intersection d'histogrammes mis en place est très efficace de manière générale et fournit une très bonne mesure de similarité ainsi qu'une complexité faible. La projection d'histogrammes quant à elle permet de réduire fortement la complexité de l'*Earth Mover's Distance* tout en fournissant des résultats équivalents. Nous avons également montré que l'extension des distances usuelles par application d'un noyau gaussien fournit une augmentation significative des résultats d'indexation quand la taille du noyau gaussien est adaptée à l'espace couleur et à la mesure de similarité utilisés.

Pour la forme, un ensemble de tests a été réalisé sur six descripteurs différents. Finalement, nous avons combiné des descripteurs fournissant des informations complémentaires. Les moments de ZERNIKE décrivent ainsi la région dans son ensemble alors que les descripteurs de FOURIER apportent une caractérisation de son contour. À cela, nous avons ajouté des descripteurs géomé-

triques simples (circularité, rectangularité, excentricité) qui fournissent une troisième indication générale sur la forme.

La texture n'étant pas pertinente dans l'application qui nous intéresse, son étude a été plus rapide et les tests réalisés quasiment inexistants. C'est tout de même les diverses statistiques calculables à partir de la matrice de cooccurrences qui ont été retenues pour cet attribut. Ce choix est motivé par la simplicité de cette statistique et par les résultats de l'étude de MEASTEX. Notons ici, que les énergies de GABOR, les ondelettes ou encore les champs de MARKOV aléatoires gaussiens fournissent généralement une caractérisation plus fine de la texture.

Notre structure de représentation est maintenant définie entièrement : structure et description des différents éléments. Nous allons donc montrer dans les deux chapitres suivants son intérêt dans le cadre de l'indexation d'images et surtout dans celui de la recherche d'objets.

CHAPITRE 7

COMPARAISON DE GRAPHES PYRAMIDAUX

Sommaire

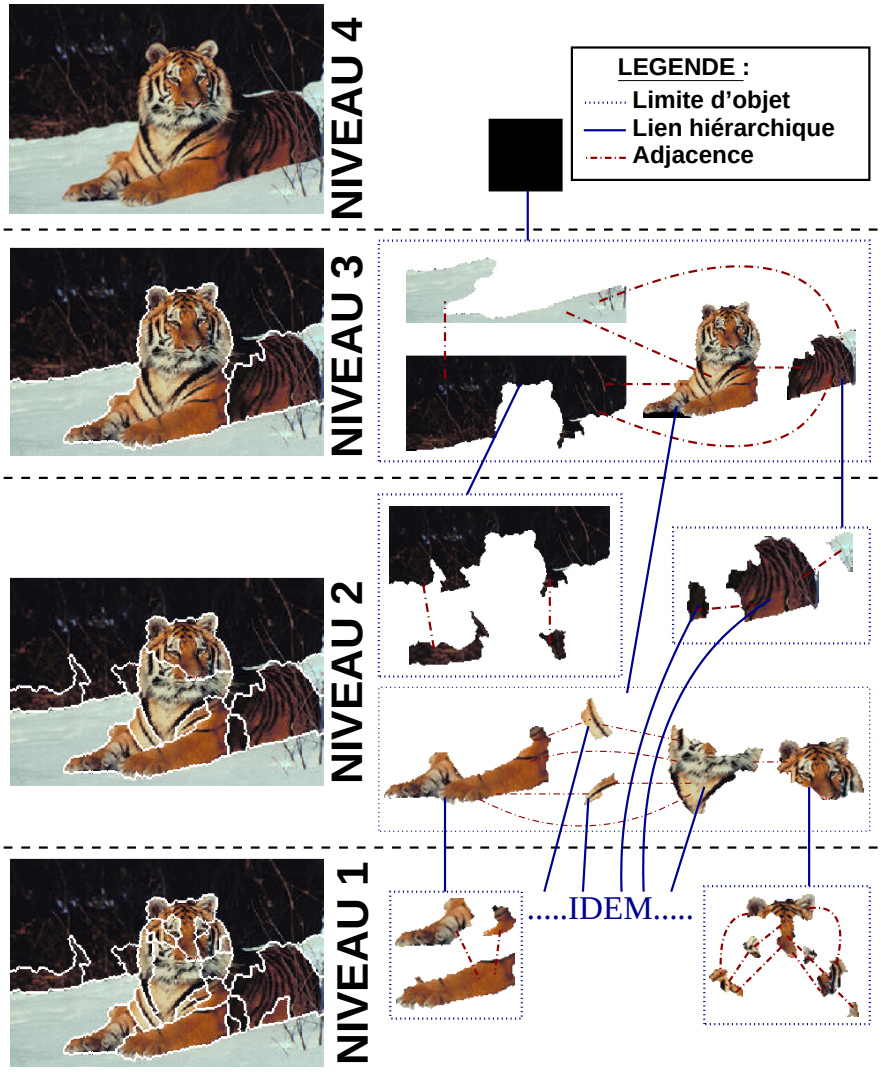


FIG. 7.1 – Graphe pyramidal obtenu à partir d'une segmentation réelle

7.1 Introduction

Au chapitre 4, nous avons présenté une nouvelle structure de représentation appelée *graphe pyramidal*. Celle-ci permet de décrire une image par une succession de décompositions. À partir d'un algorithme de segmentation fournissant plusieurs partitionnements de l'image (cf. chapitre 5), un graphe pyramidal est construit au sein duquel chaque niveau correspond au graphe d'adjacence d'une carte de segmentation (cf. figure [7.1]). Les différentes échelles sont alors liées pour coder la composition des différents objets. Chaque région de cette pyramide est caractérisée par divers attributs de couleur, de forme et de texture définis au chapitre 6.

Ce chapitre a pour but de montrer l'intérêt d'un telle structure de représentation dans le cadre de l'indexation d'images. Le schéma d'indexation proposé est celui présenté sur la figure [7.2]. Une fois les graphes pyramidaux générés la comparaison des images est effectuée par trois modules dépendants gérant chacun une spécificité du graphe pyramidal :

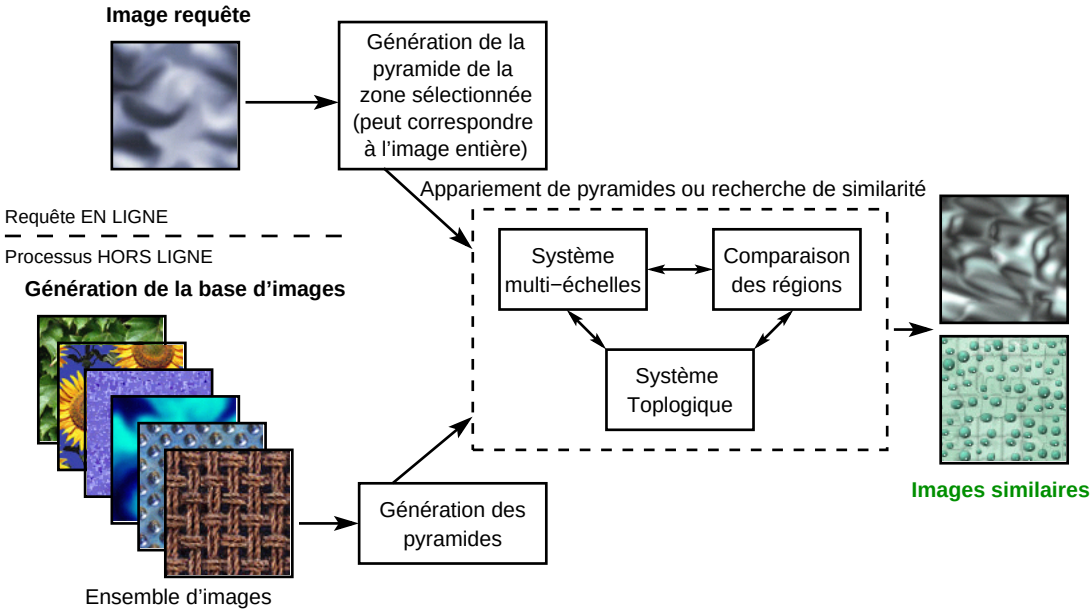


FIG. 7.2 – Schéma de principe du système d'indexation mis en place

– la comparaison des régions est réalisée par une mesure de similarité entre les attributs ;
– le module topologique définit l'utilisation de l'information spatiale des différents niveaux et permet la comparaison de deux ensembles de régions arrangées spatialement ou non ;
– le module multi-échelles gère la hiérarchie de la représentation et rend possible l'utilisation combinée de tous les niveaux de la pyramide.

Ce chapitre présente les différentes techniques mises en place au sein des modules topologiques et multi-échelles et met en évidence leurs intérêts dans le cadre de la recherche globale d'images. Par quelques exemples concrets, nous présentons ensuite les apports d'un tel système pour la recherche d'objets. Enfin, nous comparons cette méthode avec quelques systèmes existants pour lesquels nous avons pu obtenir les résultats.

7.2 Intérêt de la structure spatiale

Dans un premier temps, nous allons présenter en quoi la structure spatiale apporte un gain d'information intéressant pour l'indexation d'images. Nous utilisons simplement la décomposition de l'image en plusieurs régions sans prendre en compte leur arrangement spatial. Le formalisme des graphes nous permet ensuite d'utiliser cette propriété. Ce dernier aspect est surtout intéressant dans le cadre de la recherche d'objets car ces algorithmes sont généralement trop complexes pour être utilisés sur les niveaux entiers.

7.2.1 Utilisation de la décomposition

7.2.1.1 Algorithme développé

Pour montrer que les statistiques localisées portent plus d'informations que le même descripteur global, nous avons tout d'abord travaillé sur la notion de décomposition des images en plusieurs régions. Les différentes échelles du graphe pyramidal sont considérées comme des ensembles simples de régions. Par conséquent, l'arrangement spatial n'est pas pris en compte. La méthode la plus simple est alors de considérer la somme des associations minimales entre les régions pour caractériser la distance entre images.

Notons R et T les niveaux à comparer contenant respectivement $|R|$ et $|T|$ régions. Les régions de R sont notées : $R = \{R_i / i \in [1 \dots |R|]\}$. De la même manière, T se décompose en $T = \{T_j / j \in [1 \dots |T|]\}$. Nous pouvons définir la matrice de similarité entre les différents ensembles :

$$[D_{ij}] = [d(R_i, T_j)], \text{ avec } i \in [1 \dots |R|] \text{ et } j \in [1 \dots |T|]$$

qui contient toutes les distances entre les régions des deux ensembles. La distance choisie d est quelconque et peut prendre en compte les caractéristiques souhaitées (couleur et/ou texture et/ou forme...).

À partir de ces données nous voulons déterminer la similarité $S(R,T)$ entre les deux niveaux. Dans le cadre d'une comparaison globale d'images, celle-ci devant être inversible, cette mesure doit être symétrique ($S(R,T)$=$S(T,R)$). Mais, si R représente un objet qui est recherché dans T alors cette propriété n'est pas nécessaire. Dans ce dernier cas, ne voulant pas prendre en compte l'information sur les relations entre régions, nous calculons le poids minimum de projection de l'objet dans l'image par :

$$p_{NivSim}(R,T) = \frac{1}{|R|} \sum_{i=1}^{|R|} \left[\min_{j \in [1 \dots |T|]} D_{ij} \right]$$

La définition d'une telle mesure pose deux problèmes majeurs. Le premier est le choix d'une pondération des régions. En effet, nous voulons prendre en considération toutes les régions mais faut-il prendre en compte leurs tailles relatives ? Les diverses expérimentations que nous avons pu mener montrent qu'une pondération par la taille des régions n'est pas intéressante pour de telles recherches. Une telle compensation provoque en réalité la diminution de l'intérêt porté aux petites régions formées par la segmentation. Or, elles peuvent être visuellement très importantes. Le nez rouge d'un clown peut sembler petit en regard des autres éléments de l'image mais il apporte une information discriminante pour différencier un personnage standard d'un clown. Au vu des résultats expérimentaux une telle pondération n'a pas été incluse. De cette manière, toutes les régions formées par la segmentation sont prises en compte de manière équivalente et c'est là le réel apport d'une approche spatiale même si cela implique d'avoir une segmentation de bonne qualité.
Le second problème est lié à la non prise en compte de l'arrangement spatial. En effet, les régions de R se projettent indépendemment dans T. Deux régions de R peuvent ainsi être associées à la même région de T ; des régions visuellement importantes de T peuvent ainsi ne pas être prises en compte. De plus, le fait de ne pas utiliser les relations spatiales fait qu'un objet peut se retrouver projeté en plusieurs régions non adjacentes. Cette mesure semble donc fortement biaisée dans

le cadre de la recherche d'objets. Pour des recherches globales, les problèmes engendrés par les projections désordonnées peuvent être atténués par symétrisation de la mesure :

$$S_{NivSim}(R,T) = \frac{p_{NivSim}(R,T) + p_{NivSim}(T,R)}{2}$$

Comme nous le verrons dans la section suivante, cette approche fournit de bons résultats. Nous avons également essayé de faire évoluer cette évaluation de similarité pour éviter les projections multiples mais les résultats deviennent alors aberrants car, dans ce cas, les erreurs de segmentation ne peuvent pas être gommées. Effectivement, certaines régions divisées par erreur peuvent ne pas trouver de correspondants dans l'autre niveau ce qui engendre l'ajout d'une distance importante non pertinente à p_{NivSim}.

Bien que biaisée, nous verrons que cette mesure de similarité spatiale apporte un gain d'information important dans le cadre des recherches d'images et permet d'améliorer sensiblement la précision par rapport à une requête globale.

7.2.1.2 Résultats

Pour montrer expérimentalement l'amélioration apportée par l'utilisation de la décomposition de l'image en régions, nous avons travaillé sur diverses bases d'images (Columbia, SOIL-47...). Tous les résultats obtenus vont dans le même sens, nous ne présenterons par conséquent que ceux liés à la base Columbia et à celle de l'université de Washington. Pour Columbia, les courbes précision/rappel ont été calculées en utilisant une base réduite à 12 angles de vues pour chaque objet (0°, 25°, 45°, 65°, 85°, 125°, 165°, 185°, 245°, 265°, 325° et 345°). Pour rester dans un cadre général, c'est la distance *EMD* standard en $L^*a^*b^*$ qui a permis de comparer les régions après une quantification préalable des images en 16 couleurs. Cette quantification forte associée à l'*EMD* se justifie par les résultats présentés à la section 6.2.8.

Les figures [7.3] et [7.4] montrent ainsi qu'une décomposition de l'image en régions est intéressante dans le cadre de l'indexation d'images. Sur ces deux exemples, un gain de précision d'environ 10% est atteint pour le niveau de segmentation le plus fin. Cela paraît naturel car au sein de l'image entière les statistiques des composants de l'image sont mélangées. Elles sont alors toutes fusionnées dans le descripteur global ce qui les rend peu discriminantes. Le fait de les extraire et de les considérer toutes avec le même niveau d'importance permet de décrire plus finement l'image. Il faut tout de même faire attention a ne pas trop segmenter finement les images. En effet, comme nous pouvons le voir sur la figure [7.4], l'utilisation de la décomposition la plus fine fournit de moins bons résultats que le premier niveau de partitionnement. Cela est principalement dû à la présence dans les images de cette base de zones uniformes étendues qui sont généralement sursegmentées sur le niveau le plus fin engrendant ensuite plus de biais dans la comparaison.

La figure [7.5] présente également quelques requêtes significatives obtenues sur la base Columbia. Elles montrent bien qu'au sein des requêtes globales les images retrouvées sont globalement similaires par leur couleur mais le fait de ne pas prendre en compte les régions des images ne permet pas d'utiliser les spécificités des objets tels que les motifs des tasses ou l'arrière rouge de la voiture qui apparaît dans la requête du bateau vert.

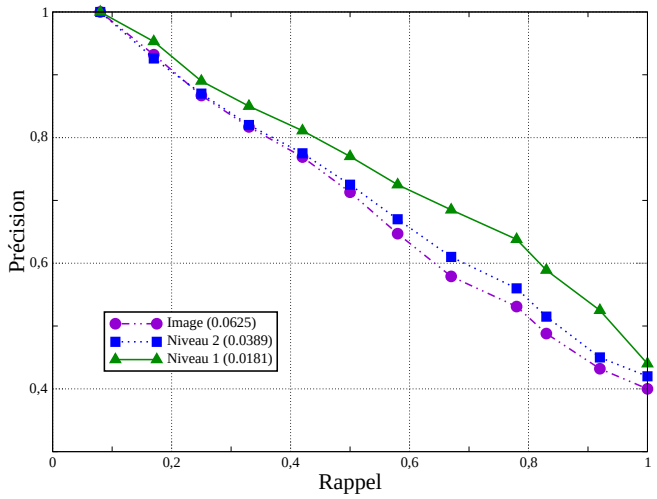


FIG. 7.3 – Mise en évidence de l'intérêt de la décomposition spatiale sur la base Columbia
Entre parenthèses dans la légende, les valeurs de $\widetilde{Rang}$.

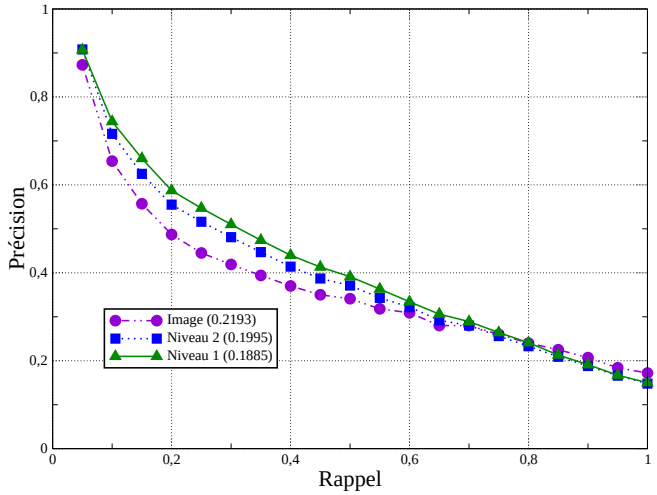


FIG. 7.4 – Intérêt de la décomposition spatiale sur la base de l'université de Washington
Entre parenthèses dans la légende, les valeurs de $\widetilde{Rang}$.

7.2.2 Prise en compte de l'arrangement spatial

La mesure définie précédemment calcule le poids minimum d'association entre les deux ensembles de régions. Or, il est obtenu sans prendre en compte l'information dont nous disposons sur l'arrangement de celles-ci. L'utilisation de cette information fera évoluer l'association des régions. La mesure de similarité résultante sera donc obligatoirement supérieure à la mesure précédente qui définit le minimum global. Cette remarque amène un premier critère d'optimisation : si la mesure de similarité simple montre que les ensembles de régions ne sont pas du tout similaires, il est alors inutile d'utiliser une technique d'appariement plus évoluée et donc plus complexe pour affiner la mesure.

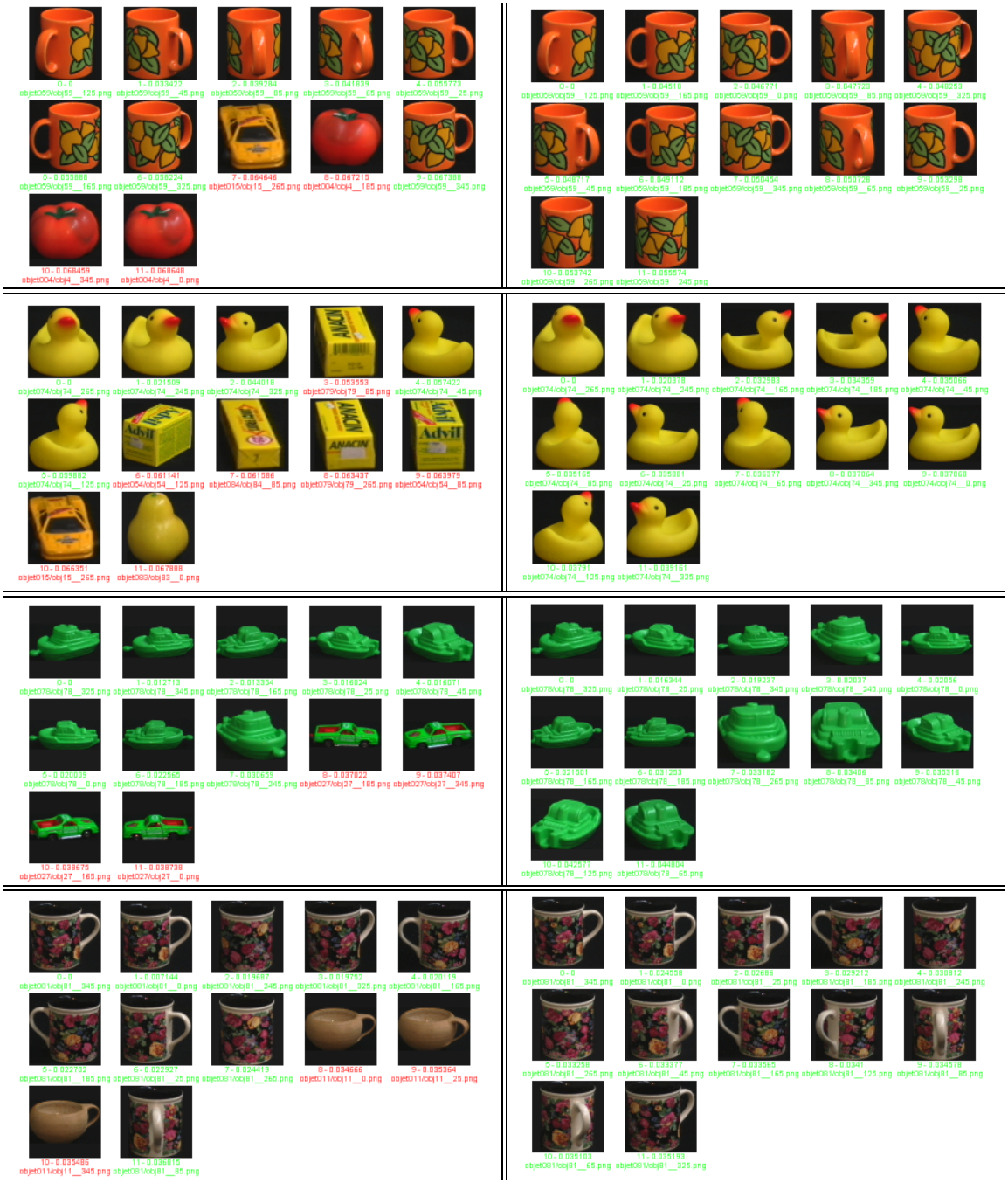


FIG. 7.5 – Mise en évidence de l'intérêt de l'approche spatiale sur quelques requêtes
À gauche : recherche globale, à droite : requête spatiale. Les images sont triées de gauche à droite et de haut en bas. L'image requête est donc située en haut à gauche.
Nos quatre requêtes spatiales présentées apportent un résultat parfait en retrouvant toutes les images de la même catégorie. Cela est possible grâce au partitionnement des images et à l'utilisation des caractéristiques couleur des différentes régions. Le fait d'isoler les différentes parties des objets engendre alors l'ajout d'une distance importante si elles ne sont pas retrouvées dans l'image comparée. Pour les tasses par exemple, les motifs sont caractéristiques et permettent de rejeter les images trouvées comme proches par le descripteur global. De même pour le bateau vert, le coffre rouge du camion retrouvé est peu important au sein du descripteur global alors que l'approche spatiale le met plus en évidence.

Pour prendre en compte l'arrangement spatial des régions, il paraît intéressant d'utiliser des méthodes de comparaison de graphes attribués. Nous présentons donc rapidement le problème général de l'appariement de graphes (ou *graph matching* en anglais) puis les deux techniques retenues pour son implantation.

7.2.2.1 Notations et terminologie

Un graphe $G = (V, E)$ est composé dans sa forme générale d'un ensemble de nœuds V (pour *vertice* en anglais) et d'un ensemble d'arêtes E (pour *edge* en anglais). Les arêtes (aussi appelées arcs ou lignes) lient deux nœuds (aussi appelés sommets ou points) non forcement distincts : $E \subset V \times V$.

