



# Analyse et organisation de corpus pour une recherche thematico-visuelle d'images

Gérald Duffing

► **To cite this version:**

Gérald Duffing. Analyse et organisation de corpus pour une recherche thematico-visuelle d'images. XVIIème Congrès d'Informatique des Organisations et Systèmes d'Information et de Décision - INFORSID'99, 1999, La Garde, France, 20 p, 1999. <inria-00098747>

**HAL Id: inria-00098747**

**<https://hal.inria.fr/inria-00098747>**

Submitted on 26 Sep 2006

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

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

# Analyse et organisation de corpus pour une recherche thématico-visuelle d'images

Gérald Duffing

UMR Loria

BP 239 — F-54506 Vandoeuvre-les-Nancy cedex, France.

e-mail : duffing@loria.fr — Tél. : +33 (0)3.83.59.20.47 ; Fax. : +33 (0)3 83.41.30.79

## Résumé

La recherche d'images dans de grandes bases d'images suppose des capacités d'interrogation fondées à la fois sur le contenu visuel et thématique. Un lourd travail d'indexation manuelle des images est un préalable à toute tentative d'interrogation «réaliste» de la base, alors que les interrogations fondées uniquement sur le contenu visuel peuvent exploiter les techniques automatiques d'analyse d'image. Il nous semble que les deux approches peuvent être efficacement combinées, de manière à permettre au système de retrouver des documents thématiquement et visuellement pertinents, et également d'autoriser le traitement de bases qui ne sont pas *totale*ment indexées thématiquement. Dans cet article, nous présentons un système réalisant une telle intégration grâce à une pré-organisation du corpus, qui agit à la fois sur les caractéristiques thématiques et visuelles des images. Un processus de recherche adapté à cette organisation est proposé ; il privilégie l'interaction avec l'utilisateur par le biais d'un bouclage de pertinence thématique et visuel. Nous exposons également les résultats de nos expérimentations et quelques pistes d'approfondissement de notre approche.

## Abstract

Retrieving images in large databases involves both visual and thematic querying capabilities. Realistic retrieval requires manual indexing task, whereas visual retrieval can take advantage of automatic indexing methods based on image analysis. We believe that both approaches can be efficiently combined, in order to achieve relevant visual *and* thematic retrieval, and to allow partially indexed corpora to be exploited. This article presents such a system, based on a prior corpus organization relying on both thematic and visual image descriptions. An adapted retrieval process, including relevance feedback capabilities, is proposed.

**Mots-clés** : Recherche d'image fondée sur le contenu ; Analyse d'image ; Indexation thématique et visuelle d'images ; Clustérisation.

**Keywords** : Content-based image retrieval; Image Analysis; Thematic and visual indexing; Clustering.

**Catégorie** : Jeune chercheur.

# 1 Introduction

La constitution de corpus regroupant de grandes quantités d'images ne pose plus de problèmes techniques ou économiques majeurs. Essentielle à la recherche efficace des images, la tâche consistant à examiner chaque image de manière à l'indexer à l'aide de mots-clés reste cependant manuelle et longue. Nous pensons qu'un corpus partiellement indexé peut néanmoins être exploité par un système de recherche d'images, pourvu qu'il ait été organisé préalablement, et que le processus de recherche soit également adapté à cette organisation.

Les systèmes actuels peuvent être classés en deux catégories principales [AZP96].

- Les systèmes permettant une *interrogation par mots-clés* ne considèrent évidemment que les images indexées par ces mots-clés, de sorte que la qualité des résultats dépend fortement du soin apporté à l'indexation. Les modalités d'indexation des images sont variées, et de nombreux modèles ont été proposés, permettant une représentation par vecteurs de mots-clés (RIVAGE [Hal90], Cabri-n [Sma94]), ou par des structures de données plus complexes (VIMSYS [GWJ91], MMIS [GOC<sup>+</sup>92], MULTOS [RS92], système fondé sur EMIR2 [Mec95b, CM97]). Il est clair que la tâche d'indexation sera d'autant plus difficile que le modèle est complet.
- Les systèmes *fondés sur le contenu*<sup>1</sup> autorisent une recherche fondée sur le calcul d'une mesure de similarité visuelle. Le calcul des indices visuels permettant de caractériser les images est effectué lors de l'introduction de l'image dans le corpus. On pourra ainsi, avec ces systèmes, soumettre des requêtes sous forme d'esquisse, ou d'image exemple. Les aspects purement thématiques de la requête sont souvent peu ou pas pris en compte par ce type de systèmes. Citons les systèmes TradeMark [WLS95], Art-Museum [Kat92], VisualSEEk [SC97].