L'ordre d'un graphe G est le nombre de sommets le composant ; il est noté $|V|$. Le nombre d'arcs est noté de manière similaire $|E|$.

Si deux nœuds u et $v \in V$ de G sont connectés par une arête $e \in E$, les deux sommets sont dits adjacents ou voisins et cette relation est notée $e = (u, v)$. Les arêtes peuvent être non orientées si elles n'ont pas de direction ; le graphe G contenant de telles arêtes est dit non orienté. Par contre, si (u, v) et (v, u) peuvent être distingués, les arcs sont alors orientés et le graphe est dit orienté. Un graphe G est également dit complet quand il existe une arête entre toutes les paires de nœuds.

Les éléments du graphe peuvent être porteurs d'information. Si cette information est un label, le graphe est alors dit labellisé. Si l'information portée est plus complexe, on parle alors de graphe attribué.

Un chemin entre deux sommets u et $u' \in V$ est une séquence non vide de k sommets différents $< v_1, \ldots, v_k >$ où $u = v_1$, $u' = v_k$ et $(v_{i-1}, v_i) \in E$, $\forall i \in [2, \ldots, k]$.

7.2.2.2 Les différents types de *graph matching*

De très nombreux domaines d'applications tels que la vision par ordinateur, la chimie ou la biologie moléculaire considèrent les images comme un ensemble de régions que l'on cherche à identifier ou à comparer avec d'autres. Pour que cela puisse être réalisé automatiquement sans intervention humaine, les informations présentes au sein des images sont généralement représentées sous forme de graphes. Ce système a déjà montré son efficacité dans la caractérisation d'objets [ESHERA84].

Pour décrire des objets ou des images, les nœuds représentent généralement leurs régions ou leurs caractéristiques. Par exemple, la représentation d'un vélo peut être celle présentée à la figure [7.6] : les différentes parties sont symbolisées par les sommets du graphe et les arêtes marquent les relations d'adjacence les liant. Des graphes similaires correspondent ainsi au même objet. Si les graphes sont non orientés, ils indiquent uniquement l'existence ou non de relations entre les sommets. D'un autre côté, les arêtes orientées permettent de symboliser des relations non symétriques. Sur cet exemple, le graphe est non orienté car la relation prise en compte est l'adjacence qui est symétrique.

Considérons que nous voulons déterminer si un graphe de données G_D est similaire à un graphe modèle G_M. De manière générale, le problème d'appariement de graphes s'exprime par : soit deux graphes $G_M = (V_M, E_M)$ et $G_D = (V_D, E_D)$ avec $|V_M| = |V_D|$, nous voulons trouver la fonction

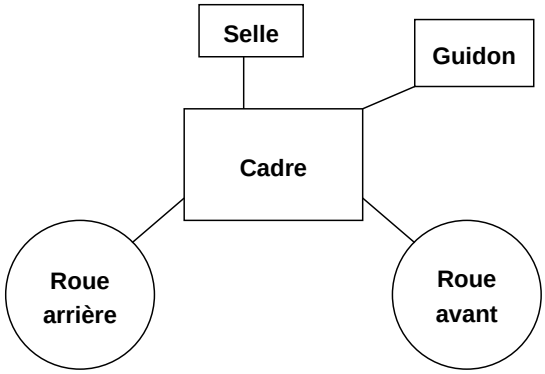


FIG. 7.6 – Illustration de la représentation de l'architecture d'un vélo par un graphe

d'association bijective $f : V_D \rightarrow V_M$ telle que $(u, v) \in E_D \Leftrightarrow (f(u), f(v)) \in E_M$. Si une telle fonction f existe alors c'est un isomorphisme et G_D est dit isomorphique à G_M. Cette catégorie de problème est dite *mise en correspondance exacte de graphes*.

Le terme d'appariement *inexact* s'applique alors à tous les autres problèmes. Ainsi, si la condition $|V_M| = |V_D|$ n'est pas satisfaite, il sera toujours possible de rechercher un isomorphisme entre des sous-graphes des deux structures mais il sera impossible de mettre en correspondance les deux graphes complètement. Cette catégorie de problèmes peut toujours être considérée comme un appariement exact car une fonction bijective est recherchée. De plus, dans le cadre des graphes attribués, même si l'ordre des deux graphes est identique cela n'implique pas qu'il existe un isomorphisme entre les graphes car la correspondance exacte entre les caractéristiques des nœuds n'est pas forcément satisfaite. Le cadre de mise en correspondance inexacte de graphes a pris son essor dans les 20 dernières années. En effet, les systèmes de segmentation automatique fournissent un nombre de régions variables. Par conséquent, seul le meilleur appariement possible peut être déterminé. Celui-ci peut être défini comme l'optimum d'une certaine fonction mesurant la similarité entre les nœuds et les arêtes appariés.

L'appariement exact ou inexact de graphes apparaît généralement dans la littérature sous les noms respectifs d'isomorphisme et d'homomorphisme de graphes. La figure [7.7] présente une classification de ces méthodes : d'un côté se trouve les techniques de recherche de fonctions bijectives et de l'autre celles qui tentent de trouver la meilleure mise en correspondance possible. Dans notre cas, nous désirons mesurer la ressemblance entre des graphes de tailles quelconques, nous nous situons donc dans la branche correspondant à l'appariement de sous-graphes attribués (grisé sur la figure).

Ce type de problèmes est considéré comme un des plus complexes en vision par ordinateur. Sa complexité est due à sa nature combinatoire. Dans la plupart des cas, ils sont au pire NP-complets. Pour certains types de graphes cette complexité devient linéaire mais il a été montré que le cas qui nous intéresse est NP-complet [GAREY79, ABDULK98].

7.2.2.3 Les systèmes d'appariement de sous-graphes attribués

Deux approches principales ont été développées pour résoudre ce problème. La première implique la construction d'un espace d'états[TSAI83, ESHERA84]. C'est globalement un ensemble

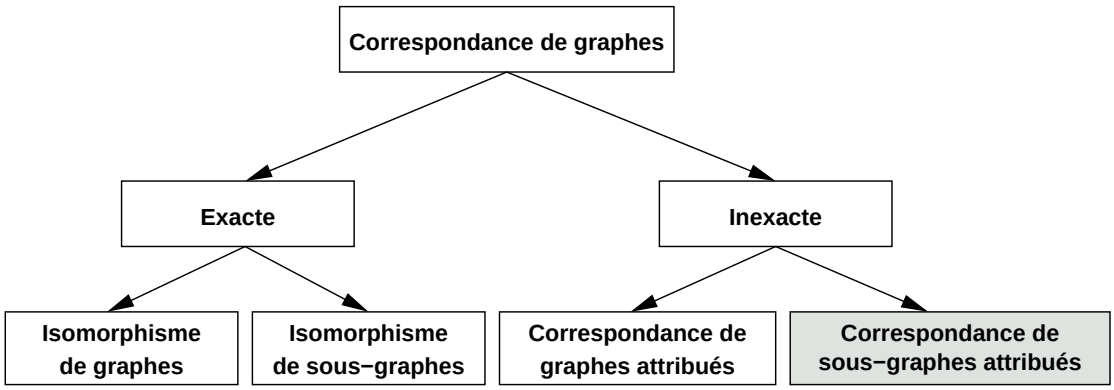


FIG. 7.7 – Classification des problèmes de mise en correspondance de graphes
En grisé, le problème qui nous intéresse.

de méthodes « brutales » qui explorent toutes les associations possibles. Le premier graphe G_D est transformé séquentiellement pour obtenir le graphe G_M. De nombreuses séquences de manipulations permettent de passer de G_D à G_M ; celle pouvant être réalisée avec un coût minimum est finalement retenue. Son coût définit alors la similarité entre les graphes. ESHERA propose ainsi de mettre en parallèle les constructions des deux graphes à partir de leurs éléments de base [ESHERA84]. Un arbre d'états est construit pour coder les créations des graphes réalisées en parallèle par ajouts successifs d'éléments de base. Un poids est affecté à chaque transition en fonction du type d'ajout et de la ressemblance entre les éléments ajoutés. Cela permet finalement de trouver les deux sous-graphes fournissant la similarité la plus importante. Cet algorithme a une complexité en $\mathcal{O}(|V_D|^3|V_M|^2)$ même si quelques cas spéciaux sont calculables en $\mathcal{O}(|V_D|^4)$

La seconde catégorie d'approches emploie des méthodes d'optimisation non linéaires pour approximer l'appariement optimal entre les graphes. Contrairement à la première famille de techniques, la recherche n'est pas exhaustive mais permet de s'approcher de la solution optimale en un temps raisonnable. Diverses techniques ont été utilisées dans cette voie : relaxation floue[DAVIS79, RANGAN92], réseaux de neurones[KUNER88, YOUNG97], techniques de programmation linéaire [ALMOHA93], algorithmes génétiques[KRCMAR94] et assignement gradué[GOLD96B] pour n'en donner que les principaux exemples. Nous n'entrerons pas ici dans le détail de toutes ces méthodes ce qui pourrait être très fastidieux et sans réel intérêt au sein de ce manuscrit. De plus, nous présenterons dans les sections suivantes les deux méthodes que nous avons implantées :

- la plus ancienne qui est la relaxation floue et en particulier la version de RANGANATH et CHIPMAN ;
- l'assignement gradué proposé par GOLD.

Ces famille de techniques ont une complexité moindre que la première car en $\mathcal{O}(|V_D||V_M|)$.

Ces divers algorithmes sont très intéressants pour réduire l'influence des erreurs de segmentation. En effet, à cause du bruit, des mauvaises conditions de prises de vue et des limitations de l'algorithme utilisé, cinq types d'erreurs sont possibles :

Erreurs de mesure : la présence de bruit peut engendrer des erreurs dans la détermination des attributs des régions. Les descripteurs de forme sont les plus sensibles à ce problème car une

variation locale de la position du contour peut entraîner un changement significatif dans la statistique calculée et par conséquent une mauvaise classification de la région ;

Objets non détectés : un ou plusieurs objets peuvent ne pas apparaître dans la segmentation. Ce problème peut arriver si l'objet n'est pas clairement visible dans l'image à cause, par exemple, d'une ombre ou d'une occultation. Les nœuds correspondants aux objets non visibles seront donc absents du graphe de l'image ;

Faux objets : C'est le cas inverse du précédent. Des ombres ou d'autres marques peuvent être segmentées sans pour autant correspondre à un objet réel ;

Sur-segmentation et objets fragmentés : Le bruit peut introduire dans la segmentation des frontières indésirables au sein des objets qui sont alors représentés au sein du graphe par plusieurs nœuds ;

Sous-segmentation et regroupement d'objets : C'est le cas contraire du précédent. La distinction entre plusieurs objets ne peut être réalisée, par exemple, si les frontières entre objets sont floues à cause du bruit. Ils se retrouvent alors dans une même région et donc dans un même sommet du graphe.

7.2.2.4 Méthode de relaxation floue

Nous présentons ici la méthode de relaxation floue que nous avons implantée. Celle-ci a été choisie pour sa simplicité et son utilisation importante dans différents domaines. Elle a été définie par RANGANATH et CHIPMAN [RANGAN92].

Considérons deux graphes attribués non orientés G et H composés respectivement de A et I nœuds. $\{G_a/a \in [1 \dots A]\}$ et $\{H_i/i \in [1 \dots I]\}$ représentent les ensembles des nœuds des deux graphes. De même, $\{G_{ab}/(a,b) \in [1 \dots A]^2\}$ et $\{H_{ij}/(i,j) \in [1 \dots I]^2\}$ sont les ensembles d'arêtes. Le nombre total d'associations possibles entre les objets est donc $A * I$. Pour cette présentation, la notion de sommet des graphes sera confondue avec celle de région de l'image. La distance globale définie entre les attributs de la a^{e} région de G et ceux de la i^{e} région de H s'écrira alors $D(G_a, H_i)$. Nous considérerons aussi que cette mesure est normalisée et comprise entre 0 et 1. Cette méthode effectue alors une relaxation floue sur le graphe d'association des sommets des deux graphes. La figure [7.8] résume les différentes étapes de ce processus.

Génération des nœuds du graphe d'association À la paire (G_a, H_i) est associée un poids $S(a,i)^{(0)}$ correspondant à la similarité entre les régions. Il peut être calculé par :

$$S(a,i)^{(0)} = 1 - D(G_a, H_i)$$

Cette première mesure permet de construire un graphe d'association entre les différents sommets. Ainsi, si $S(a,i)^{(0)}$ est supérieur à un seuil prédéterminé, la paire (G_a, H_i) apparaîtra comme une association possible et sera symbolisée par un nœud de ce graphe.

Génération des arêtes du graphe d'association Le poids de compatibilité entre les associations des différents nœuds définit alors les arêtes de ce graphe. Entre les sommets (G_a, H_i) et (G_b, H_j) une arête est créée à laquelle est associée C_{aibj} la mesure de similarité des relations G_{ab} et H_{ij}.

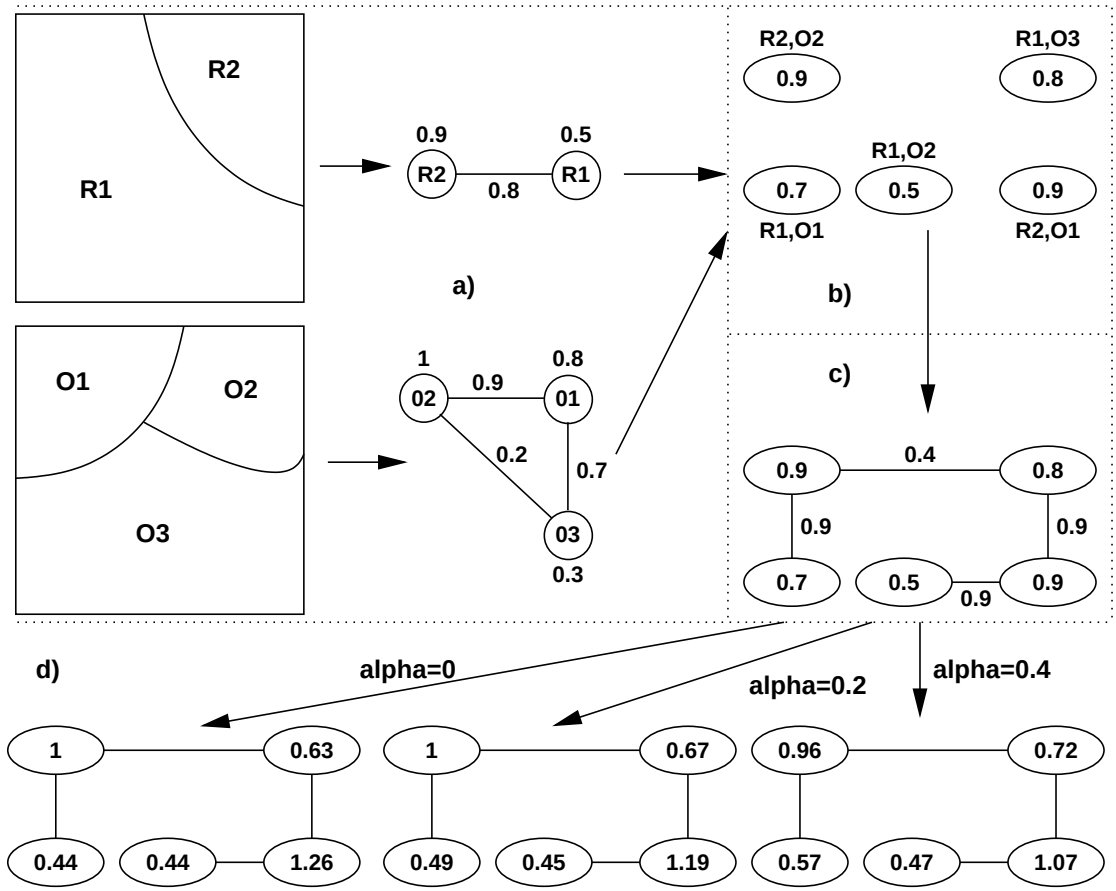


FIG. 7.8 – Schéma de principe de la relaxation floue

a) Pour illustrer le principe de la relaxation floue, nous disposons de deux images segmentées dont nous construisons les graphes d'adjacences attribués. Pour cet exemple, les nœuds et les arêtes sont caractérisés par un attribut réel.

b) Nous définissons alors la mesure de similarité entre sommets pour construire les nœuds du graphe d'association. La mesure $D(R_a, O_i) = 1 - |R_a - O_i|$ *permet d'obtenir les nœuds présentés en ne gardant que ceux strictement supérieurs à* 0.4*.*

c) Les attributs des différentes arêtes sont calculés avec la même mesure de ressemblance que pour les sommets.

d) Différentes valeurs de a *sont alors utilisées pour effectuer la relaxation ce qui permet d'être plus ou moins attaché aux données initiales. L'association finale retenue sera donc :* (R_1, O_3) *et* (R_2, O_1)*.*

C_{aibj} peut, par exemple, être calculée en comparant les attributs des arcs liants les différents sommets.

Pour ne pas prendre en compte l'association d'un nœud avec plusieurs nœuds, les valeurs de type C_{aiaj} et C_{aibi} sont mises à zéro. Les arêtes correspondantes ne sont donc pas générées.

Mise à jour itérative du graphe d'association par relaxation La méthode la plus répandue pour mettre à jour les poids d'association est la suivante :

$$S(a,i)^{(r+1)} = \frac{1}{A}\sum_{b=1}^{A}\left[\max_{j=1}^{I} S(a,i)^{(r)} C_{aibj}\right]$$

Une fois cette opération effectuée, les poids sont normalisés pour forcer la somme des poids à être constante :

$$S(a,i)^{(r+1)} = \frac{S(a,i)^{(r+1)}\sum_{b=1}^{A}\sum_{j=1}^{I} S(b,j)^{(0)}}{\sum_{b=1}^{A}\sum_{j=1}^{I} S(b,j)^{(r+1)}}$$

Le problème de cette formule est que l'attachement aux données initiales est très faible. Ainsi, même si un nœud a initialement un poids faible, il peut obtenir finalement un poids très important grâce aux diverses relations dont il dispose. Il paraît important d'intégrer les données initiales dans ce processus d'évolution. La formule devient alors :

$$S(a,i)^{(r+1)} = \alpha S(a,i)^{(0)} + (1-\alpha)\frac{1}{A}\sum_{b=1}^{A}\left[\max_{j=1}^{I} S(a,i)^{(r)} C_{aibj}\right]$$

où α est le facteur d'attachement aux données qui est compris entre 0 et 1. Si $\alpha = 0$, cette écriture est équivalente à la précédente. La figure [7.8] montre ainsi comment différentes valeurs de α permettent de contrôler les résultats en prenant plus ou moins en compte les poids initiaux.

Condition d'arrêt Le nombre d'itérations nécessaires à la convergence de l'algorithme de relaxation floue dépend des poids initiaux des nœuds et des arêtes du graphe d'association ainsi que de sa taille. Grâce à la normalisation des poids à chaque itération, le processus converge naturellement. Une condition de stabilité peut donc faire office de condition d'arrêt. Si, pour tous les couples (a,i), $|S(a,i)^{(r+1)} - S(a,i)^{(r)}| < \delta$ avec δ un seuil prédéterminé alors le processus est dit stable et il est arrêté.

Association finale des nœuds À partir des poids finaux du graphe d'association, pour chaque nœud du graphe de données le sommet correspondant du graphe modèle est défini simplement comme celui fournissant l'association maximale. À la fin du processus itératif, il peut exister des associations multiples au même sommet. Celles-ci ne sont pas problématiques dans notre cas car cette souplesse permet de gommer les erreurs de segmentation.

7.2.2.5 Méthode d'assignement gradué

Cette méthode est une des dernières méthodes d'optimisation non linéaire proposée par GOLD pour résoudre le problème d'appariement inexact de graphes attribués[GOLD96B]. Comme dans la plupart des systèmes d'optimisation, cette technique définit une fonction objectif comportant des contraintes permettant de résoudre un problème quelconque d'appariement. Comme exemple, nous présentons cette approche dans le cadre de mise en correspondance de graphes attribués.

Considérons deux graphes attribués non orientés G et H composés respectivement de A et I nœuds. $\{G_a/a \in [1 \dots A]\}$ et $\{H_i/i \in [1 \dots I]\}$ représentent les ensembles des nœuds des deux graphes. De même, $\{G_{ab}/(a,b) \in [1 \dots A]^2\}$ et $\{H_{ij}/(i,j) \in [1 \dots I]^2\}$ sont les ensembles d'arêtes. Nous voulons alors trouver la meilleure matrice de correspondances M minimisant la fonction objectif suivante :

$$E(M) = -\frac{1}{2}\sum_{a=1}^{A}\sum_{i=1}^{I}\sum_{b=1}^{A}\sum_{j=1}^{I} M_{ai}M_{bj}C^{(2)}_{aibj} + \alpha\sum_{a=1}^{A}\sum_{i=1}^{I} M_{ai}C^{(1)}_{ai} \tag{7.1}$$

soumise aux contraintes : $\forall a,\ \sum_{i=1}^{I} M_{ai} < 1$, $\forall i, \sum_{a=1}^{A} M_{ai} < 1$ et $\forall (a,i), M_{ai} \in [0 \dots 1]$.

$C^{(2)}_{aibj}$ correspond à la similarité entre les arêtes et peut être calculée par :

$$C^{(2)}_{aibj} = \begin{cases} 0 & \text{si } G_{ab} \text{ ou } H_{ij} \text{ n'existent pas} \\ D^{(2)}(G_{ab}, H_{ij}) & \text{sinon} \end{cases}$$

où $D^{(2)}$ est une mesure de similarité quelconque permettant de comparer les arêtes. GOLD utilise pour cela la notion de présence des arêtes. De cette manière, $D^{(2)}(G_{ab}, H_{ij}) = 1$ si les arêtes G_{ab} et H_{ij} existent. De même, si $D^{(1)}$ est la mesure de similarité entre les sommets :

$$C^{(1)}_{ai} = D^{(1)}(G_a, H_i)$$

$C^{(1)}_{ai} = 0$ si les attributs des nœuds ne sont pas semblables du tout et $C^{(1)}_{ai} = 1$ s'ils sont identiques.

À la convergence du processus, $M_{ai} = 1$ si le nœud G_a est associé au nœud H_i et $M_{ai} = 0$ dans le cas contraire. Initialement l'équation 7.1 est écrite par GOLD avec seulement le premier terme qui permet simplement de définir l'énergie d'association des nœuds. Dans le cas des graphes attribués, il ajoute ensuite le second terme pour avoir un terme quantifiant exclusivement la similarité entre les attributs.

La résolution de ce problème proposée par GOLD et RANGARAJAN est alors basée sur une technique de programmation linéaire. Étant donnée les conditions initiales M^0 pour la matrice d'appariements, la fonction objectif peut être développée autour de celles-ci par une série de TAYLOR :

$$\begin{aligned} E(M) &= -\frac{1}{2}\sum_{a=1}^{A}\sum_{i=1}^{I}\sum_{b=1}^{A}\sum_{j=1}^{I} M_{ai}M_{bj}C^{(2)}_{aibj} + \alpha\sum_{a=1}^{A}\sum_{i=1}^{I} M_{ai}C^{(1)}_{ai} \\ &\approx -\frac{1}{2}\sum_{a=1}^{A}\sum_{i=1}^{I}\sum_{b=1}^{A}\sum_{j=1}^{I} M^0_{ai}M^0_{bj}C^{(2)}_{aibj} - \sum_{a=1}^{A}\sum_{i=1}^{I} Q_{ai}(M_{ai} - M^0_{ai}) \end{aligned} \tag{7.2}$$

Algorithme 3 : Algorithme d'assignement gradué de GOLD

Données : M^0 la matrice d'appariement initiale
C^1 *et* C^2 les matrices de similarité entre nœuds et arêtes
Résultat : La matrice d'appariement finale
début
1 - Initialisation de M : $M = M^0$;
tant que *la convergence ou le nombre maximum d'itérations ne sont pas atteints* **faire**
2 - Réaliser le développement de TAYLOR présenté à l'équation 7.2 ;
3 - Résoudre le problème d'assignement résultant (cf. équation 7.3) en utilisant la méthode choisie ;
4 - Prendre le résultat, i.e. M, et l'injecter dans l'équation 7.2 ;
fin
fin

où

$$Q_{ai} = -\frac{\delta E(M)}{\delta M_{ai}} = \sum_{b=1}^{A}\sum_{j=1}^{I} M_{bj} C^{(2)}_{aibj} + \alpha C^{(1)}_{ai} \tag{7.3}$$

La minimisation de $E(M)$ revient donc à maximiser l'expression $\sum_{a=1}^{A}\sum_{i=1}^{I} Q_{ai} M_{ai}$.

Ce terme représente un problème d'assignement qui peut être résolu par différentes techniques. La solution proposée par GOLD est très efficace. De plus, la convergence de cette technique sous les contraintes de l'étude a été prouvée[SINKHO64]. L'algorithme [3] présente les grandes lignes de cette méthode. L'ensemble des détails sur l'algorithme d'assignement employé par GOLD sont présentés dans [GOLD96B].

Finalement, la distance entre les structures appariées peut ne prendre en compte que les distances entre attributs. Mais, il est également possible d'introduire un facteur quantifiant la similarité des relations. La matrice d'associations finale ne comprend que des 1 et des 0 pour symboliser l'association ou non entre les nœuds. Les mesures possibles sont alors :

$$\begin{aligned} S_{AG1}(G,H) &= \sum_{a=1}^{A}\sum_{i=1}^{I} M_{ai}\left(1 - C^{(1)}_{ai}\right) \\ S_{AG2}(G,H) &= \sum_{a=1}^{A}\sum_{i=1}^{I} M_{ai}\left(1 - C^{(1)}_{ai}\right) + \beta \sum_{a=1}^{A}\sum_{b=1}^{A} V_{ab} \sum_{i=1}^{I}\sum_{j=1}^{I} M_{ai} * M_{bj} * (1 - C^{(2)}_{aibj}) \end{aligned}$$

avec $V_{ab} = 1$ si les nœuds a et b sont adjacents et $V_{ab} = 0$ sinon. β est le poids alloué à la similarité entre arêtes.

Les résultats présentés par GOLD sont très intéressants tant en temps de calcul qu'en précision des comparaisons. Nous verrons que c'est également le cas dans le cadre de nos expérimentations.

7.2.2.6 Résultats

La complexité de ces deux techniques, bien qu'inférieure aux techniques brutales, est toujours fortement supérieure aux méthodes simples présentées précédemment. Il semble donc irréaliste de les utiliser de manière étendue sur des niveaux entiers caractérisés par des graphes d'ordres importants. Nous ne présentons donc pas de résultats pour ces deux méthodes pour la recherche de similarité entre images. En revanche, l'intérêt de tels algorithmes est illustré plus loin pour la recherche d'objets (cf. section 7.4).

7.3 Intérêt de la structure pyramidale

Nous avons déjà montré l'intérêt de la décomposition de l'image en objets pour effectuer une recherche d'images par le contenu. Nous allons maintenant voir en quoi la structure hiérarchique est intéressante. Il paraît évident que la détection d'un objet est possible ou non suivant le niveau de finesse de la segmentation. Ainsi, un objet ou une partie d'objet de petite taille ne sera isolé que dans les niveaux inférieurs de la pyramide. Il semble donc pertinent d'utiliser les différents niveaux pour caractériser un objet et par conséquent une image.

7.3.1 Algorithmes développés

Nous présentons dans cette section deux méthodes simples mises en place pour utiliser la structure multi-échelles dans la comparaison de graphes pyramidaux.

Cas des bases spécialisées Au sein des bases spécialisées telles que Columbia ou SOIL-47 (cf. annexe B), les prises de vues sont généralement très semblables. Pour ces deux bases par exemple, les images des objets ont toutes la même taille et les prises de vues sont effectuées à une distance quasi-constante des objets. Dans ces conditions, l'algorithme de segmentation multi-échelles que nous avons défini fournit sur l'ensemble de la base des niveaux indépendants et cohérents. Par indépendant nous entendons que chaque échelle met en évidence des éléments différents. Le premier niveau, le plus grossier, sépare ainsi l'objet du fond alors que le niveau le plus fin extrait les composants des objets. Par cohérent il faut comprendre que pour chaque image l'indépendance des niveaux est obtenue. Par ces deux remarques, il semble que pour des bases spécialisées la comparaison d'échelles différentes n'est pas pertinente. Il est donc préférable de comparer les images en mesurant la similarité entre les niveaux équivalents (cf. figure [7.9]).

Soit deux graphes pyramidaux $P = \{P_i \ / \ i \in [1 \dots N]\}$ et $Q = \{Q_i \ / \ i \in [1 \dots N]\}$ composés de N niveaux et S la mesure de similarité entre niveaux. Dans ce cas, la ressemblance entre pyramides peut s'exprimer de différentes manières :

$$\begin{aligned} S_{SN1}(P,Q) &= \min_{i \in [1 \dots N]} S(P_i, Q_i) \\ S_{SN2}(P,Q) &= \frac{1}{N} \sum_{i=1}^{N} S(P_i, Q_i) \end{aligned}$$

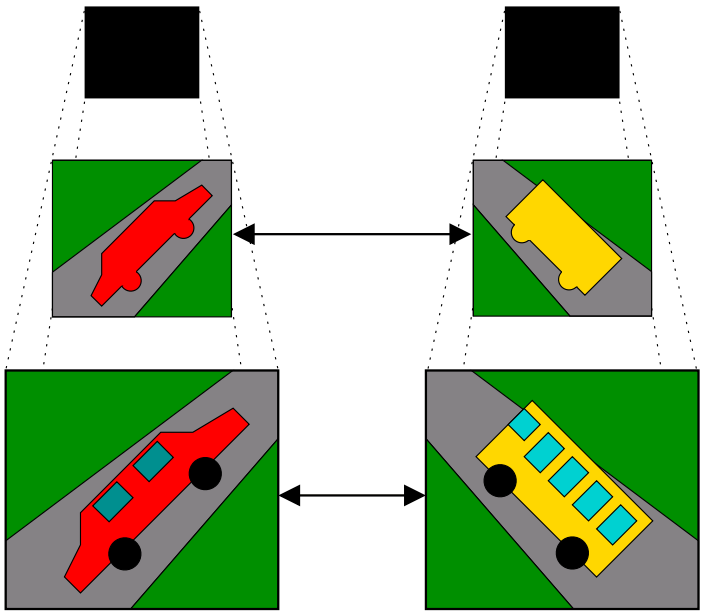


FIG. 7.9 – Utilisation simple de la structure pyramidale
Les flèches symbolisent les comparaisons réalisées entre les niveaux.

Suivant l'application, le choix des niveaux pris en compte peut varier. Ainsi, comme nous le verrons dans les résultats, l'utilisation du niveau correspondant à l'image n'est généralement pas pertinent. Avec ces mesures, le coût calculatoire est simplement N fois plus important que pour la comparaison d'un niveau simple.

Cas des bases génériques Pour des bases plus génériques pour lesquelles la composition n'est pas homogène, la méthode précédente n'est pas adaptée car elle ne permet pas de détecter des objets vus à des échelles différentes. Pour la base de fresques médiévales du CESCM par exemple, avec la distance précédente il serait impossible de trouver un objet présent à différentes résolutions. Il est donc nécessaire de comparer tous les niveaux entre eux (cf. figure [7.10]).

Si P et Q sont composés respectivement de N et M niveaux, les distances définies précédemment s'étendent de façon suivante :

$$\begin{aligned} S_{CN1}(P,Q) &= \min_{\substack{i\in[1\ldots N]\\ j\in[1\ldots M]}} S(P_i,Q_j) \\ S_{CN2}(P,Q) &= \frac{1}{N}\sum_{i=1}^{N} \min_{j\in[1\ldots M]} S(P_i,Q_j) \end{aligned}$$

S_{CN1} correspond à la recherche de la meilleure similarité entre niveaux. Dans le cas d'une vue d'une même scène à plusieurs échelles comme à la figure [7.10], la similarité minimale sera trouvée pour des niveaux différents.

Ces évolutions augmentent le coût calculatoire qui est alors $N * M$ fois plus important que pour la comparaison d'un niveau simple. Cette complexité peut être réduite si les images ne sont

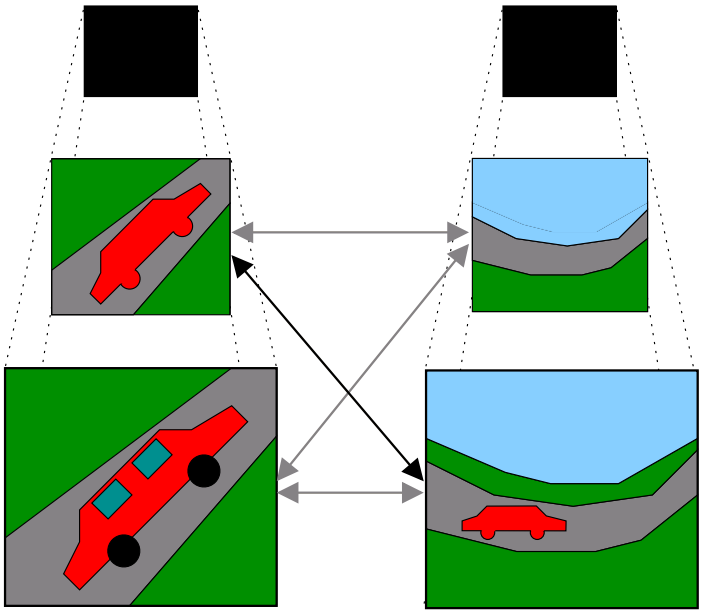


FIG. 7.10 – Utilisation complète de la structure pyramidale
Les flèches symbolisent les comparaisons réalisées entre les niveaux.
La flèche noire montre quant à elle la similarité la plus forte qui est trouvée.

pas du tout similaires. Dans ce cas, les premières comparaisons peuvent permettre de dire que les images ne sont pas semblables et qu'il est inutile d'effectuer le reste des calculs.

7.3.2 Résultats

Tout comme pour la décomposition spatiale pour rester dans un cadre général, nous utilisons une comparaison de la composition colorée des régions fondée sur l'*Earth Mover's Distance* utilisant une distance euclidienne dans l'espace $L^*a^*b^*$. Globalement, tous les tests menés ont montré l'intérêt d'utiliser l'ensemble des niveaux de la pyramide plutôt qu'un seul. La base Columbia et la base de l'université de Washington ont ainsi permis d'obtenir les graphes précision/rappel présentés aux figures [7.11] et [7.12]. Entre parenthèses, dans la légende, sont indiquées les valeurs de $\widetilde{Rang}$. Sur ces deux bases, l'ensemble des niveaux permet d'accroître légèrement la précision générale du système. Il faut également remarquer que l'utilisation du niveau correspondant à l'image dans ce type de comparaison réduit la performance du système en introduisant une mesure de ressemblance moins fiable. En effet, celle-ci correspond à une description globale de l'image qui, comme nous l'avons montré précédemment, est beaucoup moins précise qu'une caractérisation spatiale.

L'utilisation combinée des différentes échelles permet surtout de réduire l'impact des erreurs de segmentation. Ainsi, si à un niveau, la segmentation n'est pas adaptée au contenu, elle pourra être bien meilleure pour les autres niveaux qui fourniront alors une mesure de similarité beaucoup plus pertinente. La figure [7.15] présente en fin de chapitre montre quelques exemples de ce phénomène.

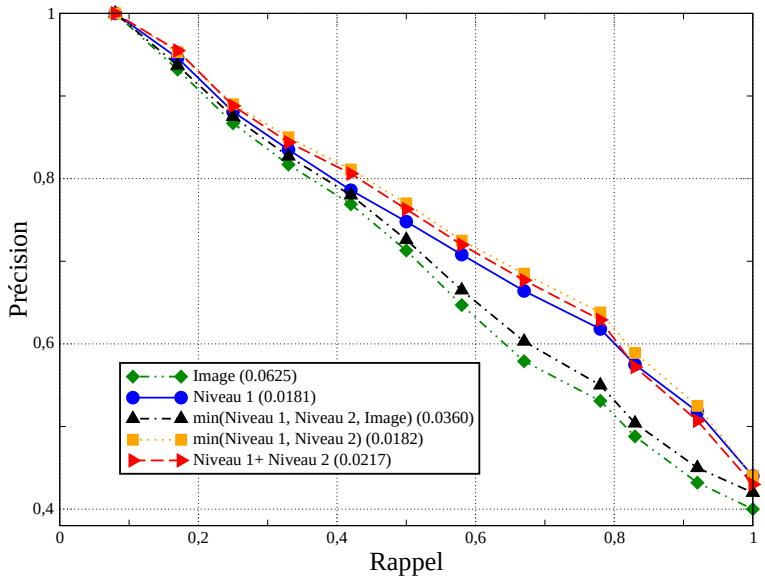


FIG. 7.11 – Mise en évidence de l'intérêt de l'aspect pyramidal sur la base Columbia

7.4 Système de recherche d'objets

7.4.1 Principe développé

La recherche d'objets utilise les mêmes algorithmes que la recherche de similarité entre images. L'objet est alors décrit par la sous-pyramide extraite de l'image d'où l'objet est tiré ou bien celle générée à partir de l'esquisse définissant l'objet. C'est alors ce graphe pyramidal de l'objet qui est utilisé pour effectuer la recherche tout comme pourrait l'être une image entière (cf. figure [7.13]). L'interface produite durant cette thèse permet de sélectionner les régions générées par l'algorithme de segmentation et ainsi délimiter l'objet à rechercher (cf. figure [7.14]). Dans ce cas, les systèmes d'appariement de graphes permettent de garder une cohésion entre les régions au sein de la requête et améliorent sensiblement les résultats obtenus.

L'interface permet aussi d'effectuer en cascade tous les traitements définis dans ce manuscrit : segmentation multi-échelles, création des graphes pyramidaux et calcul de la similarité entre structures suivant la méthode choisie. Les images servant à l'étude sont présentées sous forme d'un arbre symbolisant leur appartenance à une base d'images, à une catégorie et à une sous-catégorie si nécessaire. Pour chacune d'elle, une icône différente présente son état de traitement (segmentation réalisée ou non et graphe pyramidal généré ou non). Dès la sélection d'une image, elle est présentée dans la fenêtre centrale ainsi que sa segmentation multi-échelles si elle est disponible. Il est alors possible de ne voir que certains niveaux de segmentations et d'effectuer un zoom. La sélection de l'objet à rechercher est très simple. Après avoir choisi le niveau d'étude, les régions sélectionnées sont mises en évidence.

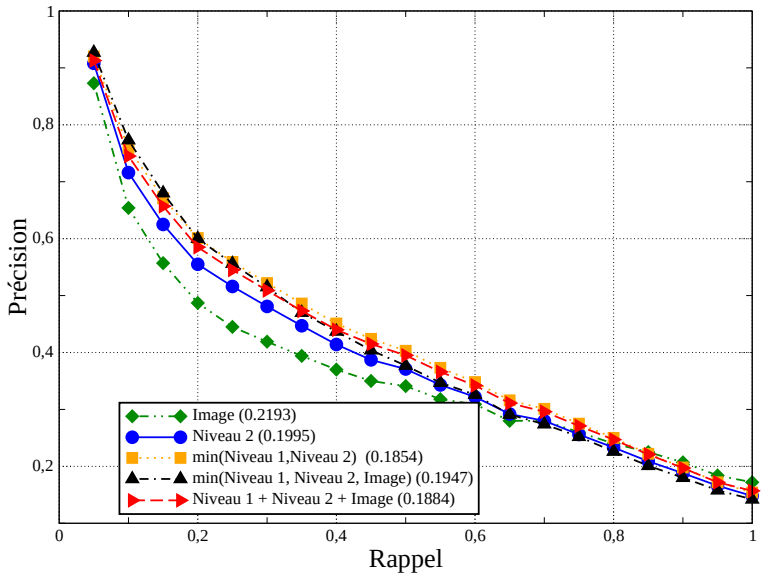


FIG. 7.12 – Intérêt de l'aspect pyramidal sur la base de l'université de Washington

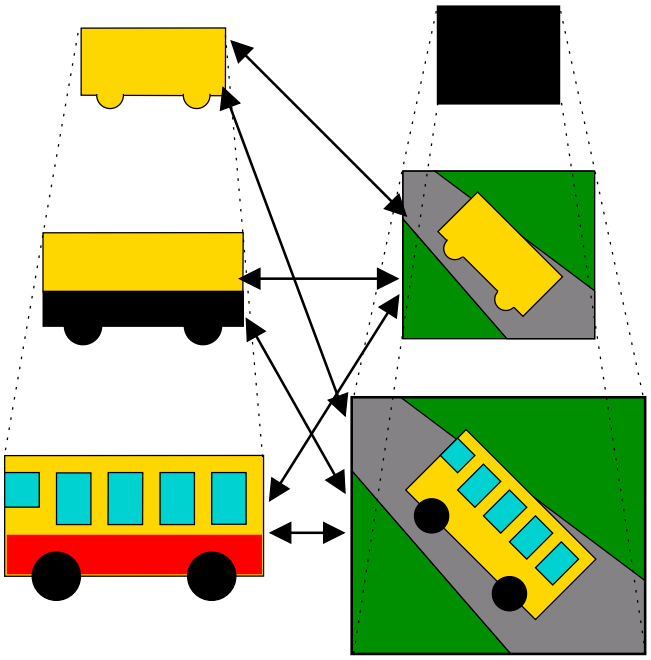


FIG. 7.13 – Recherche d'objets par recherche de sous-pyramide
À gauche : l'objet extrait d'une image ; à droite l'image dans laquelle l'objet est recherché.

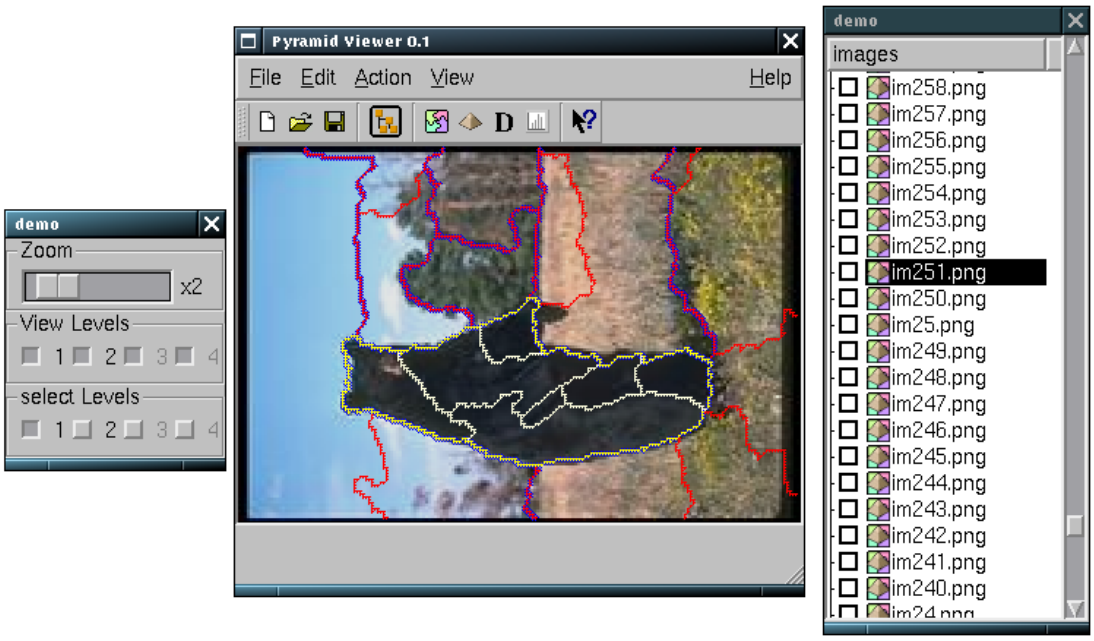


FIG. 7.14 – Sélection d'un objet à partir de l'interface homme-machine
Les traits bleus et rouges montrent les divisions des deux niveaux de segmentation calculés. Les lignes jaunes symbolisent la sélection globale effectuée à une échelle donnée. Les traits blancs permettent de visualiser les régions sélectionnées bien qu'intérieures à la sélection globale. À gauche est présentée la fenêtre permettant de zoomer et de choisir les échelles à voir et celle sur laquelle les régions sont sélectionnées. À droite, un arbre présente les différentes images disponibles ainsi que leur niveau de traitement.

7.4.2 Résultats

Les requêtes présentées ont été obtenues sur la base d'images de PICTOSEEK (cf. section B.2). Pour chaque image, trois types de résultats sont présentés (cf. figures [7.16] et [7.17] en fin de chapitre) :

- le premier n'utilisant que la statistique couleur globale de l'image ;
- le deuxième en ne prenant en compte que la décomposition de l'image et l'aspect pyramidal avec une statistique combinant la couleur, la texture et la forme ;
- le troisième utilisant en plus l'information spatiale grâce à l'un des deux algorithmes d'appariement de graphes (cf. section 7.2.2).

Malgré le contexte qui apporte beaucoup d'information à la recherche d'objets, ces résultats montrent que l'utilisation de l'objet uniquement est très intéressante. Ainsi, dans les diverses requêtes présentées, les images retournées par les requêtes partielles sont plus pertinentes. Globalement, l'utilisation de l'information spatiale apporte également un gain d'information non négligeable sur la plupart des requêtes ce qui permet d'obtenir des résultats encore plus précis. Ces techniques permettent ainsi d'exclure les images pour lesquelles les projections des régions apportent une faible distance mais sont complètement désordonnées. Ce phénomène est illustré par exemple sur les figures [7.16] et [7.17].

7.5 Comparaisons avec d'autres systèmes

7.5.1 Une comparaison difficile

Les diverses publications présentées en indexation d'images par le contenu sont généralement illustrées par des exemples de requêtes (bien choisis ?) mais des indices de performances globales sont souvent absents ou calculés sur un nombre restreint d'images (bien choisies ?). Il apparaît donc très difficile de comparer objectivement les résultats obtenus par les différents systèmes.

De plus, très peu de systèmes sont actuellement disponibles en ligne car commerciaux ou simplement sans démonstrateur. La comparaison effectuée est donc très restreinte.

Nous n'avons ainsi pu comparer nos méthodes de manière qualitative qu'avec trois systèmes d'indexation et quantitativement qu'avec un seul.

7.5.2 Tests réalisés

Tous les résultats de recherche présentés montrent les images les plus proches de l'image requête de gauche à droite et de haut en bas. L'image requête se trouve donc en haut à gauche et la plus éloignée en bas à droite. Pour plus de lisibilité, l'ensemble de ces figures sont reportées à la fin de ce chapitre.

7.5.2.1 PICTOSEEK

PICTOSEEK est le système développé au sein du laboratoire ISIS[1] de l'université d'Amsterdam. Ce projet est principalement dirigé par THEO GEVERS[GEVERS99B, GEVERS00]. La recherche est réalisée par comparaison d'invariants colorimétriques définis globalement pour les images.

Nous avons mis en concurrence ce système avec la version de recherche spatiale avec comparaison complète des niveaux présentée à la section 7.3.1. Seul l'attribut de couleur est pris en compte par l'intersection d'histogrammes à supports colorimétriques différents (cf. section 6.2.4.3). La figure [7.18] met en parallèle les résultats des deux méthodes et montre bien l'apport qualitatif de notre système. Les images retrouvées sont globalement plus pertinentes que ce soit pour un coucher de soleil, une voiture ou des éléphants dans la brousse.

7.5.2.2 IKONA

IKONA est le projet global d'indexation d'images par le contenu de l'équipe IMEDIA de l'INRIA de Rocquencourt. Les statistiques utilisées ne sont pas précisées explicitement sur le site du démonstrateur et dans les divers articles présentés. D'après [BOUJEM01], des histogrammes couleurs pondérés sont utilisés pour caractériser les images par une statistique de type couleur/texture. Les figures [7.19] et [7.20] présentent des requêtes posant quelques problèmes engendrés par des images très proches visuellement dans la base Columbia mais représentant des objets différents. Elles

[1] *Intelligent Sensory Information Systems*.

montrent que notre système se comporte aussi bien voire mieux dans certains cas qu'IKONA en utilisant une approche spatiale et pyramidale.