Des approches hybrides ont été expérimentées, visant à intégrer les deux types de systèmes. CHABOT [OS95] permet une association d'un mot-clé à un ensemble de caractéristiques de couleurs ; QBIC [NBE<sup>+</sup>93] autorise à la fois dans les requêtes l'utilisation de conditions exprimées sur la couleur, la forme, la texture, mais aussi sur la thématique au moyen de mots-clés.

Nous pensons cependant que les systèmes permettant une recherche par le contenu visuel ne prennent pas suffisamment en compte l'aspect thématique des requêtes, tout au moins lorsque l'indexation des images d'un point de vue thématique n'est pas exhaustive et correcte. Les relations sémantiques existant entre les mots ne sont pas prises en compte (CHABOT), et ceux-ci sont parfois utilisés comme un simple outil de pré-filtrage des images.

Il nous semble pourtant que le mot-clé est le médiateur le plus efficace entre l'utilisateur et le système, car il est à la fois précis et familier à l'utilisateur. Pour permettre une utilisation conjointe de conditions thématiques et visuelles dans une même requête, et tout en accordant une importance

---

1. Nous faisons ici référence au contenu *visuel* des images.

particulière à la thématique, nous proposons une nouvelle stratégie de recherche qui s'appuie sur un corpus pré-organisé. Nous exposons tout d'abord les méthodes de caractérisation d'images en terme d'indices visuels, ainsi que les modalités d'organisation du corpus qui en découlent. Nous présentons alors un processus de recherche adapté au corpus ainsi construit. Enfin, nous détaillons les résultats de nos expérimentations.

## 2 Analyse et organisation du corpus

### 2.1 La classification par clustérisation

Nous organisons le corpus par un mécanisme de classification visant à regrouper les images similaires : la clustérisation<sup>2</sup>. Comme le suggère l'hypothèse de clustérisation [Rij79], une amélioration de la qualité de la recherche peut être attendue : «*Les documents étroitement associés tendent à être pertinents pour les mêmes requêtes*». De plus, la clustérisation conduisant à un pré-calcul des similarités entre images, seuls les représentants des groupes constitués sont mis en correspondance avec la requête, et la recherche est accélérée. En revanche, le processus de clustérisation lui-même à la mauvaise réputation d'être lent lorsqu'il est appliqué à de grands corpus, et de ne pas toujours améliorer les performances de la recherche [SM83].

Un corpus peut être clustérisé sous forme de partition, ou sous une forme hiérarchique [EW89, Wil88]. Dans ce travail, nous l'utilisons comme un outil permettant d'organiser le corpus en construisant deux classifications différentes – une hiérarchie thématique et une hiérarchie visuelle, et nous considérons que des liens peuvent être établis entre elles, établissant ainsi une synergie entre deux points de vue différents sur le corpus. Nous utilisons la clustérisation agglomérative hiérarchique, malgré sa complexité algorithmique, car elle autorise différentes techniques de recherche, elle est simple à implémenter, et elle s'est montrée efficace dans le domaine de la recherche d'information [Rij79, CKPT92]. Une méthode accélérée [SM83] permet de construire en une seule passe une structure hiérarchique, tout en contrôlant la taille des clusters. Aux niveaux inférieurs de la hiérarchie, les clusters ne contiennent qu'une seule image. Au niveau le plus haut, le cluster constituant la racine de la hiérarchie contient toutes les images du corpus. A chaque itération du processus de clustérisation, les deux clusters jugés les plus similaires sont fusionnés, et un nouveau représentant est calculé. Cela suggère une méthode simple pour privilégier le rappel ou la précision<sup>3</sup> : les clusters formés durant les premières itérations contiennent des images très similaires, assurant ainsi la précision des résultats lorsqu'un cluster est sélectionné. Cette précision décroît au fur et à mesure que la clustérisation progresse, si bien que l'on peut décider d'arrêter le processus quand on juge que la précision n'est plus suffisante. La détermination d'un tel seuil n'est cependant pas aisée. Cette méthode de clustérisation aboutit à une structure hiérarchique appelée «dendrogramme». Avec un

---

2. Le terme de «clustérisation» a été choisi comme traduction du terme anglais «clustering».

3. La *précision* est définie comme la proportion de documents retrouvés et pertinents, tandis que le *rappel* est la proportion de documents pertinents qui sont effectivement retrouvés.

corpus de  $N$  images, on obtient  $N - i$  clusters à la  $i$ -ème itération : on peut ainsi aisément contrôler le nombre de clusters.

## 2.2 Description thématique des images

Notre corpus est composé d'environ 2500 images hétérogènes, dont 30% sont indexées thématiquement (les images et les index sont un sous-ensemble de ceux précédemment utilisés pour les travaux exposés dans [SQ96]). L'indexation peut faire usage de termes différents, faisant pourtant référence à une thématique commune, pour exprimer certaines nuances. Nous devons donc disposer d'une mesure de similarité thématique  $sim(a,b)$  entre deux mots  $a$  et  $b$ . Nous utilisons WordNet [Fel98] comme source de connaissance permettant de calculer une telle mesure : il s'agit d'un système de référence lexical de langue anglaise, dans lequel les mots sont organisés en ensembles de mots synonymes appelés «synsets». Chaque synset représente un concept lexical et peut être connecté avec d'autres synsets par différents types de relations sémantiques. Citons les relations de spécificité/généricité, d'antonymie, etc. Nous n'utilisons que les noms, bien que WordNet traite également les verbes, les adjectifs et les adverbes. De même, nous utiliserons les relations faisant apparaître les liens de généralité/spécificité et les relations de type «partie de» (méronyme et holonyme).

La méthode de calcul de similarité est simple. Si deux mots appartiennent au même synset, ils sont considérés comme équivalents, et nous définissons  $sim(a,b) = 1$ . Sinon, nous recherchons une forme plus faible de similarité, en essayant d'établir une liaison entre deux mots en utilisant les relations sémantiques. Si  $a$  est un spécifique de  $b$ , ou l'inverse, nous définissons  $sim(a,b) = 0.75$ .

## 2.3 Description visuelle des images

Le domaine de l'analyse d'image fournit un ensemble de techniques permettant de calculer automatiquement des indices qui vont décrire les caractéristiques visuelles de chaque image du corpus. La texture, la couleur et la forme, notamment, peuvent être caractérisées par des indices [Pra91, HS92]. Tous présentent un inconvénient majeur : ils ne sont pas *robustes*, c'est-à-dire que considérant un indice visuel  $v$ , deux images visuellement assez similaires pourront donner lieu à deux valeurs numériques assez différentes pour cet indice  $v$ . A l'inverse, deux valeurs très proches pour un indice  $v$  n'assurent pas que les images à partir desquelles  $v$  a été calculé soient visuellement similaires. Le calcul d'indices invariant [RW95] par rapport à l'échelle, la rotation, la translation ou la variation d'intensité permet d'apporter une réponse partielle à ce problème, bien qu'aucune solution ne soit à ce jour universellement applicable [MCL97]. En effet, certains indices tendent à être très efficaces sur un type d'image particulier, tandis que leur performance se dégrade avec d'autres images. Les images de notre corpus ne font pas partie d'un domaine d'application précis, ce qui interdit toute utilisation de connaissance liée au type des images traitées. Les indices que nous utilisons sont donc volontairement très génériques.