7.5.2.3 FIDS de l'université de Washington

FIDS (pour *Fast Image Database System*) est développé par l'équipe *Computer Graphics, Computer Vision and Animation* de l'université de Washington et particulièrement par LINDA SHAPIRO. Deux démonstrateurs sont disponibles. Le premier utilise uniquement une statistique couleur appelée *Color Orientation* (cf. figure [7.21]) et le second permet de combiner plusieurs descripteurs (cf. figure [7.22]). Les résultats présentés pour le second ont été obtenus après optimisation manuel des poids des différentes statistiques. Globalement ces différentes approches fournissent de bons résultats qui sont globalement équivalents à nos résultats. Sur certaines requêtes, les images retournées par notre système apparaissent tout de même plus cohérentes.

7.5.2.4 SOIL-47

Dans [KOUBAR02], KOUBAROULIS et al. présentent une nouvelle méthode nommée *Multimodal Neighbourhood Signature* (MNS). C'est une méthode de reconnaissance et de recherche d'objets multicolores basée sur la couleur. La structure couleur locale est représentée par des attributs invariants à l'illumination calculés à partir des voisinages de l'image avec une fonction de densité couleur multimodale. Les positions des différents modes utilisés pour le calcul des invariants sont établis efficacement dans l'espace RGB.

Les expérimentations effectuées ont été réalisées à partir de la base d'images SOIL-47 (cf. annexe B.4). Les requêtes sont les vues frontales des différents objets avec comme ensemble de recherche les images d'un même angle de vue. Le pourcentage de requêtes fournissant l'image du même objet pour l'angle de vue étudié en première position sont alors calculés (ou dans les trois premières positions). Les résultats sont présentés à la figure [7.23] et montrent que notre système de recherche fondée sur une recherche simple des niveaux pour une base spécialisée (cf. section 7.3) obtient une précision de 30 à 40% supérieure quand seule la première image retrouvée est prise en compte et de 20% pour les 3 premières. Cette comparaison montre bien l'intérêt d'une approche spatiale par rapport à une description globale des images.

7.6 Conclusion

Ce chapitre a présenté les différentes techniques développées pour comparer deux graphes pyramidaux. Nous avons tout d'abord montré l'intérêt de cette représentation du point de vue de la décomposition des images à un niveau donné. L'aspect pyramidal de la structure permet alors d'améliorer encore la précision du système par des algorithmes simples. Ces méthodes fonctionnent très bien pour des recherches d'images globales mais il est nécessaire, pour rendre la recherche d'objets plus fiable, d'introduire des algorithmes d'appariement de graphes attribués. Dans ce cas, l'arrangement spatial des régions permet d'affiner la recherche et de ne garder que les images où la projection des régions de l'objet n'est pas désordonnée.

Nous avons enfin comparé qualitativement et quantitativement notre méthode avec d'autres systèmes disponibles. Les résultats montrent globalement que cette structure apporte un plus non négligeable.

Le chapitre suivant se propose enfin d'appliquer ces différentes techniques au problème des fresques médiévales du CESCM.

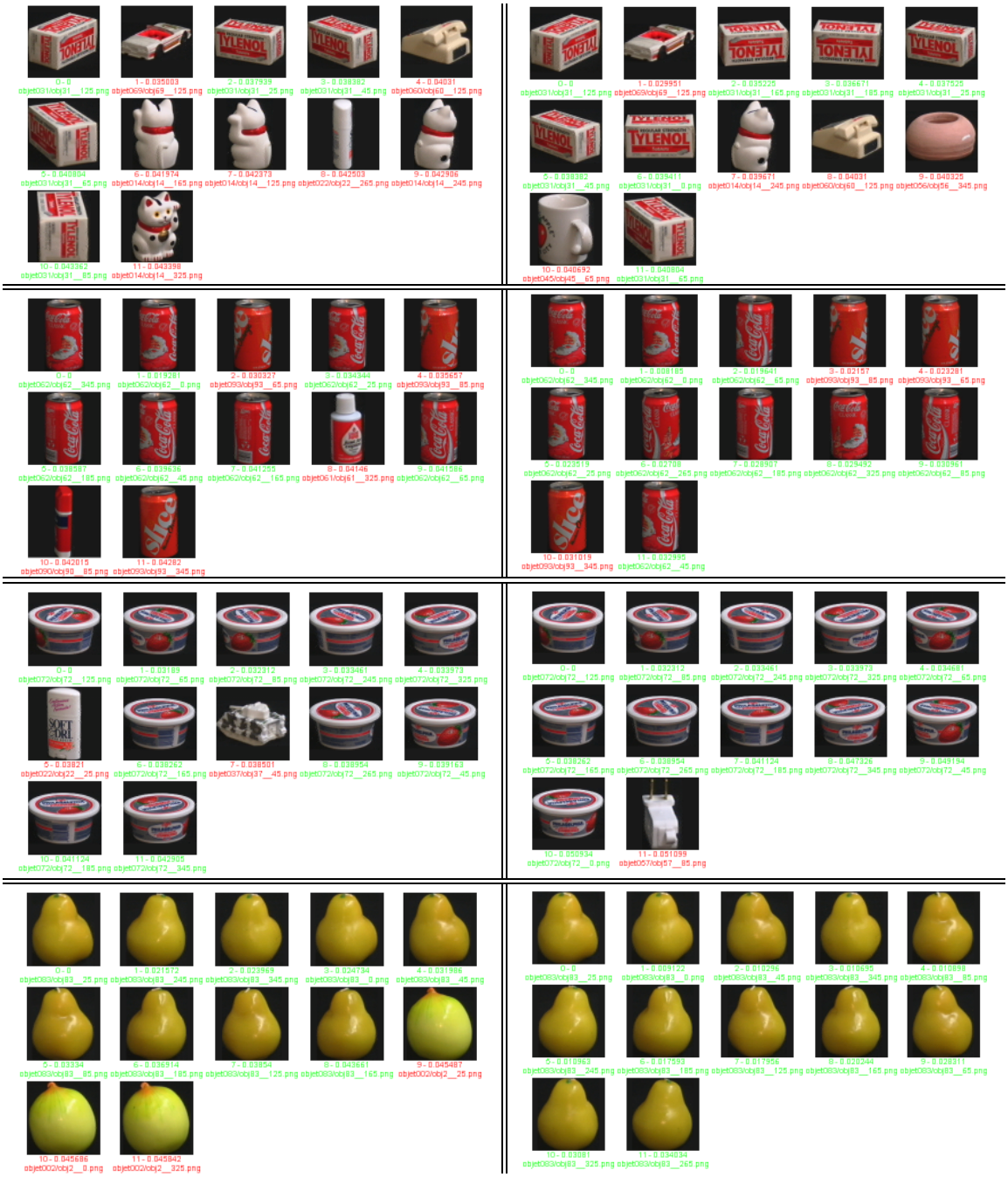


FIG. 7.15 – Mise en évidence de l'intérêt de l'approche pyramidale sur quelques requêtes
À gauche : recherche sur un niveau de décomposition, à droite : requête utilisant tous les niveaux. Les images sont triées de gauche à droite et de haut en bas. Même si l'algorithme de segmentation choisi est globalement cohérent pour l'ensemble des images, il engendre quelques segmentations peu pertinentes. Ces erreurs provoquent alors des recherches spatiales moins fiables. L'utilisation combinée des différents niveaux réduit ces problèmes. En effet, même si pour un niveau le partitionnement est aberrant, il l'est rarement pour toutes les échelles. Prendre en compte l'ensemble de la pyramide permet alors de choisir le meilleur niveau et ainsi de gommer ces erreurs. Dans les quatre exemples proposés, les requêtes sur un niveau ont été effectuées sur une échelle mal partitionnée. L'utilisation des autres échelles (à droite) réduit alors fortement l'influence de cette erreur même si elle n'est pas complètement supprimée. La segmentation nécessite donc encore d'être améliorée pour rendre plus précise notre représentation.

FIG. 7.16 – Quelques recherches d'objets au sein de la base de PICTOSEEK (1/2)
En haut : recherche globale, au milieu : utilisation de l'aspect pyramidal et de la décomposition, en bas : utilisation des aspects pyramidal et spatial.
Les images sont triées de gauche à droite et de haut en bas.

Sur ces deux exemples de recherche de voitures, nous pouvons tout d'abord remarquer que la recherche globale fondée sur les histogrammes couleur fournit des résultats pour lesquels certaines images retrouvées n'ont une ressemblance forte que sur une zone de l'image (généralement le fond ou l'objet principal). Le fait de limiter la recherche à l'objet d'intérêt permet alors d'affiner la recherche tant du point de vue de la zone prise en compte que de la qualité de la description. En effet, l'ajout de l'attribut de forme semble très important dans ce cas car il permet de retrouver des voitures de formes similaires mais de couleurs différentes. Finalement l'ajout de la contrainte spatiale par un algorithme d'appariement de graphes attribués permet d'ordonner la projection de l'objet et d'exclure ainsi certaines images. À droite les deux couchers de soleil sont ainsi exclus et une image de la même voiture apparaît en huitième position.

FIG. 7.17 – Quelques recherches d'objets au sein de la base de PICTOSEEK (2/2)
En haut : recherche globale, au milieu : utilisation de l'aspect pyramidal et de la décomposition, en bas : utilisation des aspects pyramidal et spatial.
Les images sont triées de gauche à droite et de bas en haut.
Les mêmes remarques effectuées sur la figure précédente sont applicables sur celle-ci. Ici, bien que la recherche globale fournissent des images très proches colorimétriquement et même par leur contenu, le fait de restreindre la zone de recherche à l'objet d'intérêt permet d'améliorer la qualité des résultats. Les objets sont alors décrits plus finement.

FIG. 7.18 – Comparaison de recherches de PICTOSEEK avec notre système
À gauche : PICTOSEEK*, à droite : notre système.*

Les requêtes proposés pour PICTOSEEK *montrent les limites de l'utilisation de statistiques globales. Sur ces exemples, celle-ci fournit des résultats peu pertinents du point de vue du contenu des images. Malgré l'utilisation de l'image dans son ensemble, notre approche présente quant à elle des images plus proches par les objets qui les composent. Ainsi, pour les trois exemples proposés, le contenu des images retournées sont plus en accord avec celui de l'image requête que ce soit pour une voiture, un éléphant ou un coucher de soleil. Nous pouvons tout de même remarquer que le fait de ne pas limiter la recherche à l'objet d'intérêt fait que certaines images proposées sont proches colorimétriquement de la requête sans pour autant contenir des objets du même type.*

FIG. 7.19 – Comparaison de recherches d'IKONA avec notre système (1/2)
À gauche : IKONA*, à droite : notre système.*

*Bien que ces trois requêtes montrent l'apport de notre système par la nature spatiale de notre approche, le premier exemple est assez étonnant. En effet, l'intérieur rouge de la voiture semble peu discriminant. Ceci s'explique en partie par le fait qu'il correspond à une unique région au sein des différents niveaux alors que le reste de la voiture est sur-segmenté. Par contre, les différentes tasses retrouvées (même la première) sont composées de régions colorimétriquement proches de la voiture. La tasse est grise et le motif est rouge. Sur la première tasse même si elle est peu visible, une petite portion du symbole est apparent et segmenté ce qui explique sa présence à cette position. Pour l'exemple du milieu, les spatules retrouvées en premier sont celles de même orientation car, dans ce cas, l'algorithme de segmentation produit des segmentations très similaires engendrant des pyramides quasi-identiques. Pour le dernier exemple, les résultats d'*IKONA *sont peu cohérents ce qui paraît anormal au vue de notre résultat pour lequel la recherche est parfaite. Ce problème s'explique en partie par l'aspect spatial de notre approche qui délimite bien les différents éléments du chat permettant ainsi de la caractériser plus finement et d'exclure les tasses de forme voisine.*

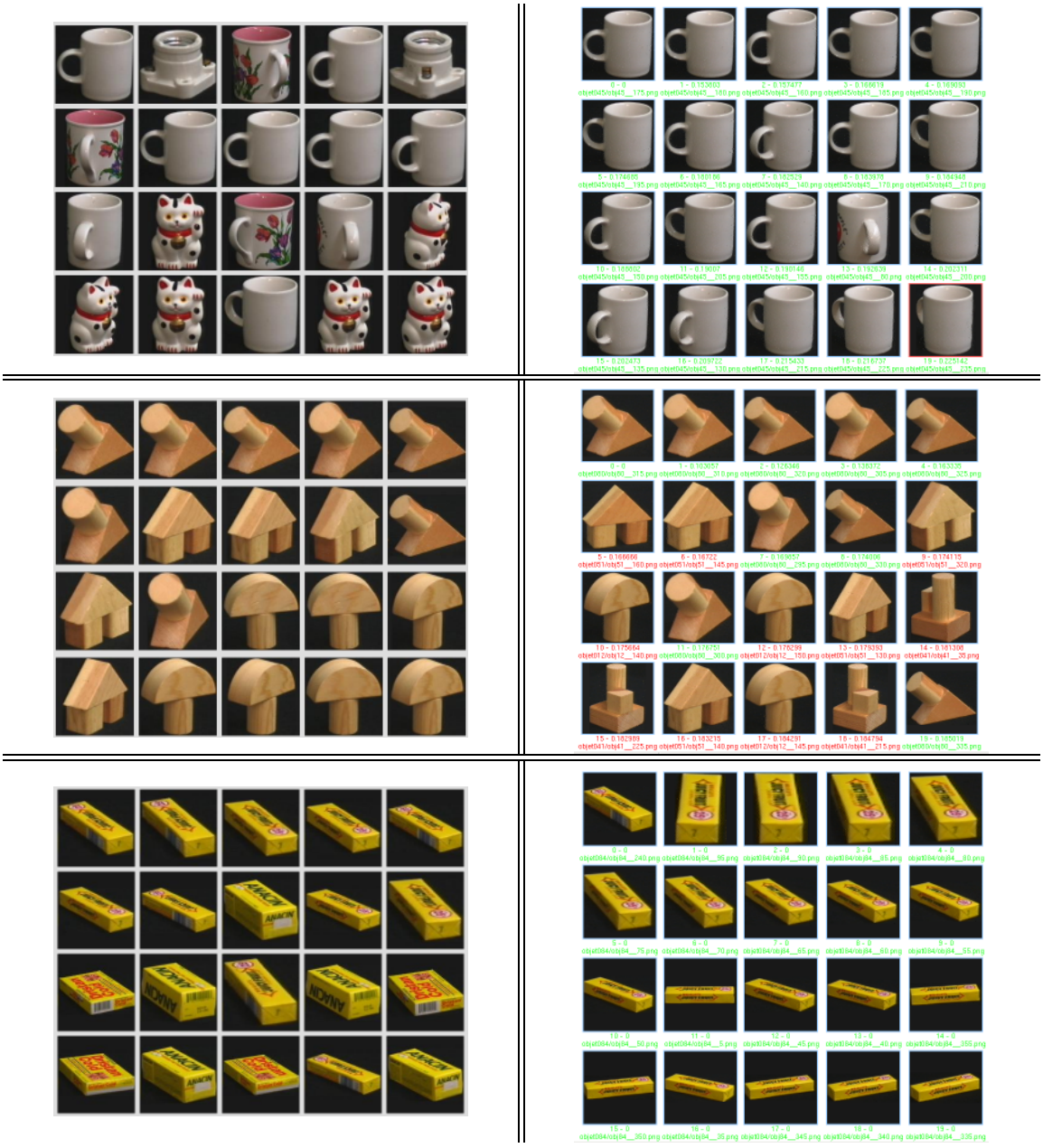


FIG. 7.20 – Comparaison de recherches d'IKONA avec notre système (2/2)
À gauche : IKONA*, à droite : notre système.*

De même que sur la dernière requête de l'exemple précédent, les images retournées par IKONA *pour la première requête de cette figure semblent bizarres comparées à nos résultats. Cette recherche ne semble en effet pas poser de gros problèmes vu que la tasse est colorimétriquement uniforme. Il semble ainsi peu normal de retrouver des images comportant des zones rose ou rouge. Les résultats pour la seconde requête sont sensiblement identiques car pour les deux systèmes les objets en bois sont très semblables. Quant à la dernière requête, il est à noter que les premières images retrouvées sont celles vues de face. Elle sont considérées comme les plus proches par notre système car leur segmentation est quasiment parfaite et permet de les décrire très fidèlement. Pour les autres vues, les limites des boîtes sont moins précises d'où de petites erreurs dans la description. Cet exemple est encore une illustration intéressante de l'intérêt d'une approche spatiale qui permet de caractériser les motifs des boîtes. Cette finesse de la description fait que les autres boîtes jaunes ne disposant pas des mêmes motifs sont considérées comme moins proches.*

FIG. 7.21 – Comparaison de recherches du 1er démonstrateur de FIDS avec notre système
À gauche : FIDS, *à droite : notre système.*

L'ensemble des résultats obtenus avec FIDS *sont très satisfaisants et il n'existe globalement que très peu de différences entre les deux systèmes. Sur certaines requêtes nous pouvons tout de même mettre en évidence l'apport d'une approche spatiale par rapport à une description globale. Sur le premier exemple, les images retrouvées par* FIDS *sont majoritairement composées d'arbustes alors que la scène requête montre des joggeurs. Même si pour notre approche une image ne correspond pas à cela, toutes les autres sont composées de routes bordées d'arbres sur lesquelles on trouve des personnes. De même, sur les deux requêtes suivantes, en proposant un environnement enneigé ou gelé,* FIDS *retourne à chaque fois une image montrant un ciel gris vu à travers des arbres.*

FIG. 7.22 – Comparaison de recherches du 2e démonstrateur de FIDS avec notre système
À gauche : FIDS*, à droite : notre système.*

De même que pour les exemples précédent, le système FIDS *fournit globalement des images très similaires à l'image recherchée. Sur le deuxième exemple, il est tout de même possible de noter la différence de couleur entre l'arbuste recherché et ceux renvoyés par le système. De plus, la dernière image ne correspond pas du tout alors que pour notre système les arbustes retrouvés sont tous d'une teinte colorée quasiment identique à celle de l'image originale. Pour les deux dernières requêtes, la demande de vues de geysers du parc de Yellowstone conduit* FIDS *à fournir des images de montagnes majoritairement blanches alors que l'image demandée est plutôt bleu-vert. Avec notre système même si certaines images n'appartiennent pas à la même catégorie, elles sont visuellement très proches.*

Angle de vue (en degrés)	Notre système	MNS
-90	42.6	36.2
-81	63.8	51
-72	74.5	48.9
-63	87.2	55.3
-54	93.6	59.6
-45	91.5	57.5
-36	93.6	42.6
-27	95.7	55.3
-18	93.6	70.2
-9	91.5	57.4
9	93.6	38.3
18	89.4	44.7
27	93.6	40.4
36	89.4	44.7
45	91.5	57.4
54	83	55.3
63	87.2	61.7
72	80.9	53.2
81	70.2	63.8
90	53.2	55.3
Moyenne (rang 1)	83	52.5
Moyenne ± 20	91.7	52.0
Moyenne ± 60	92.0	52.7
Moyenne (rang 1 à 3)	95.6	78
Moyenne ± 20	100	78.7
Moyenne ± 60	99.5	78.7

FIG. 7.23 – Comparaison quantitative des résultats obtenus avec SOIL-47

CHAPITRE 8

APPLICATION AUX FRESQUES MÉDIÉVALES

Sommaire

8.1 Introduction

Jusqu'ici les différentes méthodes développées durant cette thèse ont été illustrées avec diverses bases d'images même si elles étaient initialement conçues dans le cadre de notre partenariat avec le Centre d'Études Supérieures des Civilisations Médiévales de l'université de Poitiers (CESCM). Ainsi, le but final de cette étude est de fournir aux historiens des outils d'aide à la recherche et à la caractérisation d'objets au sein de leurs bases d'images.

Aujourd'hui, ils disposent d'un ensemble très vaste de clichés (plus de 150000) sous divers formats (papier, diapositives...) mais seulement quelques milliers sont numérisés et surtout indexés. La description utilisée pour leur indexation est essentiellement textuelle. Pour chaque image, en plus des données géographiques et topographiques du cliché, la scène est caractérisée par un ensemble de mots clefs choisis dans un thésaurus (celui du GAHOM[1]). Cette description étant limitative et subjective (cf. chapitre 3), notre approche de par son analyse directe du contenu des images est intéressante pour aller plus loin dans les recherches. Ainsi, l'association de paramètres objectifs sur la forme, la couleur et la texture nous permet de fournir automatiquement l'ensemble des images contenant les objets qu'ils manipulent usuellement.

Deux applications principales ont été finalisées durant ces trois années. La première est un système d'aide à la localisation d'images inconnues. Le seconde est liée à la problématique de recherche d'objets et permet de retrouver les objets visuellement proches présents dans l'ensemble des images étudiées.

[1] Groupe d'anthropologie historique de l'Occident médiéval.

8.2 Recherche globale pour la détection de site

8.2.1 Méthode employée

Le but de cette application est de déterminer la provenance géographique d'une image inconnue. Cette application « grand public » est un peu en marge des objectifs initiaux fixés avec le CESCM. En effet, elle n'a pas été mise en place sur la demande des historiens de l'art mais elle nous a surtout permis de développer un démonstrateur simple et convivial. De plus, nous avons ainsi pu tester la fiabilité de nos méthodes sur une application simple. Ce système pourra également être utilisé au sein de nombreux autres projets liés au patrimoine de la région Poitou-Charentes pour mettre en évidence de manière « pédagogique » les liens entre les différents édifices anciens.

Pour chaque site répertorié nous disposons d'une vingtaine d'images caractéristiques constituant la base de référence. L'ensemble des images est donc fonction du nombre de sites définis. L'image inconnue est alors recherchée au sein de cette base grâce à la mise en concurrence de deux méthodes. La première méthode utilise une comparaison globale de la couleur des images alors que la seconde est fondée sur le système de comparaison d'images pour les bases généralistes (cf. section 7.3.1). Pour la première, seule la couleur est utilisée alors qu'au sein de la seconde la texture et la forme complètent la description des régions.

Les deux types de requêtes sont alors calculés. Pour chacune d'elles, le site détecté est le site majoritaire dans les 5 premières images trouvées. Si plusieurs sites ont la même occurrence, ils sont considérés comme tous probables. Deux cas sont ensuite possibles :

- les deux méthodes amènent au même site. Nous pouvons donc émettre l'hypothèse sur ce lieu avec un indice de confiance tout en prenant garde à afficher les images correspondantes pour permettre à l'expert de vérifier notre hypothèse ;
- les lieux déterminés par les deux méthodes sont contradictoires. Aucune hypothèse ne peut être formulée. Nous affichons tout de même les images similaires que nous avons trouvées pour que l'expert puisse déterminer visuellement le site exact.

8.2.2 Résultats

Globalement avec les deux méthodes choisies nous obtenons un pourcentage de mauvaise classification de 7%. Il y a également 10% d'images pour lesquelles nous ne pouvons prendre une décision et pour les 83% restantes la recherche fournit une localisation exacte du site. Pour les mêmes descripteurs utilisés seuls, les pourcentages de bonnes détections sont de l'ordre de 85% mais il semble hasardeux d'émettre une hypothèse avec une seule source d'information. L'utilisation combinée des deux méthodes réduit ainsi légèrement le pourcentage de bonne détection mais réduit aussi significativement celui de fausse détection (divisé par 2).

Les figures [8.1], [8.2] et [8.3] présentent des exemples des trois cas possibles. Pour chacun d'eux nous pouvons remarquer que les images similaires retrouvées sont toujours présentées pour que l'expert puisse vérifier visuellement le résultat fournit.