Pour modéliser l'image, c'est-à-dire en donner une description aussi précise que possible, plusieurs

modèles de haut niveau ont été élaborés. Ils permettent de mémoriser le plus souvent à la fois des informations numériques (les indices) et des connaissances plus symboliques (l'identification des objets présents dans l'image, des relations spatiales entre objets, etc.). Citons les modèles EMIR [Mou93], EMIR-2 [Mec95a], le modèle de Meghini [Meg95], de Kiyoki [KKH94], ainsi que les modèles utilisés dans les systèmes VIMSYS [GWJ91] et CORE [WDM<sup>+</sup>95]. Nous ne détaillerons pas ces modèles, qui ne sont utilisables que lorsqu'on dispose effectivement de connaissances précises sur le contenu thématique et symbolique de l'image que l'on veut modéliser. Dans ce travail, nous nous appuyons sur des structures assurant le calcul et la comparaison d'indices visuels associés à des images ou à des parties d'images. Nous présentons ci-après quelques-uns des indices les plus couramment rencontrés.

**Analyse de texture.** — La texture est calculée le plus souvent sur une image en niveaux de gris (bien qu'une image couleur soit aussi exploitable) en faisant appel à des méthodes statistiques ou encore à des méthodes structurales, qui supposent une grande régularité des textures présentes dans l'image. Parmi les indices les plus connus, citons les attributs stochastiques [HS80, HY82, PH95]; les matrices de co-occurrences [HSD73] ou de co-occurrences généralisées [DJA79, WC92]; les approches spectrales, les indices multi-échelle, souvent fondés sur les ondelettes [MM96, CK93]; et enfin les attributs géométriques [CNT95, DP96].

**Analyse de couleur.** — La couleur est représentée à l'aide d'un vecteur dans un espace couleur donné. L'espace le plus souvent rencontré est le système RGB, utilisé notamment par les écrans des ordinateurs. D'autres modèles, liés à la perception humaine de la couleur (HSV) ou indépendants des matériels utilisés ( $L^*a^*b^*$ ,  $L^*u^*v^*$ ) existent. Dans le domaine de la couleur, il n'existe pas non plus de modèle universel et applicable avec la même efficacité à tout type d'application. Les histogrammes de couleurs constituent l'indice le plus simple à calculer, et ils ont été abondamment utilisés dans les applications d'analyse et de recherche d'image [SB91]. Il a été également suggéré d'utiliser des moments statistiques [SO95], ou encore des indices invariants par rapport aux variations d'illumination d'une scène [FCF96, HS94].

**Analyse de forme.** — Une forme peut être décrite à l'aide de nombreux indices différents [MKF97, Lon98]. Le réel problème est lié à la détermination de ce qu'est effectivement une forme «intéressante» aux yeux de l'utilisateur. On peut utiliser une frontière enveloppant la forme, ou bien un objet solide, une forme géométrique correspondant à la forme. Dans le premier cas, les coefficients de Fourier ou des indices géométriques peuvent être calculés sur certains points de la frontière [SM97]. Dans le deuxième cas, on peut utiliser les moments, qui caractérisent la distribution d'une variable aléatoire... Il existe de nombreuses possibilités [PR92].

## 2.4 Une intégration thématico-visuelle pour la description d'image

Lorsque les images ont été caractérisées de manière à la fois thématique et visuelle, nous organisons le corpus dans le but de faciliter la comparaison des images entre elles, et a fortiori la recherche.

Considérons un exemple simple où l'utilisateur recherche des images représentant un paysage maritime. Sa requête va probablement consister en un mot-clé «mer», ce qui provoquera la récupération de toutes les images indexées par ce mot-clé (ou des mots-clés similaires au sens de la mesure de similarité thématique évoquée dans la section précédente). Les images non indexées textuellement ne le seront cependant pas. Supposons maintenant que nous avons le moyen de savoir de quelles manières est représentée la mer dans notre corpus, en terme d'indices visuels. Le système dispose alors de suffisamment d'information pour effectuer une nouvelle recherche fondée non plus sur les mots-clés indexant l'image, mais sur les indices visuels qui ont été calculés sur chaque image. Un même concept – «mer» ici – peut être représenté de multiples façons, ce qui suggère la mémorisation de plusieurs versions, c'est-à-dire de plusieurs groupes d'indices. Pour modéliser cela, nous proposons le concept de **réalisation**, dérivé du «*schema*» initialement proposé par [DCB<sup>+</sup>89]. Pour nous, la réalisation sera *un ensemble d'indices visuels apparaissant simultanément dans une image ou une partie d'image, et éventuellement associé à un mot-clé.*

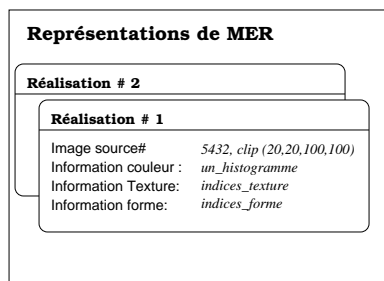


FIG. 1 – *Les représentations du mot «mer».*

Une réalisation mémorise différents indices calculés à partir d'un morceau d'image, mais un même type d'indice apparaît une seule fois dans la définition (c'est-à-dire que pour une réalisation  $r$ , on ne mémorise qu'un seul histogramme de couleur, par exemple). La zone de l'image ayant permis de calculer les indices est mémorisée également, pour que l'utilisateur puisse visualiser directement la réalisation. Puisque les réalisations peuvent être associées à un mot-clé, nous pouvons dire que l'ensemble de toutes celles qui sont associées à un même mot-clé (ou concept) constitue pour le système une sorte de connaissance à propos des propriétés visuelles de la représentation de ce concept dans le corpus. La figure 1 illustre cette définition. L'association d'un mot-clé à des réalisations doit être vue comme un processus d'indexation visuelle, qui reste cependant manuel (l'utilisateur doit déterminer la zone à caractériser, et décider du mot-clé à lui associer).

La réalisation est utilisée pour la comparaison d'image, c'est-à-dire pour établir si une image est visuellement similaire à une autre. En pratique, il s'agit d'évaluer dans quelle mesure les indices contenus dans la réalisation que l'on considère se retrouvent dans les images du corpus. Une réalisation est une agrégation de plusieurs indices, et chaque indice dispose de sa propre méthode de comparaison. Il est donc nécessaire d'élaborer une mesure globale de similarité, et nous proposons une somme pondérée :

$$s_g(a,b) = \frac{1}{N} \sum_{f=1}^N w_f * s_f(a,b)$$

Chaque indice  $f$  calcule une mesure de similarité propre  $s_f$ , et participe à la mesure globale  $s_g$  selon un poids  $w_f$  ( $\sum_f w_f = 1$ ). Cette méthode permet de donner plus d'importance aux indices jugés les plus pertinents.

## 2.5 Organisation du corpus

A partir d'un ensemble de propriétés thématiques et visuelles d'images et de leurs mesures de similarité associées, des clusters d'images peuvent être construits. La clustérisation peut ne pas être optimale, puisque nous ne disposons pas de connaissances précises sur le type des images du corpus, qui nous permettrait d'optimiser le processus. On formule cependant les hypothèses suivantes :

- La clustérisation va permettre une accélération du processus de recherche, puisque les requêtes ne sont mises en correspondance qu'avec les représentants (centroïdes) des clusters, et que seules les images contenues dans les clusters ainsi pré-sélectionnés sont analysées plus précisément.
- Les images visuellement similaires sont systématiquement clustérisées, et il suffit qu'une seule image dans un cluster soit également thématiquement indexés pour que toutes les images du cluster aient une chance d'être retrouvées et présentées à l'utilisateur, selon une méthode que nous présentons dans la section suivante.

L'organisation initiale du corpus se décompose en deux phases :

1. **Clustérisation thématique.** — Seules les images textuellement indexées sont prises en compte ici. WordNet est utilisé pour calculer les similarités thématiques entre les clusters, comme nous l'avons exposé dans la section 2.2. Cette phase aboutit à une hiérarchie d'images particulièrement importante, puisque nous donnons la priorité à la recherche thématique.
2. **Clustérisation visuelle.** — Les images sont classées à partir de leur similarité visuelle pour construire une seconde hiérarchie. Actuellement, la similarité entre deux images est fondée sur une comparaison des couleurs présentes dans les images, car l'histogramme de couleur se prête bien à un calcul rapide, et à la fusion<sup>4</sup>.

Par construction, les images indexées textuellement vont apparaître dans les deux hiérarchies, qui se retrouvent ainsi connectées grâce à ces images, comme l'illustre la figure 2. Dans cet exemple, deux seuils ont été définis, ce qui donne six clusters pour le dendrogramme thématique ( $\{C,E\}$ ,  $\{G,H\}$ ,

---

4. En effet, la réunion de deux clusters impose le calcul d'un nouveau représentant pour le cluster, et donc la combinaison des deux indices des clusters sources.