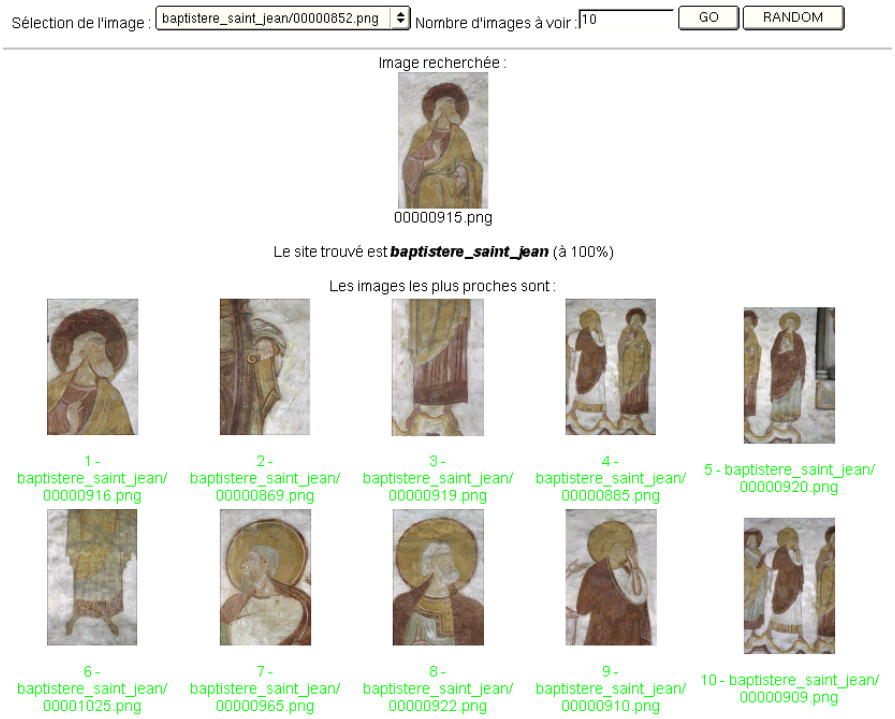


FIG. 8.1 – Application de localisation d'images : résultat pertinent

Sur cet exemple, l'image recherchée ne pose aucun problème, les images les plus proches contiennent même des parties de l'objet proposé. La décision peut donc être prise avec certitude sur sa localisation géographique.

8.3 Recherche partielle pour rechercher des objets - Application aux apôtres

Pour montrer l'intérêt de notre système de recherche d'objets, nous nous sommes focalisés sur la recherche d'apôtres aux sein des images de fresques médiévales que nous a fournies le CESCM. Ce choix est motivé principalement par le nombre important de vues d'apôtres dont nous disposions. Une étude sur les instruments de musique ou les animaux a également été discutée mais le nombre de clichés disponibles était beaucoup trop faible pour être significatif. De même, il est envisagé d'étendre ce système pour s'attaquer aux problèmes plus complexes que sont les drapés des toges ou encore les mains.

La méthode employée est celle présentée à la section 7.4 développée pour la recherche d'objets. Les statistiques prises en compte sont la couleur, la forme et la texture. L'aspect texture étant peu pertinent sur ce type d'images, il a été pondéré très faiblement alors que la forme et la couleur ont une importance similaire. La combinaison des statistiques est donc $\frac{2(\text{couleur}+\text{forme})+\text{texture}}{5}$ avec chacun des trois descripteurs ayant des distributions semblables.

FIG. 8.2 – Application de localisation d'images : mauvaise décision

Pour cette image d'une fresque d'Ebreuil, le site détecté est le baptistère Saint-Jean de Poitiers avec une certitude de 50%. Le pourcentage montre déjà que nous ne sommes pas vraiment sûr de notre décision. En fait, cette erreur est principalement due au fait qu'Ebreuil est un site caractérisée par la présence de pigment bleu. Or, sur la photo présentée, celui-ci est très peu visible. De plus, une grande partie de la fresque est effacée entraînant une couleur dominante correspondant à celle du mur. L'étude des images proposée permet tout de même à l'expert de mettre en cause notre diagnostic grâce aux images présentées comme semblables.

8.3.1 Résultats

Les figures [8.4], [8.5] et [8.6] présentent diverses requêtes effectuées sur l'ensemble des images que nous a fourni le CESCM, soit un total de 463 clichés. Pour chaque exemple, nous proposons les résultats de la requête globale ne prenant en compte que l'aspect couleur, puis une requête utilisant l'aspect pyramidal et la décomposition et finalement la version finale du système de recherche d'objets intégrant la prise en compte de l'aspect spatial par un algorithme d'appariement de graphes attribués.

L'ensemble des résultats présentés montre bien l'apport de notre structure de représentation pour ces images. En effet, sur chacune de ces requêtes, les résultats sont plus pertinents en utilisant la pyramide et encore plus cohérents quand nous utilisons la contrainte spatiale. Les différents résultats montrent ainsi que l'aspect multi-échelles permet de retrouver des objets vus sous différents points de vue. Cette approche permet même d'obtenir des parties des objets recherchés. Les gros plans sur les apôtres sont alors retournés principalement quand la contrainte spatiale est utilisée.

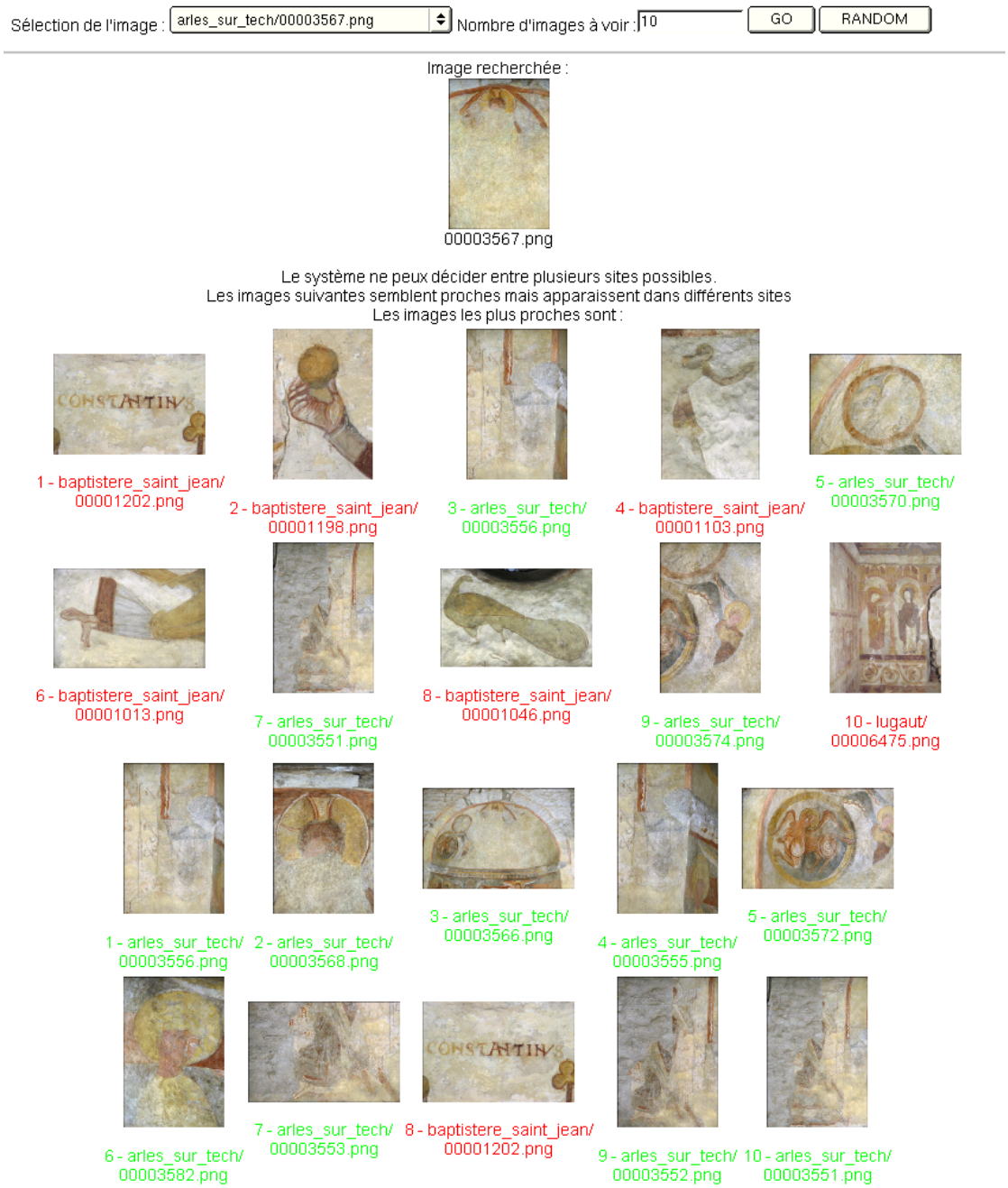


FIG. 8.3 – Application de localisation d'images : décision non prise

Sur cet exemple, aucune décision n'est prise quant à la localisation géographique de la fresque. La première méthode nous oriente vers le baptistère Saint-Jean alors que la deuxième fournit la bonne localisation qui est Arles-sur-Tech. Cette requête est fortement biaisée à cause de l'effacement de la quasi-totalité de la fresque. L'étude des images ressemblantes permet alors de retrouver une partie de cette image dans les images similaires (2^e image de la deuxième série) ce qui permet alors à l'expert d'être certain de sa provenance. Dans ce cas, la recherche serait fortement améliorée si l'image proposée au système de recherche se focalisait sur la zone non effacée.

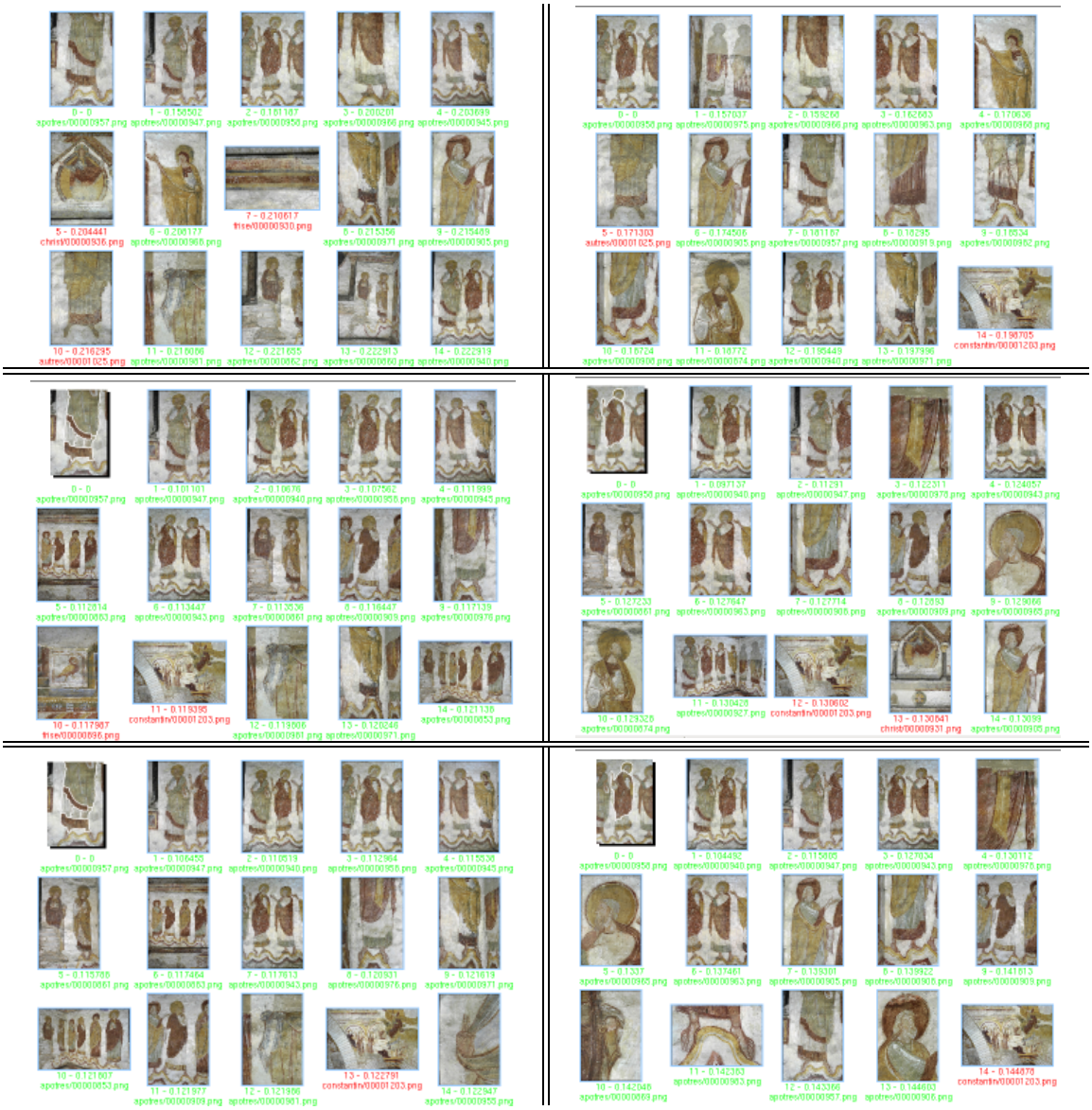


FIG. 8.4 – Quelques recherches d'apôtres au sein de la base du CESCM (1/3)
En haut : recherche globale, au milieu : utilisation de l'aspect pyramidal et de la décomposition, en bas : utilisation des aspects pyramidal et spatial.
Les images sont triées de gauche à droite et de haut en bas.

Pour la recherche de la toge à gauche, il est à noter que la recherche globale fournit des résultats satisfaisants en donnant 2 autres vues de l'objet requête. Les différences entre l'utilisation du graphe pyramidal avec et sans contrainte spatiale sont faibles. Pour les deux cas, les 3 vues de l'objet sont retrouvées. Les images suivantes contiennent également des toges de même type (bandeau coloré, partie blanche et haut coloré) mais de couleurs différentes. Ici, l'apport de la contrainte spatiale est très faible même si elle permet d'exclure la vue de la frise comportant un oiseau. Pour l'exemple de droite par contre, la recherche globale ne fournit que deux autres vues de l'apôtre recherché. Dans ce cas, la contrainte spatiale n'améliore aussi que très peu la recherche. En effet, 4 vues de l'objet sont retrouvées dans les deux types de recherche. Cette information permet tout de même de classer 3 vues en premières positions au lieu de 2. Par contre, elle rejette une vue de l'apôtre à une échelle différente qui n'apparaît plus dans les résultats (10^e image). Cela est principalement dû à une mauvaise segmentation de l'apôtre sur ce cliché.

FIG. 8.5 – Quelques recherches d'apôtres au sein de la base du CESCM (2/3)
En haut : recherche globale, au milieu : utilisation de l'aspect pyramidal et de la décomposition, en bas : utilisation des aspects pyramidal et spatial. Les images sont triées de gauche à droite et de haut en bas.

Sur ces deux exemples, l'intérêt de la contrainte spatiale est plus visible. En effet, la recherche globale fournit toujours des images ressemblantes mais peu cohérentes par rapport à l'objet qui est recherché. Dans les deux cas, un algorithme d'appariement de graphes attribués permet d'augmenter le nombre de vues retournées de l'objet d'intérêt. De plus, elles sont mieux classées, ce qui semble logique du fait que la contrainte spatiale permet d'augmenter les mesures de similarité pour les images pour lesquelles la projection de l'objet est désordonnée. Pour la recherche de droite, il est également à noter que des vues très différentes de l'apôtre sont retournées : 2 gros plans de son buste, 3 clichés à la même échelle et 1 plan large sont ainsi retrouvés.

FIG. 8.6 – Quelques recherches d'apôtres au sein de la base du CESCM (3/3)
En haut : recherche globale, au milieu : utilisation de l'aspect pyramidal et de la décomposition, en bas : utilisation des aspects pyramidal et spatial. Les images sont triées de gauche à droite et de haut en bas.
Sur l'exemple de gauche, nous avons tenté de rechercher 2 objets disjoints en même temps. Dans ce cas, les résultats sont peu probants car il n'existe que très peu d'images où les deux objets sont présents simultanément. La requête de droite est similaire au deux de la figure précédente. La recherche globale n'est pas adaptée et la contrainte spatiale permet d'affiner la recherche en rejetant certaines images non pertinentes ce qui fait apparaître d'autres vues de l'apôtre recherché ainsi que des apôtres similaires (auréole orangée et toge marron).

8.4 Conclusion

Dans ce chapitre, nous avons présenté les résultats des méthodes mises en place durant cette thèse appliquées sur les images de fresques médiévales du CESCM. Les deux applications développées ont ainsi été décrites. La première nous a permis de tester la fiabilité de nos techniques sur ce type d'images tout en créant un démonstrateur simple. De plus, celui-ci n'est pas limité aux fresques médiévales et peut être utilisé à plus vaste échelle ainsi que dans d'autres domaines.

La seconde application est quant à elle celle qui apporte le plus aux historiens de l'art en leur permettant de rechercher des objets au sein de leurs bases d'images. Elle permet ainsi de retrouver des éléments qui n'auraient pas été décrits textuellement en utilisant simplement le contenu des images. Nous avons ainsi montré encore une fois l'intérêt de la structure proposée pour décrire précisément les images et les objets qu'elles contiennent.

Bien que les résultats de cette étude soient déjà très intéressants, de nombreuses voies peuvent encore être explorées pour améliorer le système. La première, et non la moindre, consisterait à développer un algorithme de segmentation multi-échelles spécifique à cette problématique. Cela permettrait ainsi d'obtenir une description plus pertinente du contenu des fresques en termes d'objets. De plus, les statistiques employées sont celles qui ont été déterminées comme les plus performantes dans un cadre général mais il serait sûrement possible d'affiner ce choix pour cette application précise.

CHAPITRE 9

CONCLUSION ET PERSPECTIVES

Tout au long de ce manuscrit nous avons présenté les divers volets du travail réalisé pendant les trois années de ce doctorat. Les questions abordées ont été très diverses : de la perception des images couleur, à la recherche d'une représentation haut-niveau pour l'indexation d'images en passant par les outils permettant de séparer les objets présents au sein d'une image.

Ce manuscrit commence par une présentation des divers phénomènes intervenant dans le processus de la vision humaine. En effet, il nous paraît important de comprendre comment nous percevons le monde qui nous entoure et en particulier des images pour tenter d'imiter ces mécanismes dans le cadre de la vision par ordinateur. Ce premier chapitre met ainsi en évidence les aspects encore mystérieux de la perception humaine tout en exposant les solutions mises en place actuellement pour acquérir et reproduire des images. La présentation des espaces de représentation de la couleur est l'exemple le plus frappant montrant les limitations et les problèmes liés à de telles méthodes.

Le deuxième chapitre rentre ensuite dans le problème de notre étude et propose un état de l'art des méthodes d'indexation d'images par le contenu. Une fois définis les multiples problèmes de ce domaine, il apparaît évident qu'il est nécessaire de traiter l'image brute pour réduire le volume d'information à analyser tout en gardant une représentation la plus fidèle possible. Les premières approches développées en indexation d'images calculaient ainsi des statistiques globales aux images pour les caractériser (par leur couleur ou leur texture par exemple). Depuis une dizaine d'années, des modes de représentation plus évolués sont employés. Ces méthodes dites « spatiales » prennent en compte l'arrangement des différents objets au sein des images. L'utilisation de cette information supplémentaire permet alors une caractérisation beaucoup plus fine des images même si l'aspect de composition des objets est encore inutilisé.

Ayant montré la nécessité d'une représentation des images pour l'indexation, le troisième chapitre s'intéresse aux différentes structures généralement employées en traitement et analyse d'images. Les premières structures étaient planes ce qui est très limitatif pour résoudre certains problèmes. La notion de multi-résolutions ou de multi-échelles apparaît alors dans les années 1980 avec les pyramides gaussiennes et laplaciennes. Les évolutions vers les pyramides régulières et surtout irrégulières nous amènent à proposer l'utilisation d'une nouvelle structure de représentation d'images, le graphe pyramidal. Celui-ci est très similaire aux pyramides irrégulières car composé de différents niveaux caractérisés par des graphes d'adjacences. Ces niveaux sont alors reliés pour symboliser la notion de composition des différents objets. Cette nouvelle structure permet ainsi de caractériser les objets présents dans l'image par leur couleur, leur texture et leur forme mais également par leurs parties et sous-parties. Sans aller jusqu'à une description sémantique des images qu'il paraît irréaliste d'extraire, nous essayons de cette manière d'obtenir une représentation la plus fine possible des images. Cette nouvelle approche nécessite alors deux éléments indispen-

sables pour être efficace : un outil de segmentation multi-échelles orienté objets et des méthodes de caractérisation performantes des régions. Ces deux aspects font alors l'objet des deux chapitres suivants de ce manuscrit.

La génération du graphe pyramidal nécessite de disposer de plusieurs partitions plus ou moins fines de l'image. De plus, celles-ci doivent être le plus possible liées aux objets inclus dans la scène. Au sein du quatrième chapitre, nous avons par conséquent présenté rapidement les diverses techniques de segmentation multi-échelles disponibles à l'heure actuelle puis nous nous sommes intéressés plus précisément à la technique retenue qui est fondée sur l'algorithme JSEG. Celui-ci utilise un critère de rupture de modèle textural pour localiser les frontières des différents objets. L'implantation retenue pour cet algorithme nous permet d'obtenir un ensemble de partitions de plus en plus fines de l'image qui sont toutes cohérentes entre elles. Le graphe pyramidal de l'image peut alors être construit à partir de ces segmentations.

Une fois la structure du graphe pyramidal établie, il nous faut encore caractériser les différents objets de la scène par leur couleur, leur texture et leur forme. Le cinquième chapitre a donc présenté ces trois points indépendamment et a cherché à déterminer les statistiques les plus adaptées à notre structure. Pour l'aspect couleur, notre étude nous a amené à proposer une évolution des métriques couleur standards fondée sur l'idée qu'il n'est pas pertinent de quantifier la distance entre deux couleurs complètement distinctes. L'application d'un noyau gaussien aux distances standards permet ainsi de fixer un seuil pour lequel les distances sont utilisables. Ne disposant généralement pas d'images à même support colorimétrique, nous proposons également une adaptation de l'algorithme d'intersection d'histogrammes pour ce type d'images. Les résultats obtenus pour ces deux nouvelles techniques sont très intéressants par rapport aux métriques généralement utilisées telles que l'*Earth Mover's Distance*. Pour cette méthode nous proposons d'ailleurs une implantation alternative réduisant fortement le coût calculatoire par rapport à la version originale de RUBNER. Cela rend alors possible l'utilisation d'une telle distance pour des distributions de cardinalité élevée. Pour la forme et la texture, l'étude a été beaucoup moins poussée. Pour ces deux aspects, un état de l'art est présenté ainsi qu'un comparatif des résultats expérimentaux utiles à la détermination des statistiques les plus adaptées à notre structure.

Les deux derniers chapitres s'intéressent finalement à l'utilisation du graphe pyramidal en indexation d'images par le contenu. Nous définissons alors trois modules indépendants pour comparer les structures. Le premier correspond à la comparaison des régions entre elles dont l'étude a été réalisée au chapitre précédent. Nous montrons ensuite l'intérêt de l'aspect spatial de la représentation pour des recherches globales d'images. La conclusion de cette étude est que l'image est caractérisée plus finement par les statistiques locales portées par l'ensemble de ses régions que par les mêmes descripteurs globaux. L'aspect pyramidal permet ensuite de détecter des objets présents à différentes échelles au sein des images. Les diverses techniques mises en place pour des requêtes globales peuvent également être appliquées pour rechercher des objets. Dans ce cas, nous affinons la mesure de similarité entre les niveaux pour prendre en compte plus fidèlement l'arrangement spatial des régions formant l'objet. Les deux algorithmes d'appariement de graphes attribués étudiés rendent le positionnement des régions trouvées plus cohérent. Nous comparons également notre système avec quelques systèmes dont nous avons pu nous procurer des résultats. De manière qualitative, nos requêtes retournent des images généralement plus pertinentes. L'étude quantitative menée avec le système SOIL-47 justifie également l'apport significatif de notre méthode.

Le dernier chapitre expose enfin les deux applications que nous avons pu finaliser durant ce doctorat dans le cadre de notre partenariat avec le CESCM. La première qui est sûrement la plus simple fournit une solution pour déterminer la localisation d'une image d'origine géographique inconnue. Elle présente l'intérêt de pouvoir être intégrée dans des applications ludo-éducatives de type grand public. La seconde est quant à elle le système pour lequel toute cette étude a été menée. Elle consiste en un système de recherche et de caractérisation d'objets. Les requêtes présentées pour illustrer cette partie sont focalisées sur la recherche d'apôtres au sein des fresques médiévales.

Ce travail prospectif fournit des résultats très prometteurs. De nombreuses voies d'études sont encore à explorer. La premier sujet d'étude qui est sûrement le plus important pour améliorer la robustesse du système est de définir un outil de segmentation multi-échelles orienté objets plus performant. En effet, l'algorithme présenté dans ce manuscrit n'est pas parfait. Les segmentations obtenues sont parfois peu pertinentes par rapport au contenu des images. Une voie possible d'étude serait de faire évoluer le critère de rupture de texture défini par DENG en intégrant un facteur lié à la perception colorée de la texture.

Au sein du graphe pyramidal même, une étude plus complète des descripteurs de texture et de forme serait nécessaire. Pour la forme, il serait intéressant de disposer d'une statistique additive évitant alors de la recalculer pour toutes les régions des différents niveaux. Pour la texture, l'étude réalisée a été très succincte et nécessiterait d'être approfondie, en particulier au niveau des tests sur les différents descripteurs existants. De plus, comme nous l'avons déjà évoqué au chapitre 4, l'utilisation d'une description des niveaux par des cartes généralisées rendrait la description topologique complète contrairement aux graphes d'adjacences.

Nous voyons également deux approches permettant d'améliorer la qualité et la vitesse de la comparaison des graphes pyramidaux. Les méthodes utilisées actuellement pour cette étape n'effectuent pas réellement une comparaison des pyramides dans leur ensemble mais considèrent uniquement les similarités entre les différentes échelles. Il serait intéressant de faire évoluer les algorithmes d'appariement de graphes attribués en intégrant la notion d'échelle et ainsi mettre en correspondance les pyramides entières. D'autre part, les critères de rejet rapide liés à la structure sont déjà utilisés pour réduire les temps de recherche. Mais, d'autres voies sont également à explorer telle que l'indexation directe des régions qui permettrait d'exclure directement les images n'ayant aucune région ressemblante à celles proposées dans la requête.

L'adaptation de cette méthode de représentation à la vidéo semble aussi une perspective séduisante. Ainsi, le suivi des objets au cours des images permettrait de réduire le volume utile à la description des scènes et rendrait aussi possible la localisation des changements de plans difficilement détectables tels que les fondus.

Enfin, une application spécifique devrait voir le jour d'ici peu dans le cadre du partenariat avec le CESCM. Celle-ci sera focalisée sur le problèmes de la représentation des mains au sein des fresques médiévales. Une détection semi-automatique des caractéristiques définissant la main devrait être réalisée pour permettre ensuite la recherche des mains ressemblantes à partir d'un exemple ou d'une esquisse. Cette étude pourrait être réalisée avec le système défini dans ce manuscrit mais, au vu des caractéristiques utilisées pour caractériser une main, une méthode spécifique semble plus adaptée.

En attendant, le travail réalisé a permis aux experts de se poser de nouvelles questions relatives aux couleurs, aux textures et aux formes de base des différents éléments constitutifs des fresques des bâtiments médiévaux de la région Poitou-Charentes.

ANNEXE A

ALGORITHMES LIÉS À JSEG

Algorithme 4 : Algorithme général JSEG

Données : $I =$ Image originale
Entrée : $nbc =$ Nombre de couleurs à utiliser
$N_{init} =$ Niveau initial
$N_{fin} =$ Niveau final
Résultat : $\{Seg[i], \forall i \in [N_{fin} \dots N_{init}]\}$= ensemble des segmentations obtenues

Initialisation :
$Vois[i] = 2^{i+1} + 1, \forall i \in [N_{fin} \dots N_{init}]$: taille du voisinage ;
$Echant[i] = 2^{i-1}, \forall i \in [N_{fin} \dots N_{init}]$: pas d'échantillonnage ;
$TailleValleeMin[i] = Vois[i]^2, \forall i \in [N_{fin} \dots N_{init}]$: Taille minimum des vallées ;
$Seg_{courante} =$ Segmentation courante initialisée comme une région de la taille de l'image ;
$Niv_{courant} =$ Niveau courant initialisé à N_{init} ;

début
Quantification en nbc couleurs ;
↪ Image quantifiée
tant que $Niv_{courant} \neq N_{fin}$ **faire**
Calcul_Carte_de_J(I,$Vois[Niv_{courant}]$,$Echant[Niv_{courant}]$) ;
↪ Carte_des_J
`Détection_des_Vallées`(Carte_des_J,$Seg_{courante}$,$TailleValleeMin[i]$) ;
↪ Carte_des_Vallées
`Accroissement_des_Vallées`(Carte_des_Vallées,Carte_des_J) ;
↪ Carte_des_Vallées
Regroupement des régions similaires de la carte des vallées - facultatif ;
$Seg[Niv_{courant}] =$ Carte_des_Vallées ;
$Seg_{courante} = Seg[Niv_{courant}]$;
$Niv_{courant} = Niv_{courant} - 1$;
fin
Repositionnement des segmentations $Seg[i], \forall i \in [N_{fin} + 1 \dots N_{init}]$;
fin

Fonction `Détection_des_Vallées(`*Carte_des_J*,$Seg_{courante}$,$TailleValleeMin$`)`

Entrée : $Carte_J =$ Carte des J
$Seg_{courante} =$ Segmentation du niveau supérieur
$TailleValleeMin =$ Taille minimum des vallées détectées
Résultat : Carte_des_Vallées

Initialisation :
$NombreVallees[region] = 0, \forall region$ de $Seg_{courante}$: Nombre de vallées détectées

début
  pour *toutes les régions de* $Seg_{courante}$ **faire**
    Calcul de μ_J et σ_J, respectivement la moyenne et la variance des valeurs de J ;
    Détermination des seuils $Seuil[i] = \mu_J + A * \sigma_J$
      où $A = \{-0.6; -0.4; -0.2; 0; 0.2; 0.4\}$;
  fin
  pour *tous les* $Seuil[i]$ **faire**
    Seuillage de Carte_des_J par rapport à $Seuil[i]$ défini sur chaque région ;
      $\hookrightarrow$ Carte_Vallées_temp
    Connexification des régions de Carte_Vallées_temp ;
      $\hookrightarrow$ Carte_Vallées_temp
    Suppression des vallées de tailles inférieures à $TailleValleeMin$;
      $\hookrightarrow$ Carte_Vallées_temp
    pour *toutes les régions de* $Seg_{courante}$ **faire**
      $NvNbVal =$ Nombre de vallées de la région dans Carte_Vallées_temp ;
      si $NvNbVal > NombreVallees[region]$ **alors**
        Copie de la région de Carte_Vallées_temp vers Carte_des_Vallées ;
      fin
    fin
  fin
fin

Fonction `Accroissement_des_Vallées(`*Carte_des_Vallées,Carte_des_J*`)`

Entrée : Carte_des_J : Carte des J
Carte_des_Vallées : Carte des vallées détectées
Résultat : Carte_des_Vallées

Initialisation :
Définition d'une relation de voisinage : Voisins : pixel→{pixels voisins à pixel} ;
NA = {pixels de Carte_des_Vallées non affectés à une vallée} ;
NA_a = {pixels adjacents aux vallées existantes dans Carte_des_Vallées}∩NA ;
NA_i = {pixels non adjacents aux vallées existantes dans Carte_des_Vallées}∩NA ;

début
  tant que NA *non vide* **faire**
    Recherche du pixel à affecter :
      Pixel_à_Affecter=p / p ∈ NA_a et $J[p] = min\{J[i] \text{ / } i \in Na_a\}$;
    Affectation de Pixel_à_affecter à sa région adjacente ;
    $NA = NA-$Pixel_à_Affecter ;
    Voisins_à_Ajouter=NA_i∩Voisins(Pixel_à_Affecter) ;
    NA_i=NA_i-Voisins_à_Ajouter ;
    NA_a=NA_a+Voisins_à_Ajouter ;
  fin
fin

ANNEXE B

BASES D'IMAGES

Sommaire

B.1 Base d'images naturelles de l'université de Washington

La base de l'université de Washington[1] aussi appelée *UW-groundtruth* est composée de 1224 photos de « vacances » prises dans 21 endroits différents (cf. tableau [B.1]). Certaines localisations sont aussi accompagnées d'une description des photos par des mots clefs (*house, cloud, tree, bush*...) Vu l'architecture de la base, des images complètement différentes peuvent se retrouvées dans la même catégorie (cf. figure [B.2]). Les recherches de toutes les images d'une même catégorie sont donc très difficiles.

La base SIMPLICITY[2] est du même type mais comporte beaucoup plus d'images. Trois versions sont disponibles contenant 1000, 10000 ou 60000 images.

B.2 Base d'images naturelles de PICTOSEEK

Cette base d'images naturelles est utilisées pour les démonstrateurs des systèmes d'indexation d'images par le contenu du laboratoire ISIS[3] de l'université d'Amsterdam. Elle se compose de 764 images représentant principalement des animaux (ours, éléphants, lions...), des voitures et des couchers de soleil (cf. figure [B.3]).

[1] téléchargeable à l'adresse http://www.cs.washington.edu/research/imagedatabase/groundtruth/.
[2] téléchargeable à l'adresse http://wang1.ist.psu.edu/docs/related/.
[3] *Intelligent Sensory Information Systems*.

Localisation	Nombre d'images	Fichier de description
arborgreens	47	Oui
australia	30	Oui
barcelona	48	Non
barcelona2	115	Non
campusinfall	48	Oui
cannonbeach	48	Oui
cherries	55	Oui
columbiagorge	83	Oui
football	48	Oui
geneva	25	Oui
greenlake	48	Oui
greenland	255	Non
indonesia	36	Oui
iran	49	Oui
italy	22	Oui
japan	45	Oui
leaflesstrees	48	Non
sanjuans	48	Oui
springflowers	48	Oui
swissmountains	30	Oui
yellowstone	48	Oui

FIG. B.1 – Composition de la base de l'université de Washington

FIG. B.2 – Images de la base de l'université Washington tirées de la catégorie *italy*

FIG. B.3 – Images de la base de référence de PICTOSEEK

FIG. B.4 – Images de la base de l'université Columbia, COIL-100
Deux lignes supérieures : différents objets.
Dernière ligne : différentes vues d'un même objet.

B.3 Base d'objets artificiels de l'université Columbia

Initialement composée de 20 objets en niveaux de gris, la base COIL-20 a évolué vers une version couleur comprenant 100 objets différents : COIL-100[4] (COIL pour *Columbia University Image Library*). Ces objets sont vus sous 72 points de vue différents. Certains d'entre eux peuvent ne différer que par la couleur ou la forme. Il y a par exemple des bateaux en plastique vert, bleu et jaune (cf. figure [B.4]).

B.4 Base d'objets artificiels SOIL-47

La base SOIL-47[5] est du même type que la base Columbia. Elle apporte tout de même des variations intéressantes. Deux illuminations sont proposées pour chaque objet et des vues comportant de multiples objets sont également proposées (avec en particulier des occultations). Les diverses prises de vue d'un objet sont réalisées avec un distance à l'objet variable ce qui engendre des variations de la taille de l'objet. De plus, les objets composant la base sont très similaires (des boîtes de céréales en majorité) (cf. figure [B.5]). Elle propose donc 47 objets sous 21 points de vue et 2 illuminations différents ; soit 1974 images au total. 22 scènes sont aussi proposées avec de 2 ou 3 objets différents sous 21 points de vue.

[4] téléchargeable à l'adresse http://www1.cs.columbia.edu/CAVE/research/softlib/.
[5] téléchargeable à l'adresse http://www.ee.surrey.ac.uk/Research/VSSP/demos/colour/soil47/.

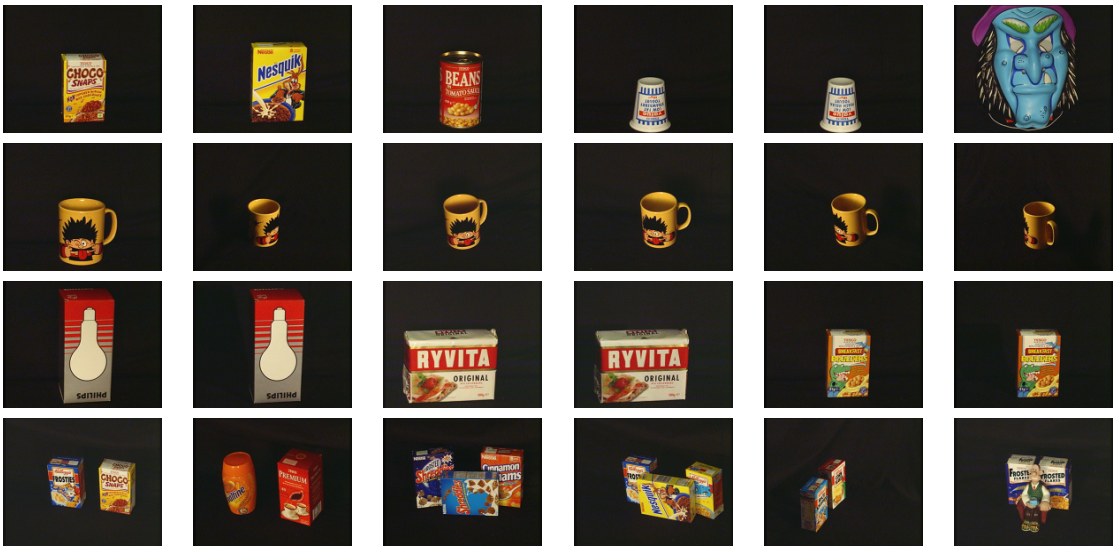


FIG. B.5 – Images de la base SOIL-47
Première ligne : différents objets.
Deuxième ligne : un objet sous différents points de vue.
Troisième ligne : des illuminations différentes d'un même objet.
Quatrième ligne : des images avec de multiples objets.

B.5 Base de formes SIID

SIID est une base de données de formes. Des versions comprenant 88 ou 216 objets sont disponibles[6]. La version comprenant 216 formes peut être divisée en 18 catégories comprenant des variations d'un même objet (cf. figure [B.6]).

B.6 Base de textures VISTEX

VISTEX est similaires à SIID mais s'applique aux textures. Elle est composée de 167 textures de référence divisées en 19 catégories[7]. 11 scènes sont aussi proposées. Pour chacune d'elles, une ou deux images globales sont disponibles ainsi que des zones texturées issues de celles-ci. Toutes ces textures sont en couleur. Les bases MEASTEX[8] et BRODATZ[9] fournissent quant à elles des textures en niveaux de gris.

[6] téléchargeables à l'adresse http://www.lems.brown.edu/vision/researchAreas/SIID/.
[7] téléchargeable à l'adresse http://whitechapel.media.mit.edu/vismod/imagery/VisionTexture/.
[8] téléchargeable à l'adresse http://www.cssip.uq.edu.au/staff/meastex/meastex.html.
[9] téléchargeable à l'adresse http://sipi.usc.edu/services/database/Database.html.

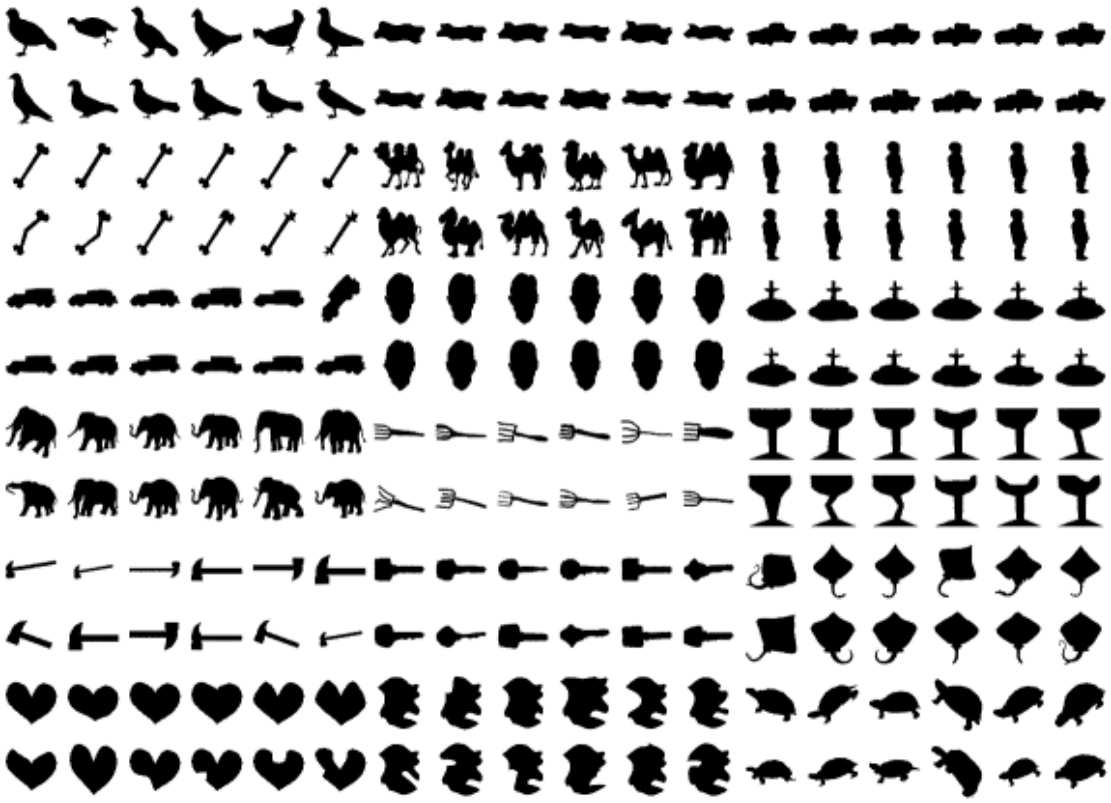


FIG. B.6 – Base de données de formes SIID

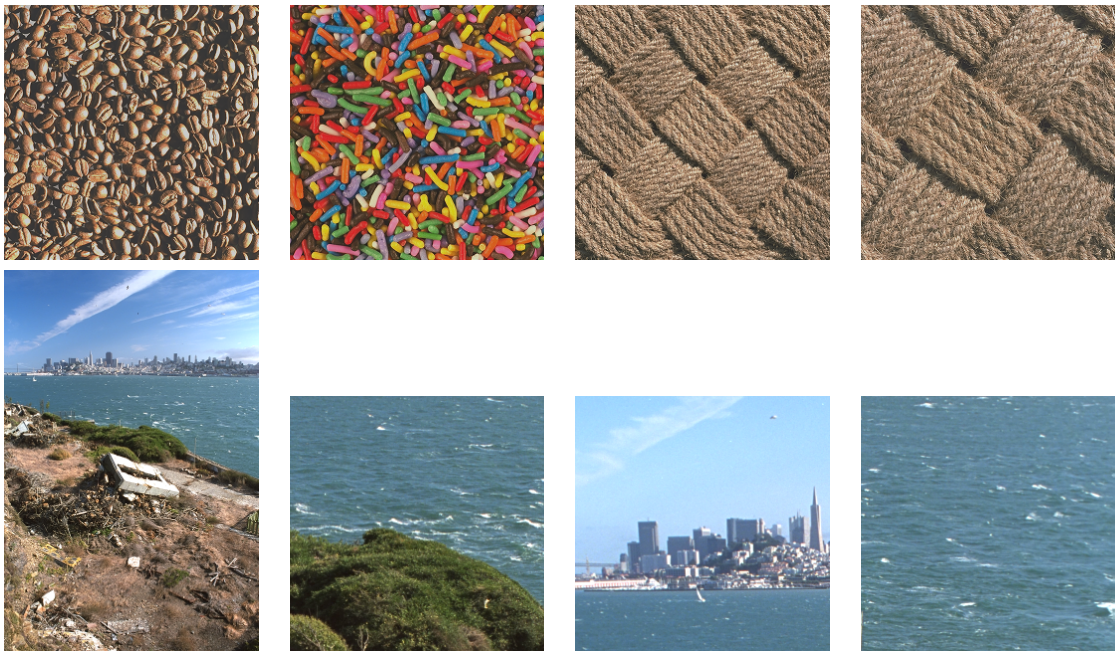


FIG. B.7 – Images de la base de textures du VISTEX
En haut : des textures de référence (2 catégories différentes).
En bas : une scène et des textures extraites de celle-ci.

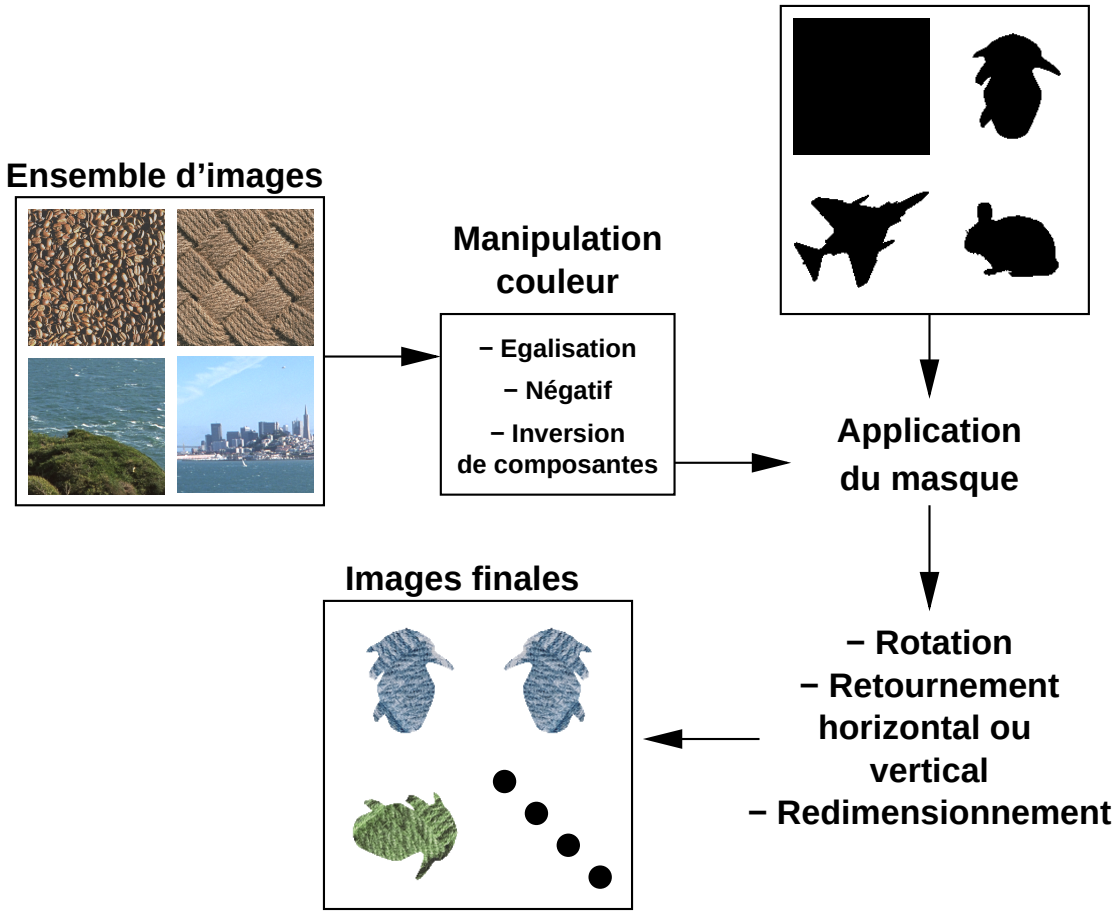


FIG. B.8 – Principe de génération de notre base de test

B.7 Base locale de textures, formes et couleurs

Cette base a été mise en place pour tester les diverses méthodes de comparaison de régions développées. À partir de i images et m masques, un ensemble d'images est alors généré (cf. figure [B.8]). Pour chaque image, c manipulations des couleurs différentes peuvent être effectuées (égalisation, négatif, inversion de composantes...). Les divers masques sont alors appliqués sur les images produites. Il est ensuite possible de réaliser r rotations ou retournement différents et d redimensionnements. La base générée contient finalement $i * m * c * r * d$ images. Elles peuvent enfin être classées de multiples manières suivant les tests à réaliser en fonction de leur forme, leur couleur ou leur texture.