$\{B\}$ ,  $\{F\}$ ,  $\{A\}$ ,  $\{D\}$ ) et trois clusters pour le dendrogramme visuel ( $\{A,H\}$ ,  $\{B,F\}$ ,  $\{D,G,E,C\}$ ). On peut déduire de ce schéma que l'image  $H$  est thématiquement similaire à  $G$ , et qu'elle est visuellement similaire à l'image  $A$ . Nous verrons plus tard comment ce «tunnel» peut être exploité dans le processus de recherche.

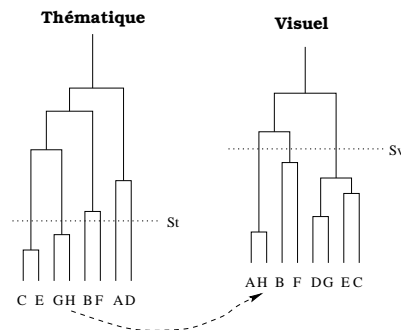


FIG. 2 – Les dendrogrammes thématique et visuels du corpus.

### 3 Un processus de recherche adapté

Thématiquement, notre corpus est partiellement indexé; visuellement tous les indices que nous calculons restent incertains, en raison du problème de la robustesse. Pour pallier à ce manque de précision, un processus adapté de recherche est proposé; il tente de profiter au mieux des interactions avec l'utilisateur. Nous décrivons ici comment une requête peut être construite, comment elle est exploitée, et comment l'utilisateur peut exprimer des jugements par rapport aux images qui lui sont proposées. Nous montrons également de quelle manière le système peut exploiter les jugements de l'utilisateur pour construire une nouvelle requête.

#### 3.1 Formulation de la requête

Le vocabulaire pouvant être utilisé en tant qu'élément de requête est limité aux noms de WordNet. Une interface graphique permet à l'utilisateur de visualiser ces mots, de suivre les liens sémantiques retenus (généricité, spécificité, «partie de»), et de choisir les mots à ajouter à la requête. Chaque élément de la requête peut être contraint: «absolument», «plutôt», «indifféremment», «présent» ou «absent». Il s'agit là d'une expression purement thématique de la requête, car aucune indication n'est donnée quant à l'aspect visuel des images à rechercher. Cependant, des réalisations sont associées à certains mots-clés, et l'utilisateur peut les visualiser et en choisir une. Dans ce cas, les conditions visuelles exprimées par la réalisation choisie sont prises en compte.

Cette première version de la requête va être soumise au système pour une première recherche. Les images présentées à l'utilisateur peuvent être jugées par lui. La connaissance de cet avis va permettre d'affiner la requête lors d'un processus de bouclage de pertinence. En particulier, elle sera enrichie par des éléments visuels extraits des images présentées à l'utilisateur. Ainsi la requête

peut-elle se composer d'une partie thématique et d'une partie visuelle.

### 3.2 Recherche dans la base d'images

Le processus de recherche se décompose en trois phases : l'extraction thématique et l'extraction visuelle selon deux méthodes (cf. figure 3).

–1– **Recherche thématique.** — La hiérarchie thématique est explorée en considérant la partie thématique de la requête (1), ce qui permet de retrouver le cluster  $C1$ , qui contient en particulier l'image  $A$ . Selon la taille du cluster (et donc la précision thématique du cluster), toutes les images ne sont pas également pertinentes par rapport à la requête, et cela motive un ordonnancement des images (2).

–2– **Recherche visuelle par les tunnels.** — La hiérarchie visuelle est ensuite explorée en utilisant comme point de départ les images retrouvées à l'étape précédente. Par les «tunnels» (c'est-à-dire le lien établi par la présence d'une même image dans les deux hiérarchies), on accède à certains clusters de la hiérarchie visuelle. Dans notre exemple,  $A$  appartient aussi au cluster  $C2$  (3). Toutes les images de ce cluster sont visuellement similaires, mais ne le sont pas pas nécessairement thématiquement. Certaines images sont indexées, ce qui permet de les ordonner grâce à la partie thématique de la requête (4).