B.8 Base d'images de fresques médiévales du CESCM

Le Centre d'Études Supérieures des Civilisations Médiévales (CESCM) de l'université de Poitiers dispose d'un fond iconographique consacré à l'art roman (du IX^{e} au XXI^{e} siècle) en France et à l'étranger (architecture, sculpture, peintures murales, inscriptions, vitraux, manuscrits, objets

FIG. B.9 – Images de la base du CESCM

d'art...). Sur divers supports, il est composé d'environ 300000 clichés noir et blanc ou couleur dont 3000 ont déjà été numérisés (cf. figure [B.9]). En plus de la numérisation, 8500 fiches descriptives sont également disponibles. Elles apportent de nombreuses informations textuelles sur les images telles que leurs localisations et les éléments les composant.

Pour chaque édifice photographié des clichés à des niveaux différents sont effectués. Des gros plans puis des vues de plus en plus rapprochées sont réalisées. Les différentes photos de détails peuvent ainsi être replacées dans leur contexte au sein de l'édifice global.

Nous avons essentiellement travaillé sur les édifices suivants :
- le baptistère Saint-Jean de Poitiers bâti vers 360 et reconstruit au VII^{e} siècle. Les peintures murales les plus anciennes datent du XI^{e} siècle ;
- l'abbaye bénédictine Saint-Léger d'Ebreuil dans l'Allier fondée vers 900 ;
- la chapelle Saint-Jean de Lugaut dans les Landes possédant des fresques du XIII^{e} siècle ;
- l'abbaye d'Arles-sur-Tech dans les Pyrénées Orientals ornées de fresques du XII^{e} siècle.

La caractéristique principale de ces images de fresques est leur forte altération. En effet, le temps peut engendrer des dégradations telles que la dépigmentation des peintures, l'apparition d'effacement partiels ou complets d'objets.

ANNEXE C

BIBLIOGRAPHIE DE L'AUTEUR

C.1 Revues

AT2002 « New use of spatial information for content-based image indexing »
J. Dombre, N. Richard et C. Fernandez-Maloigne
Annals of Telecommunications, Septembre/Octobre 2002 Vol. 57 Num. 9-10

EGC2003 « Extraction de connaissances à partie de bases d'images - Deux applications directrices »
N. Richard, M.C. Larabi, J. Dombre et C. Fernandez-Maloigne
Extraction et gestion des connaissances, RTSI série RIA-ECA Vol.17 Num. 1-2-3 Janvier 2003

C.2 Conférences internationales avec actes et comité de lecture

SPIE2002 « Content-Based Image Retrieval and high-level representations »
J. Dombre, N. Richard et C. Fernandez-Maloigne
IST/SPIE Internet Imaging III, San Jose, USA, January 20-25, 2002, pp. 134-143

CGIV2002 « A new representation system for content-based image retrieval : The Pyramidal Graph »
J. Dombre, N. Richard et C. Fernandez-Maloigne
Conference on Color in Graphics, Image and Vision, Poitiers, France, April 2-5, 2002, pp. 384-389

C.3 Conférences nationales avec actes et comité de lecture

RFIA2004 « Connaissances et Images : un formalisme commun »
N. Richard, J. Dombre, C. Larabi et C. Fernandez-Maloigne
à paraître dans RFIA 04, Workshop fouille de données. Janvier 2004, toulouse, France.

Coresa2001 « Systèmes de représentation haut-niveau pour l'indexation »
J. Dombre, N. Richard et C. Fernandez-Maloigne
Conférence sur COmpression et REprésentation des Signaux Audiovisuels, Dijon, France, Novembre 10-12, 2001, pp. 49-52

C.4 Mémoires

DEA « Outils de segmentation spatio-temporelle pour MPEG-4 »
J. Dombre
Mémoire de D.E.A. Traitement d'Images, soutenu le 10 Juillet 2000

C.5 Rapport Interne

RI09052001 « Outil de prototypage pour la manipulation et l'expérimentation sur bases de données images couleur »
J. Dombre et E. Laizé
9 Mai 2001

BIBLIOGRAPHIE

[ABDULK98] A. M. ABDULKADER. « Parallel Algorithms for Labelled Graph Matching ». Thèse de doctorat, Colorado School of Mines, 1998.

[ALMOHA93] H. A. ALMOHAMAD & S. O. DUFFUAA. « A Linear Programming Approach for the Weighted Graph Matching Problem ». *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 15(5), pages 522–525, mai 1993.

[BECK87] J. BECK, A. SUTTER & R. IVRY. « Spatial Frequency Channels and Perceptual Grouping in Texture Segregation ». *Computer Vision, Graphics, and Image Processing (CVGIP)*, 37(2), pages 299–325, février 1987.

[BELKAS91] S. O. BELKASIM, M. SHRIDHAR & M. AHMADI. « Pattern Recognition With Moment Invariants: A Comparative Study and New Results ». *Pattern Recognition*, 24(12), pages 1117–1138, 1991.

[BEZDEK81] J. C. BEZDEK. « Pattern recognition with fuzzy objective function algorithms ». Plenum Press, New York, juillet 1981. ISBN 0-306-40671-3.

[BHATTA43] A. BHATTACHARYYA. « On a measure of divergence between two statistical populations defined by their probability distributions ». *Bulletin of Calcutta Mathematical Society*, vol. 35, pages 99–109, 1943.

[BLUM67] H. BLUM. « A Transformation for Extracting New Descriptions of Shape ». Dans *Models for the Perception of Speech and Visual Form*, pages 362–380, 1967.

[BLUM78] H. BLUM & R. N. NAGEL. « Shape Description using Weighted Symmetric Axis Features ». *Pattern Recognition*, 10(3), pages 167–180, 1978.

[BOBICK96] A. F. BOBICK & J. W. DAVIS. « Real-time recognition of activity using temporal templates ». Dans *Workshop on Applications of Computer Vision*, pages 39–42, décembre 1996.

[BOUJEM01] N. BOUJEMAA, J. FAUQUEUR, M. FERECATU, F. FLEURET, V. GOUET, B. L. SAUX & H. SAHBI. « IKONA: Interactive Specific and Generic Image Retrieval ». Dans *International Workshop on Multimedia Content-Based Indexing and Retrieval*, Rocquencourt, France, 2001.

[BOVIK90] A. C. BOVIK, M. CLARK & W. S. GEISLER. « Multichannel Texture Analysis Using Localized Spatial Filters ». *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 12(1), pages 55–73, janvier 1990.

[BRAINA86] D. H. BRAINARD & B. A. WANDELL. « Analysis of the Retinex Theory of Color Vision ». *Journal of the Optical Society of America*, 3(10), pages 1651–1661, 1986.

[BRAQUE98] J.-P. BRAQUELAIRE & L. BRUN. « Image segmentation with topological maps and inter-pixel representation ». *Journal of Visual Communication and Image Representation*, 9(1), pages 62–79, mars 1998.

[BRIBIE99] E. BRIBIESCA. « A new chain code ». *Pattern Recognition*, 32(2), pages 235–251, février 1999.

[BRICE70] C. BRICE & C. FENNEMA. « Scene analysis using regions ». *Artificial Intelligence*, 1(3), pages 205–226, 1970.

[BRUN03] L. BRUN & W. G. KROPATSCH. « Contraction kernels and combinatorial maps ». *Pattern Recognition Letters*, 24(8), pages 1051–1057, mai 2003.

[BURT81] P. J. BURT, T. H. HONG & A. ROSENFELD. « Segmentation and Estimation of Image Region Properties Through Cooperative Hierarchical Computation ». *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 11(12), pages 802–809, décembre 1981.

[BURT83] P. J. BURT & E. H. ADELSON. « The Laplacian Pyramid as a compact image code ». *IEEE Transactions on Communications*, 31(4), pages 532–540, avril 1983.

[CAMPBE68] F. W. CAMPBELL & J. G. ROBSON. « Application of Fourier analysis to the visibility of gratings ». *Journal of Physiology*, vol. 197, pages 551–566, 1968.

[CARRON94] T. CARRON & P. LAMBERT. « Color Edge Detector Using Jointly Hue, Saturation, and Intensity ». Dans *International Conference on Image Processing (ICIP)*, vol. 3, pages 977–981, Austin (TX), 13–16 novembre 1994.

[CARSON02] C. CARSON, S. BELONGIE, H. GREENSPAN & 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), pages 1026–1038, août 2002.

[CHANG87] S. K. CHANG, Q. Y. SHI & C. W. YAN. « Iconic indexing by 2-D strings ». *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 9(3), pages 413–428, mai 1987.

[CHANG89] S. K. CHANG, E. JUNGERT & Y. LI. « Representation and Retrieval of Symbolic Pictures Using Generalized 2-D Strings ». Dans *SPIE Visual Communications and Image Processing*, pages 1360–1372, Philadelphia (PA), novembre 1989.

[CHANG91] C. C. CHANG & S. Y. LEE. « Retrieval of Similar Pictures on Pictorial Databases ». *Pattern Recognition*, 24(7), pages 675–680, juillet 1991.

[CHANG96] C. C. CHANG & L. L. WANG. « Color Texture Segmentation for Clothing in a Computer-Aided Fashion Design System ». *International Conference on Image and Vision Computing (IVC)*, 14(9), pages 685–702, octobre 1996.

[CHAPRO97] M. CHAPRON. « A Chromatic Contour Detector based on Abrupt Change Techniques ». Dans *International Conference on Image Processing (ICIP)*, vol. 3, pages 18–21, Santa Barbara (CA), 26–29 octobre 1997.

[CHAUDH93] B. B. CHAUDHURI, N. SARKAR & P. KUNDU. « Improved Fractal Geometry Based Texture Segmentation Technique ». *IEE Proceedings - Vision Image and Signal Processing (VISP)*, vol. 140, pages 233–241, 1993.

[CHELLA85] R. CHELLAPPA & S. CHATTERJEE. « Classification of Textures Using Gaussian Markov Random Fields ». *IEEE Transactions Acoustics, Speech, and Signal Processing*, vol. 33, pages 959–963, août 1985.

[CHONG03] C. W. CHONG, P. RAVEENDRAN & R. MUKUNDAN. « A comparative analysis of algorithms for fast computation of Zernike moments ». *Pattern Recognition*, 36(3), pages 731–742, mars 2003.

[CHU90] A. CHU, C. M. SEHGAL & J. F. GREENLEAF. « Use Of Gray Value Distribution Of Run Lengths For Texture Analysis ». *Pattern Recognition Letters*, vol. 11, pages 415–420, 1990.

[CIOCCA01] G. CIOCCA, D. MARINI, A. RIZZI, R. SCHETTINI & S. ZUFFI. « On pre-filtering with Retinex in color image retrieval ». Dans *SPIE Internet Imaging*, San Jose (CA), janvier 2001.

[CONNER80] R. W. CONNERS & C. A. HARLOW. « A Theoretical Comparison of Texture Algorithms ». *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 2(3), pages 204–222, avril 1980.

[CRAMAR97] B. CRAMARIUC, M. GABBOUJ & J. ASTOLA. « Clustering Based Region Growing Algorithm for Color Image Segmentation ». Dans *International Conference on Digital Signal Processing*, vol. 2, pages 857–860, 2–4 juillet 1997.

[CURCIO90] C. A. CURCIO, K. R. SLOAN, R. E. KALINA & A. E. HENDRICKSON. « Human photoreceptor topography ». *Journal of Comparative Neurology*, vol. 292, pages 497–523, 1990.

[DAMIAN04] G. DAMIAND, Y. BERTRAND & C. FIORIO. « Topological Model for 2D Image Representation: Definition and Optimal Extraction Algorithm ». *Computer Vision and Image Understanding*, 2004.

[DANTZI51] G. B. DANTZIG. « Application of the simplex method to a transportation problem ». JOHN WILEY and Sons, 1951.

[DARRAS95] C. DARRAS. « Eléments et réflexions d'optique physiologique ». Editions ERA Nantes, 1995.

[DAUGMA88] J. G. DAUGMAN. « Complete Discrete 2D Gabor Transforms by Neural Networks for Image Analysis and Compression ». *IEEE Transactions Acoustics, Speech, and Signal Processing*, 36(7), pages 1169–1179, juillet 1988.

[DAVIS79] L. S. DAVIS. « Shape Matching Using Relaxation Techniques ». *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 1(1), pages 60–72, janvier 1979.

[DEKKER94] A. DEKKER. « Kohonen neural networks for optimal colour quantization ». *Networks: Computation in Neural Systems*, vol. 5, pages 351–367, 1994.

[DENG99] Y. DENG, B. S. MANJUNATH & H. SHIN. « Color Image Segmentation ». Dans *IEEE Computer Vision and Pattern Recognition Conference (CVPR)*, pages 446–451, IEEE, Los Alamitos, 23–25 juin 1999.

[DIZENZ86] S. DI ZENZO. « A Note on the Gradient of a Multi-Image ». *Computer Vision, Graphics, and Image Processing (CVGIP)*, 33(1), pages 116–125, janvier 1986.

[DOWLIN87] J. E. DOWLING. « The Retina: An Approachable Part of the Brain ». Harvard University Press, Cambridge, MA, 1987.

[DUYGUL02] P. DUYGULU, K. BARNARD, J. F. G. DE FREITAS & D. A. FORSYTH. « Object Recognition as Machine Translation: Learning a Lexicon for a Fixed Image Vocabulary ». Dans *European Conference in Computer Vision (ECCV)*, vol. 4, pages 97–112, Copenhagen, Denmark, 2002.

[ESHERA84] M. A. ESHERA & K. S. FU. « A Graph Distance Measure for Image Analysis ». *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 14(3), pages 398–408, mai 1984.

[FAIRCH02] M. D. FAIRCHILD & G. M. JOHNSON. « Meet iCAM : An Image Color Appearance Model ». Dans *IS&T/SID Color Imaging Conference*, Scottsdale (AZ), 2002.

[FERNAN01] C. FERNANDEZ-MALOIGNE, N. RICHARD & E. LAIZÉ. « Content-based image indexing with color texture descriptors ». Dans *Artificial Intelligence And Cognitive Science Within Virtual Environments*, Orlando (FL), 22–25 juillet 2001.

[FINLAY96] G. D. FINLAYSON, S. S. CHATTERJEE & B. V. FUNT. « Color Angular Indexing ». Dans *European Conference in Computer Vision (ECCV)*, vol. 2, pages 16–27, 1996.

[FINLAY98] G. D. FINLAYSON, B. SCHIELE & J. L. CROWLEY. « Comprehensive Colour Image Normalization ». Dans *European Conference in Computer Vision (ECCV)*, pages 475–490, Freiburg, 1998.

[FLUSSE93] J. FLUSSER & T. SUK. « Pattern Recognition by Affine Moment Invariants ». *Pattern Recognition*, vol. 26, pages 167–174, 1993.

[FRANCO93] J. M. FRANCOS. « Orthogonal Decompositions of 2-D Random Fields and their Applications in 2-D Spectral Estimation ». Dans *Signal Processing and Its Applications Volume, Handbook of Statistics*, pages 207–227. N. K. BOSE and C. R. RAO, North-Holland Publishing Compagny, 1993.

[FREEMA61] H. FREEMAN. « On the encoding of arbitrary geometric configuration ». *IEEE Transactions on Computers*, 10(2), pages 260–268, juin 1961.

[FUNT95] B. V. FUNT & G. D. FINLAYSON. « Color Constant Color Indexing ». Dans *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 17, pages 522–529, 1995.

[FUNT96] B. V. FUNT, V. CARDEI & K. BARNARD. « Learning Color Constancy ». Dans *IS&T/SID Color Imaging Conference*, pages 58–60, novembre 1996.

[FUNT98] B. V. FUNT, K. BARNARD & L. MARTIN. « Is Machine Colour Constancy Good Enough ? ». Dans *European Conference in Computer Vision (ECCV)*, pages 445–459, 1998.

[GALLOW75] M. M. GALLOWAY. « Texture Analysis Using Gray Level Run Lengths ». *Color in Graphics and Image Processing (CGIP)*, vol. 4, pages 172–179, 1975.

[GAREY79] M. R. GAREY & D. S. JOHNSON. « Computers and intractability: A guide to the theory of NP-completeness ». W.H. Freeman and Co, 1979.

[GAUCH99] J. M. GAUCH. « Image Segmentation and Analysis via Multiscale Gradient Watershed Hierarchies ». *IEEE Transactions on Image Processing*, 8(1), page 69, janvier 1999.

[GEMAN84] S. GEMAN & D. GEMAN. « Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images ». *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 6(6), pages 721–741, novembre 1984.

[GERVAU90] M. GERVAUTZ & W. PURGATHOFER. « A Simple Method for Color Quantization: Octree Quantization ». Graphics Gems, Academic Press, 1990.

[GEVERS96] T. GEVERS & A. W. M. SMEULDERS. « A comparative study of several color models for color image invariant retrieval ». Dans *Workshop on Image Databases and Multimedia Search*, page 17, Amsterdam, Netherlands, 1996.

[GEVERS97] T. GEVERS & A. W. M. SMEULDERS. « Combining Region Splitting and Edge Detection Through Guided Delaunay Image Subdivision ». Dans *IEEE Computer Vision and Pattern Recognition Conference (CVPR)*, pages 1021–1026, 17–19 juin 1997.

[GEVERS99A] T. GEVERS & A. W. M. SMEULDERS. « Color-based object recognition ». *Pattern Recognition*, 32(3), pages 453–464, mars 1999.

[GEVERS99B] T. GEVERS & A. W. M. SMEULDERS. « The PicToSeek WWW Image Search System ». Dans *International Conference on Multimedia Computing and Systems*, pages 264–269, 1999.

[GEVERS00] T. GEVERS & A. W. M. SMEULDERS. « PicToSeek: Combining Color and Shape Invariant Features for Image Retrieval ». *IEEE Transactions on Image Processing*, 9(1), page 102, janvier 2000.

[GILLET01] A. GILLET, L. MACAIRE, C. BOTTE-LECOCQ & J.-G. POSTAIRE. « Color Image Segmentation by Fuzzy Morphological Transformation of the 3D Color Histogram ». Dans *IEEE International Conference on Fuzzy Systems*, vol. 824, 2001.

[GOLD96A] S. GOLD & A. RANGARAJAN. « A Graduated Assignment Algorithm for Graph Matching ». *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 18(4), pages 377–388, avril 1996.

[GOLD96B] S. GOLD & A. RANGARAJAN. « Graph Matching by Graduated Assignment ». Dans *IEEE Computer Vision and Pattern Recognition Conference (CVPR)*, pages 239–244, 1996.

[GOTLIE90] C. C. GOTLIEB & H. E. KREYSZIG. « Texture Descriptors Based on Co-occurrence Matrices ». *Computer Vision, Graphics, and Image Processing (CVGIP)*, vol. 51, pages 70–84, juillet 1990.

[GUDIVA95] V. N. GUDIVADA & V. V. RAGHAVAN. « Design and Evaluation of Algorithms for Image Retrieval by Spatial Similarity ». *Information Systems*, 13(2), pages 115–144, 1995.

[GUILD31] J. GUILD. « The colorimetric properties of the spectrum ». *Philosophical Transactions of the Royal Society of London*, vol. 230, pages 149–187, 1931.

[HAFNER95] J. HAFNER, H. S. SAWHNEY, W. EQUITZ, M. FLICKNER & W. NIBLACK. « Efficient Color Histogram Indexing for Quadratic Form Distance Functions ». *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 17(7), pages 729–736, juillet 1995.

[HARALI73] R. M. HARALICK, K. S. SHANMUGAN & I. DUNSTEIN. « Textural features for image classification ». *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 3(6), pages 610–621, novembre 1973.

[HARALI85] R. M. HARALICK & L. G. SHAPIRO. « Image Segmentation Techniques ». *Computer Vision, Graphics, and Image Processing (CVGIP)*, 29(1), pages 100–132, janvier 1985.

[HECKBE80] P. S. HECKBERT. « Color Image Quantization for Frame Buffer Display ». *Bachelor of Sciences*, Architecture Machine Group, MIT, mai 1980.

[HECKBE82] P. S. HECKBERT. « Color Image Quantization for Frame Buffer Display ». *Computer Graphics*, 16(3), pages 297–307, 1982.

[HEINRI99] A. HEINRICHS, D. KOUBAROULIS, B. LEVIENAISE-OBADIA, P. ROVIDA & J.-M. JOLION. « Robust Image Retrieval in a Statistical Framework ». Rapport technique RR-99-04, Laboratoire Reconnaissance de Formes et Vision, INSA Lyon, 1999.

[HERING78] E. HERING. « Zur lehre vom lichtsinne ». 1878. Traduction anglaise par L. M. HURVICH et D. JAMESON.

[HOROWI74] S. L. HOROWITZ & T. PAVLIDIS. « Picture segmentation by a directed split-and-merge procedure ». Dans *International Conference on Pattern Recognition (ICPR)*, pages 424–433, 1974.

[HU62] M.-K. HU. « Visual Pattern Recognition by Moment Invariants ». *IRE Transactions on Information Theory*, 8(2), pages 179–187, février 1962.

[HUANG97] J. HUANG, S. R. KUMAR, M. MITRA, W. J. ZHU & R. ZABIH. « Image indexing using color correlograms ». Dans *IEEE Computer Vision and Pattern Recognition Conference (CVPR)*, pages 762–768, juin 1997.

[HWANG99] W.-S. HWANG, J. J. WENG, M. FANG & J. QIAN. « A Fast Image Retrieval Algorithm with Automatically Extracted Discriminant Features ». Dans *IEEE Workshop on Content-based Access of Image and Video Libraries (CBAIVL)*, 1999.

[IIVARI97] J. IIVARINEN, M. PEURA, J. SARELA & A. VISA. « Comparison of Combined Shape Descriptors for Irregular Objects ». Dans *British Machine Vision Conference*, pages 430–439, 1997.

[JACOBS95] C. E. JACOBS, A. FINKELSTEIN & D. H. SALESIN. « Fast Multiresolution Image Querying ». Dans *ACM SIGGraph, Computer Graphics*, vol. 29, pages 277–286, 1995.

[JAIN98] A. JAIN & G. HEALEY. « A Multiscale Representation Including Opponent-Color Features for Texture Recognition ». *IEEE Transactions on Image Processing*, 7(1), pages 124–128, janvier 1998.

[JOLION92] J.-M. JOLION & A. MONTANVERT. « The Adaptive Pyramid: A framework for 2D Image Analysis ». *Computer Vision, Graphics, and Image Processing (CVGIP)*, 55(3), pages 339–348, mai 1992.

[JUDD30] D. B. JUDD. « Reduction of data on mixture of color stimuli ». *Journal of the Optical Society of America*, 4(163), pages 515–548, 1930.

[JULESZ75] B. JULESZ. « Experiments in the visual perception of texture ». *Scientific American*, 232(4), pages 34–43, avril 1975.

[KHOTAN90] A. KHOTANZAD & Y. H. HONG. « Invariant Image Recognition by Zernike Moments ». *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 12(5), pages 489–497, mai 1990.

[KOUBAR02] D. KOUBAROULIS, J. MATAS & J. KITTLER. « Evaluating Colour-Based Object Recognition Algorithms Using the SOIL-47 Database ». Dans *Asian Conference on Computer Vision (ACCV)*, (5), Melbourne, Australia, 2002.

[KRCMAR94] M. KRCMAR & A. P. DHAWAN. « Application of genetic algorithms in graph matching ». Dans *IEEE International Conference on Neural Networks (ICNN)*, vol. 6, pages 3872–3876, 1994.

[KROPAT95] W. KROPATSCH & H. MACHO. « Finding the structure of connected components using dual irregular pyramids ». Dans *Discrete Geometry for Computer Imagery (DGCI)*, pages 147–158, *invited lecture*, september 1995.