–3– **Recherche visuelle par les réalisations.** — L'étape précédente a retrouvé des images pouvant *éventuellement* correspondre à la partie visuelle de la requête. Ici, nous recherchons des images en nous fondant uniquement sur cette partie visuelle. Cette recherche (7) retrouve le cluster  $C3$ <sup>5</sup>.

Toutes les images retrouvées en (2), (4) et (5) peuvent maintenant être ordonnées visuellement (6) grâce à la partie visuelle de la requête, et être rassemblées en une suite ordonnée d'images à présenter à l'utilisateur. La formule suivante est utilisée pour ordonner les images :  $S = w_t S_t + w_v S_v$ , avec  $w_t + w_v = 1$ . Les scores visuel ( $S_v$ ) et thématique ( $S_t$ ) peuvent être pondérés de manière à donner plus ou moins d'importance à l'une ou à l'autre des méthodes de recherche.

Le tableau 1 résume les avantages et les inconvénients de chaque étape du processus de recherche. L'intérêt de cette approche est d'une part que les images non indexées peuvent être retrouvées. D'autre part, lorsqu'une image non indexée a été retrouvée, visualisée et jugée par l'utilisateur, alors certaines suppositions peuvent être formulées quant à la thématique représentée dans l'image. En effet, si l'image a été jugée pertinente d'un point de vue thématique, on peut générer une hypothèse d'indexation. Il est souhaitable cependant que la requête soit suffisamment précise et concise, de manière à ne pas sur-indexer l'image avec des mots-clés non pertinents. En ce sens, le système améliore incrémentalement l'indexation des images, ce qui doit conduire à moyen terme à réduire le bruit et à augmenter la précision des recherches.

---

5. Notons que les clusters  $C3$  et  $C2$  sont *distincts*, bien qu'une seule représentation apparaisse sur la figure 3 pour des raisons de lisibilité.

Phase	Avantages	Inconvénients
-1-	Retrouve des images thématiquement similaires. Prend en compte des relations sémantiques.	Le problème de l'ambiguïté entre les mots n'est pas abordé.
-2-	Ebauche une première recherche visuelle alors qu'aucune condition visuelle n'apparaît dans la requête. Découvre d'éventuelles nouvelles représentations visuelles d'un concept exprimé dans la requête, enrichissant ainsi la prochaine recherche.	Le bruit : le rappel est favorisé par rapport à la précision, puisque nous considérons des images non indexées textuellement, et donc potentiellement non pertinentes par rapport à la requête.
-3-	Retrouve des images non indexées isolées, en considérant uniquement la partie visuelle de la requête.	Là encore, on ne peut pas assurer la pertinence thématique des images retrouvées.

TAB. 1 – *Analyse des étapes du processus de recherche.*

### 3.3 Interaction avec l'utilisateur : le bouclage de pertinence

Dans notre contexte, les jugements que l'utilisateur porte sur les images sont particulièrement importants. En effet, le processus de recherche a pu retrouver des images pour lesquelles un point de vue seulement (thématique ou visuel) était pertinent ; l'utilisateur va en revanche pouvoir juger les images retrouvées selon les deux points de vue. L'analyse de ces avis est le seul moyen de s'assurer de la pertinence des images retrouvées, afin de permettre une reformulation correcte de la requête. Pour exprimer cet avis, nous proposons un jugement à trois facettes.

**Le thème.** – C'est la facette la plus importante puisqu'on privilégie la thématique par rapport à l'aspect visuel de l'image. Elle permet de déterminer les thèmes à rechercher et à éviter. Une image peut représenter plus de thèmes que ceux qui constituent la requête. L'injection ou le retrait de ces nouveaux thèmes dans la requête seront assurés par le mécanisme de reformulation, au vu des jugements de l'utilisateur.

**La couleur.** – Nous l'avons choisie parce qu'elle est l'indice visuel le plus représentatif parmi ceux que nous avons implémentés. C'est une propriété très importante de l'image, et l'oeil humain y est particulièrement sensible. La couleur présente de plus l'avantage d'avoir un grand pouvoir discriminant [SB91].