[KUNER88] P. KUNER & B. UEBERREITER. « Pattern Recognition by Graph Matching Combinatorial Versus Continuous Optimization ». *International Journal on Pattern Recognition and Artificial Intelligence*, vol. 2, pages 527–542, 1988.

[LAND71] E. H. LAND & J. J. MCCANN. « Lightness and Retinex Theory ». *Journal of the Optical Society of America*, 61(1), pages 1–11, 1971.

[LARABI00] M. C. LARABI, N. RICHARD & C. FERNANDEZ. « A Fast Color Quantization Using a Matrix of Local Pallets ». Dans *IEEE Applied Imagery Pattern Recognition*, pages 136–140, Washington, octobre 2000.

[LARABI02] M. C. LARABI. « Codage et analyse d'images couleur : application à l'indexation de bases dimages réparties ». Thèse de doctorat, Université de Poitiers, 9 décembre 2002.

[LE GRA72] Y. LE GRAND. « Optique Physiologique : Lumière et Couleur II ». Masson, Paris, 1972.

[LEE89] S. Y. LEE, M. K. SHAN & W. P. YANG. « Similarity retrieval of iconic image database ». *Pattern Recognition*, 22(6), pages 675–682, 1989.

[LEE90] S. Y. LEE & F.-J. HSU. « A New Spatial Knowledge Representation for Image Database Systems ». *Pattern Recognition*, 23(10), pages 1077–1087, 1990.

[LI92] Y. LI. « Reforming the Theory of Invariant Moments for Pattern Recognition ». *Pattern Recognition*, vol. 25, pages 723–730, 1992.

[LI02] C. LI, M. R. LUO, R. W. G. HUNT, N. MORONEY, M. D. FAIRCHILD & T. NEWMAN. « The Performance of CIECAM02 ». Dans *Color Image Conference: Color Science, Systems and Applications*, pages 51–54, 2002.

[LIENHA91] P. LIENHARDT. « Topological models for boundary representation: a comparison with n-dimensional generalized maps ». *Computer-Aided Design (CAD)*, 23(1), pages 59–82, janvier 1991.

[LINDEB90] T. LINDEBERG. « Scale-Space for Discrete Signals ». *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 12(3), pages 234–254, 1990.

[LIU96] F. LIU & 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), pages 722–733, 1996.

[MA99] W.-Y. MA & B. S. MANJUNATH. « NeTra: A Toolbox for Navigating Large Image Databases ». *Multimedia Systems*, 7(3), pages 184–198, 1999.

[MACADA42] D. L. MACADAM. « Visual Sensitivities to Small Color Differences in Daylight ». *Journal of the Optical Society of America*, vol. 32, pages 247–274, 1942.

[MACADA85] D. L. MACADAM. « Color measurement: theme and variation ». Springer Verlag, 2e édition, 1985.

[MAHALA36] P. C. MAHALANOBIS. « On the Generalized Distance in Statistics ». *National Institute of Sciences in India*, vol. 12, pages 49–55, 1936.

[MANDAL96] M. K. MANDAL, T. ABOULNASR & S. PANCHANATHAN. « Image indexing using moments and wavelets ». Dans *IEEE Transactions on Consumer Electronics*, 42(3), pages 557–, 1996.

[MANJUN96] B. S. MANJUNATH & W. Y. MA. « Texture Features For Browsing And Retrieval Of Image Data ». *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 18(8), pages 837–842, août 1996.

[MAO92] J. MAO & A. K. JAIN. « Texture Classification and Segmentation using Multiresolution Simultaneous Autoregressive Models ». *Pattern Recognition*, 25(2), pages 173–188, 1992.

[MARINI00] D. MARINI, A. RIZZI & L. D. CARLI. « Multiresolution retinex: comparison of algorithms ». Dans *Color in Graphics and Image Processing (CGIP)*, Saint-Étienne, France, 1–4 mars 2000.

[MEYER90] F. MEYER & S. BEUCHER. « Morphological Segmentation ». *Journal of Visual Communication and Image Representation*, 1(1), pages 21–46, septembre 1990.

[MOKHTA88] F. MOKHTARIAN & A. K. MACKWORTH. « The Renormalized Curvature Scale Space and the Evolution Properties of Planar Curves ». Dans *IEEE Computer Vision and Pattern Recognition Conference (CVPR)*, pages 318–326, 1988.

[MOKHTA95] F. MOKHTARIAN. « Silhouette-Based Isolated Object Recognition through Curvature Scale-Space ». *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 17(5), pages 539–544, mai 1995.

[MOKHTA96] F. MOKHTARIAN, S. ABBASI & J. V. KITTLER. « Robust and Efficient Shape Indexing through Curvature Scale Space ». Dans *British Machine Vision Conference*, 1996.

[MONTAN91] A. MONTANVERT, P. MEER & A. ROSENFELD. « Hierarchical Image Analysis Using Irregular Tessellations ». *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 13(4), pages 307–316, avril 1991.

[MULLER99] H. MULLER, W. MULLER, D. M. SQUIRE & T. PUN. « Performance evaluation in content based image retrieval: Overview and proposals ». Rapport technique 99.05, Computer Vision Group, Computing Centre University of Geneva, décembre 1999.

[NIBLAC93] W. NIBLACK, R. BARBER, W. EQUITZ, M. D. FLICKNER, D. GLASMAN, D. PETKOVIC & P. YANKER. « The QBIC Project: Querying Image by Content Using Color, Texture, and Shape ». *SPIE Storage and Retrieval for Image and Video Databases*, vol. 1908, pages 173–187, février 1993.

[OHANIA92] P. P. OHANIAN & R. C. DUBES. « Performance Evaluation for Four Classes of Textural Features ». *Pattern Recognition*, vol. 25, pages 819–833, 1992.

[OZER99] B. OZER, W. WOLF & A. N. AKANSU. « A graph based object description for information retrieval in digital image and video libraries ». Dans *IEEE Workshop on Content-based Access of Image and Video Libraries (CBAIVL)*, pages 79–83, 1999.

[PADGHA75] C. A. PADGHAM & J. E. SAUNDERS. « The Perception of Light and Color ». Academic Press, London, 1975.

[PAL93] N. R. PAL & S. K. PAL. « A review on image segmentation techniques ». *Pattern Recognition*, 26(9), pages 1277–1294, 1993.

[PALM02] C. PALM & T. M. LEHMANN. « Classification of color textures by Gabor filtering ». *Machine Graphics and Vision*, 11(2/3), pages 195–219, 2002.

[PAPPAS88] T. N. PAPPAS & N. S. JAYANT. « An Adaptive Clustering Algorithm for Image Segmentation ». Dans *International Conference on Computer Vision (ICCV)*, pages 310–315, 1988.

[PARK91] R. H. PARK & P. MEER. « Edge-Preserving Artifact-Free Smoothing with Image Pyramids ». *Pattern Recognition Letters*, vol. 12, pages 467–475, 1991.

[PARK98] S. PARK, I. YUN & S. LEE. « Color Image Segmentation Based on 3-D Clustering: Morphological Approach ». *Pattern Recognition*, 31(8), pages 1061–1076, août 1998.

[PARK99A] J. S. PARK & D. H. CHANG. « 2-D Invariant Descriptors for Shape-Based Image Retrieval ». Dans *KISS Fall Conference*, 26(2), pages 554–556, Kwangwoon University, Korea, octobre 1999.

[PARK99B] K. PARK, I. YUN & S. LEE. « Color Image Retrieval Using a Hybrid Graph Representation ». *Journal of Image and Vision Computing (IVCJ)*, vol. 17, pages 465–474, 1999.

[PASS96] G. PASS, R. ZABIH & J. MILLER. « Comparing Images Using Color Coherence Vectors ». Dans *ACM Multimedia*, pages 65–74, ACM Press, New York, NY, USA, novembre 1996. ISBN 0-201-92140-X.

[PELEG84] S. PELEG, J. NAOR, R. HARTLEY & D. AVNIR. « Multiple Resolution Texture Analysis and Classification ». *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 6(4), pages 518–523, juillet 1984.

[PENTLA84] A. P. PENTLAND. « Fractal-Based Description of Natural Scenes ». *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 6(6), pages 661–674, novembre 1984.

[PERONA90] P. PERONA & J. MALIK. « Scale-space and edge detection using anisotropic diffusion ». *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 12(7), pages 629–639, juillet 1990.

[PERSOO77] E. PERSOON & K. S. FU. « Shape discrimination using Fourier descriptors ». *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 7(3), pages 170–179, mars 1977.

[PETRAK93] E. G. M. PETRAKIS & S. C. ORPHANOUDAKIS. « Methodology for the Representation, indexing and retrieval of images by content ». *International Conference on Image and Vision Computing (IVC)*, 11(8), pages 504–521, octobre 1993.

[PETRAK96] E. G. M. PETRAKIS & S. C. ORPHANOUDAKIS. « A Generalized Approach to Image Indexing and Retrieval Based on 2-D Strings ». *Intelligent Image Database Systems. World Scientific Publishing Co.*, pages 197–218, 1996.

[PETRAK97] E. G. M. PETRAKIS & C. FALOUTSOS. « Similarity Searching in Large Image Databases ». Dans *IEEE Transactions on Knowledge and Data Engineering (KDE)*, vol. 9, pages 435–447, 1997.

[PETRAK02] E. G. M. PETRAKIS. « Design and Evaluation of Spatial Similarity Approaches for Image Retrieval ». *International Conference on Image and Vision Computing (IVC)*, 20(1), pages 59–76, janvier 2002.

[PORTER97] R. PORTER & N. CANAGARAJAH. « Robust Relation-Invariant Texture Classification: Wavelet, Gabor Filter and GMRF Based Schemes ». *IEE Proceedings - Vision Image and Signal Processing (VISP)*, 144(3), pages 180–188, juin 1997.

[PREWIT66] J. M. S. PREWITT & M. L. MENDELSOHN. « The Analysis of Cell Images ». *Annals of New York Academu of Sciences*, vol. 128, pages 1035–1053, 1966.

[PUZICH99] J. PUZICHA, J. M. BUHMANN, Y. RUBNER & C. TOMASI. « Empirical Evaluation of Dissimilarity Measures for Color and Texture ». Dans *International Conference on Computer Vision (ICCV)*, pages 1165–1173, 1999.

[RANGAN92] H. S. RANGANATH & L. J. CHIPMAN. « Fuzzy Relaxation Approach for Inexact Scene Matching ». *International Conference on Image and Vision Computing (IVC)*, 10(9), pages 631–640, 1992.

[RAUBER94] T. W. RAUBER. « Two-diemnsional shape description ». Rapport technique GR UNINOVA-RT-10-94, UNINOVA - Intelligent Robotics Center Quinta da Torre, Monte de Caparica, Portugal - Universidade Nova de Lisboa, 1994.

[ROSENF69] A. ROSENFELD. « Picture Processing by Computer ». Academic Press, 1969.

[RUBNER98A] Y. RUBNER & C. TOMASI. « Comparing the EMD to other Dissimilarity Measures for Color Images ». Dans *DARPA Image Understanding Workshop*, pages 331–339, 1998.

[RUBNER98B] Y. RUBNER & C. TOMASI. « Texture Metrics ». Dans *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 4601–4607, San Diego (CA), octobre 1998.

[RUBNER98C] Y. RUBNER, C. TOMASI & L. J. GUIBAS. « A Metric for Distributions with Applications to Image Databases ». Dans *International Conference on Computer Vision (ICCV)*, pages 59–66, 1998.

[RUI96A] Y. RUI, A. SHE & T. HUANG. « Modified Fourier descriptors for shape representation-a practical approach ». Dans *Workshop on Image Databases and Multimedia Search*, Amsterdam, Netherlands, 1996.

[RUI96B] Y. RUI, A. SHE & T. S. HUANG. « Automated Region Segmentation Using Attraction-Based Grouping in Spatial-Color-Texture Space ». Dans *International Conference on Image Processing (ICIP)*, vol. 1, pages 53–56, septembre 1996.

[RUSSEL69] E. J. RUSSEL. « Extension of Dantzig's algorithm to finding an initial near optimal-basis for the transportation problem ». *Operations Research*, vol. 17, pages 187–191, 1969.

[SABER95] E. SABER, A. M. TEKALP, R. ESCHBACH & K. KNOX. « Annotation of Natural Scenes Using Adaptative Color Segmentation ». Dans *SPIE - The International Society for Optical Engineering, Image and Video Processing*, pages 72–80, SPIE, San Jose (CA), février 9–10 1995.

[SALTON92] G. SALTON. « The State of Retrieval System Evaluation ». *Information Processing and Management*, 28(4), pages 441–450, 1992.

[SAMET80] H. SAMET. « Region Representation: Quadtree from Binary Arrays ». *CGIP*, 13(1), pages 88–93, May 1980.

[SANGWI79] S. J. SANGWINE. « Fourier transforms of colour images using quaternion or hypercomplex numbers ». *Electronics Letters*, 32(21), 1979.

[SANGWI00] S. J. SANGWINE. « Colour Image Processing ». *Electronics and Communication Engineering Journal*, 15(5), pages 211–219, octobre 2000.

[SCHETT94] R. SCHETTINI & M. SUARDI. « A Low-Level Segmentation Procedure for Color Images ». Dans *European Signal Processing Conference (EUSIPCO)*, vol. 1, pages 26–29, Lausanne Switzerland, 13–16 septembre 1994.

[SEBAST01] T. B. SEBASTIAN, P. N. KLEIN & B. B. KIMIA. « Recognition of Shapes by Editing Shock Graphs ». Dans *International Conference on Computer Vision (ICCV)*, pages 755–762, 2001.

[SHAFAR98] L. SHAFARENKO, M. PETROU & J. V. KITTLER. « Histogram Based Segmentation in a Perceptually Uniform Color Space ». *IEEE Transactions on Image Processing*, 7(9), pages 1354–1358, septembre 1998.

[SHARVI98] D. SHARVIT, J. CHAN, H. TEK & B. B. KIMIA. « Symmetry-based indexing of image databases ». Dans *IEEE Workshop on Content-based Access of Image and Video Libraries (CBAIVL)*, pages 56–62, 23–25 janvier 1998.

[SINKHO64] R. SINKHORN. « A relationship between arbitrary positive matrices and doubly stochastic matrices ». *Annals of Mathematical Statistics*, vol. 35, pages 876–879, 1964.

[SMITH94] J. SMITH & S. F. CHANG. « Transform Features for Texture Classification and Discrimination in Large Image Database ». Dans *International Conference on Image Processing (ICIP)*, 1994.

[SMITH95] J. SMITH & S. CHANG. « Single color extraction and image query ». Dans *International Conference on Image Processing (ICIP)*, 1995.

[SMITH96] J. SMITH & S.-F. CHANG. « VisualSEEk: A fully automated content-based image query system ». Dans *ACM Multimedia*, pages 87–98, ACM Press, New York, NY, USA, novembre 1996. ISBN 0-201-92140-X.

[SMITH97] J. R. SMITH. « Integrated Spatial and Feature Image Systems: retrieval, Analysis and Compression ». Thèse de doctorat, Columbia University, USA, 1997.

[STRICK95] M. STRICKER & M. ORENGO. « Similarity of color images ». Dans *SPIE Storage and Retrieval for Image and Video Databases*, vol. 2420, pages 381–392, février 1995.

[SWAIN91] M. J. SWAIN & D. H. BALLARD. « Color indexing ». *International Journal on Computer Vision (IJCV)*, 7(1), pages 11–32, novembre 1991.

[TAKAHA95] K. TAKAHASHI, H. NAKATANI & K. ABE. « Color image segmentation using ISODATA clustering method ». Dans *Asian Conference on Computer Vision*, vol. 1, pages 523–527, Singapore, 1995.

[TAUBIN92] G. TAUBIN & D. B. COOPER. « Object Recognition Based on Moment (or Algebraic) Invariants ». Dans J. MUNDY & A. ZISSERMAN, rédacteurs, *Geometric Invariance in Computer Vision*, pages 375–397, MIT Press, 1992.

[TREMBL02] B. TREMBLAIS. « De la résolution numérique des EDP à l'extraction de caractéristiques linéiques dans les images : application à la détection multi-échelle d'un arbre vasculaire ». Thèse de doctorat, Université de Poitiers, décembre 2002.

[TRÉMEA97] A. TRÉMEAU & N. BOREL. « A Region Growing and Merging Algorithm to Color Segmentation ». *Pattern Recognition*, 30(7), pages 1191–1203, 1997.

[TRÉMEA03] A. TRÉMEAU, C. FERNANDEZ-MALOIGNE & P. BONTON. « Image numérique couleur ». Editions DUNOD, à paraître 1er novembre 2003.

[TSAI83] W. H. TSAI & K. S. FU. « Subgraph Error-Correcting Isomorphisms for Syntatic Pattern Recognition ». *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 13(1), pages 48–62, janvier 1983.

[TSCHUM02] D. TSCHUMPERLE. « PDE-Based Regularization of Multivalued Images and Applications ». Thèse de doctorat, University of Nice-ophia Antipolis, December 2002.

[VALKEA98] K. VALKEALAHTI & E. OJA. « Reduced Multidimensional Cooccurrence Histograms in Texture Classification ». *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 20(1), pages 90–94, janvier 1998.

[VANDEN00] N. VANDENBROUCKE. « Segmentation d'images couleur par classification de pixels dans des espaces d'attributs colorimétriques adaptés. Application à l'analyse d'images de football ». Thèse de doctorat, Université des Sciences et Technologies de Lille, décembre 2000.

[VANHAM03] I. VANHAMEL, I. PRATIKAKIS & H. SAHLI. « Multiscale gradient watersheds of color images ». *IEEE Transactions on Image Processing*, 12(6), pages 617–626, juin 2003.

[VEREVK95] O. VEREVKA. « Color image quantization in window systems with local K-means algorithm ». Dans *Western Computer Graphics Symposium*, Banff (AL), 1995.

[VINCEN91] L. VINCENT & P. SOILLE. « Watersheds in Digital Spaces: An Efficient Algorithm Based on Immersion Simulations ». *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 13(6), pages 583–598, juin 1991.

[VONHEL67] H. VON HELMHOLTZ. « Manuel d'optique physiologique ». V. Masson et fils, 1867. Traduit de l'allemand par E. JAVAL et N. T. KLEIN.

[WAN90] S. J. WAN, P. PRISINKIEWICZ & S. K. M. WONG. « Variance-Based color image quantization for frame buffer display ». *Color Research and Application*, 15(1), pages 52–58, 1990.

[WATSON87] A. I. WATSON. « A New Method of Classification for Landsat Data Using the "Watershed" Algorithm ». *Pattern Recognition Letters*, vol. 6, pages 15–19, 1987.

[WEEKS97] A. R. WEEKS & G. E. HAGUE. « Color segmentation in the HSI color space using the k-means algorithm ». Dans *SPIE Nonlinear Image Processing*, pages 143–154, San Jose (CA), février 10–11 1997.

[WEICKE98] J. WEICKERT. « Fast Segmentation Methods Based on Partial Differential Equations and the Watershed Transformation ». Dans *DAGM-Symposium*, pages 93–100, 1998.

[WITKIN83] A. P. WITKIN. « Scale-space filtering ». Dans *International Joint Conference on Artificial Intelligence*, pages 1019–1021, 1983.

[WONG98] Y. Y. WONG, P. C. YUEN & C. S. TONG. « Segmented snake for contour detection ». *Pattern Recognition*, 31(11), pages 1669–1679, novembre 1998.

[WRIGHT29] W. D. WRIGHT. « A re–determination of the trichromatic coefficients of the spectral colours ». *Transactions of the Optical Society*, vol. 30, pages 141–164, 1929.

[XU00] Y. XU, P. DUYGULU, E. SABER, M. TEKALP & F. YARMANVURAL. « Object Based Image Retrieval Based On MultiLevel Segmentation ». Dans *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, juin 2000.

[YANG94] L. YANG & F. ALGRETSEN. « Fast computation of invariant moments: A new method giving correct results ». Dans *International Conference on Image Processing (ICIP)*, 1994.

[YOUNG07] T. YOUNG. « Lectures on natural philosophy and mechanical arts ». Thoemmes Press, 1807.

[YOUNG91] R. A. YOUNG. « Oh say, can you see ? The physiology of vision ». Dans *SPIE Human Vision, Visual Processing and Digital Display*, vol. 1453, pages 92–123, San Jose (CA), 1991.

[YOUNG97] S. S. YOUNG, P. D. SCOTT & N. M. NASRABADI. « Object Recognition Using Multilayer Hopfield Neural-Network ». *IEEE Transactions on Image Processing*, 6(3), pages 357–372, mars 1997.

[ZHANG01] D. S. ZHANG & G. LU. « A Comparative Study on Shape Retrieval Using Fourier Descriptors with Different Shape Signatures ». Dans *International Conference on Intelligent Multimedia and Distance Education*, pages 1–9, Fargo, ND, USA, 1–3 juin 2001.

[ZHANG02] D. S. ZHANG & G. LU. « A comparative study of curvature scale space and Fourier descriptors for shape-based image retrieval ». *Journal of Visual Communication and Image Representation*, 14(1), pages 39–57, mars 2002.

[ZILIAN98] F. ZILIANI & B. JENSEN. « Unsupervised Image Segmentation Using the Modified Pyramidal Linking Approach ». Dans *International Conference on Image Processing (ICIP)*, vol. 3, pages 303–307, Chicago (IL), 4–7 octobre 1998.

Résumé Nos travaux ont été motivés par un problème pratique visant à rechercher des objets au sein d'images et en particulier dans des bases d'images de fresques médiévales. Les outils développés devront aider les historiens de l'art dans leur travail quotidien en leur fournissant la possibilité de retrouver et de caractériser des images ou des objets similaires dans l'ensemble des images dont ils disposent. Cette étude se situe par conséquent dans le cadre de la recherche d'images par le contenu. Pour résoudre ce problème, les méthodes classiques sont inefficaces car elles caractérisent les images globalement sans prendre en compte l'arrangement des objets qui les compose.
La méthode proposée décrit l'image par un graphe pyramidal considérant alors l'image composée d'objets complexes en relation les uns avec les autres. Un algorithme de segmentation multi-échelles génère un ensemble de partitions cohérentes entre elles isolant ainsi les objets puis leurs parties. Les graphes d'adjacence des différents niveaux sont alors liés les uns avec les autres pour obtenir le graphe pyramidal de l'image. Au sein de cette structure de représentation de haut-niveau, nous avons ensuite cherché à caractériser chaque région le plus correctement possible à partir de sa couleur, sa texture et sa forme.
Divers algorithmes sont enfin proposés pour utiliser de manière combinée ces description des objets ainsi que l'information spatiale et pyramidale de la représentation pour déterminer la similarité entre images et détecter des objets semblables au sein de la base d'images. De nombreux exemples expérimentaux illustrent ce manuscrit et montrent que cette approche est globalement plus performante que les systèmes existants.

Multi-scale representation systems for indexing and restoring color medieval archives

Abstract Our work has been motivated by a practical problem aiming at seeking objects within images and in particular in image databases of medieval frescos. The developed tools will have to help the art historians in their daily work in enabling retrieving of similar images or objects in the group of the images they have. This study is consequently within the framework of content based image retrieval. In order to solve this problem, the traditional methods are ineffective because they characterize the images as a whole without taking the arrangement of the objects which composes them into account.
Suggested method describes the image by a pyramidal graph. The images are considered to be made up complex objects in relationship. A algorithm computes several coherent segmentations isolating the objects and theirs parts. Then, the adjacency graphs of the various levels are linked in order to obtain the pyramidal graph of the image. Within this high-level representation structure, we characterize each region most correctly as possible using its color, its texture and its shape.
Various algorithms are finally proposed to use these descriptions of the objects, the spatial and the pyramidal informations of the representation in order to determine the similarity between images or to detect similar objects within the database. Many experimental examples illustrate this manuscript and show that this approach is overall more powerful than the existing systems.

Discipline : traitement du signal et des images.

Mots clés : indexation d'images par le contenu, structure multi-échelles, segmentation, couleur, forme, texture, médiévale.

Laboratoire IRCOM-SIC, UMR 6615 CNRS
Bât SP2MI - Téléport 2, Boulevard Marie et Pierre Curie
BP 30179 86962 Futuroscope Chasseneuil Cedex, France