**La composition.** – La composition correspond à l'aspect général de l'image, en termes de disposition spatiale de zones homogènes de couleur ou de texture sur le canevas qu'est l'image. C'est un moyen commode de capturer les formes, sans pour autant tenter d'associer une forme à un objet du monde réel. Il est en effet très difficile d'isoler un objet dans une scène.

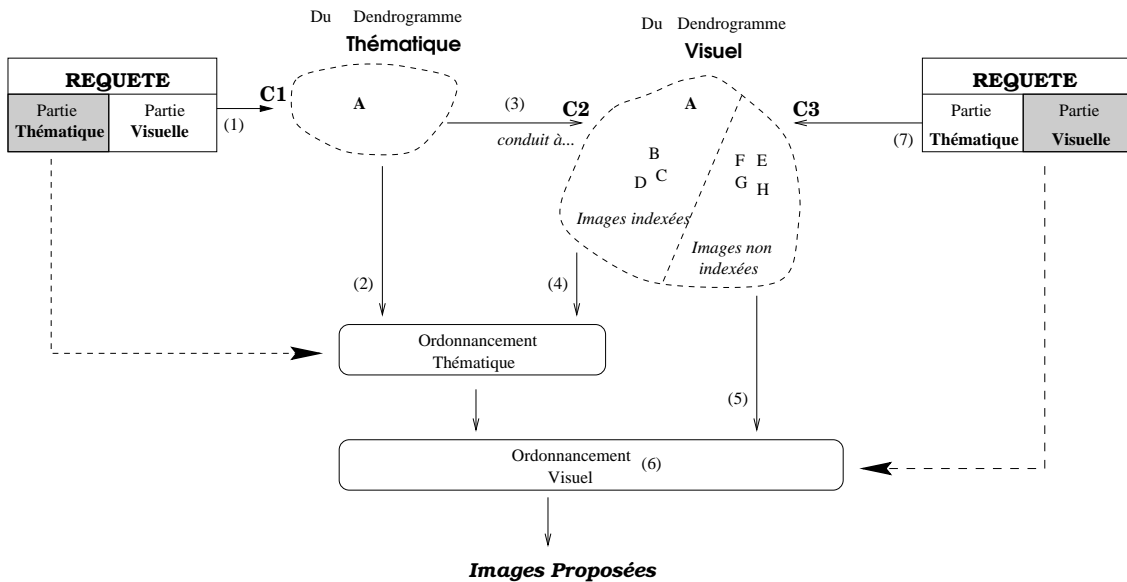


FIG. 3 – *Le processus de recherche.*

Pour chaque facette, l'utilisateur peut accepter, rejeter ou encore ne pas avoir d'avis sur l'image considérée. Ce vote peut être modéré par un coefficient allant de un à dix.

### 3.4 Comprendre le besoin d'information de l'utilisateur

A partir de cette collection de jugements, le système a pour tâche de se construire une représentation du besoin de l'utilisateur. Chaque itération «recherche-présentation-jugement-reformulation» apporte plus d'information sur les thèmes que l'image devrait représenter pour répondre à la demande, sur l'aspect que l'image devrait avoir. En pratique, c'est une information aussi discriminante que possible que l'on souhaite, de manière à rechercher et à ordonner correctement les images du corpus. Il s'agit donc de trouver deux ensembles de thèmes et d'indices visuels, l'un étant commun au groupe des images pertinentes, et l'autre étant commun à celui des images à éviter.

Pour modéliser ces ensembles, nous introduisons le concept d'**image virtuelle**, défini comme une collection de réalisations pondérées. Les éléments de requête purement thématique, purement visuel, et les éléments mixtes peuvent tous être représentés par une réalisation. A partir des jugements de l'utilisateur et des indices visuels correspondant aux images retrouvées, on peut construire deux ensembles de réalisations. La figure 4 les visualise, l'un correspondant aux caractéristiques des images à rechercher, et l'autre à celles des images à éviter.

Durant la session de recherche, plusieurs itérations de bouclage de pertinence vont avoir lieu, chacune apportant de nouvelles réalisations. Il est important de compacter ces réalisations, notamment en fusionnant celles qui sont très similaires. Les réalisations sont responsables du déclenchement et de la mise en oeuvre de cette opération.

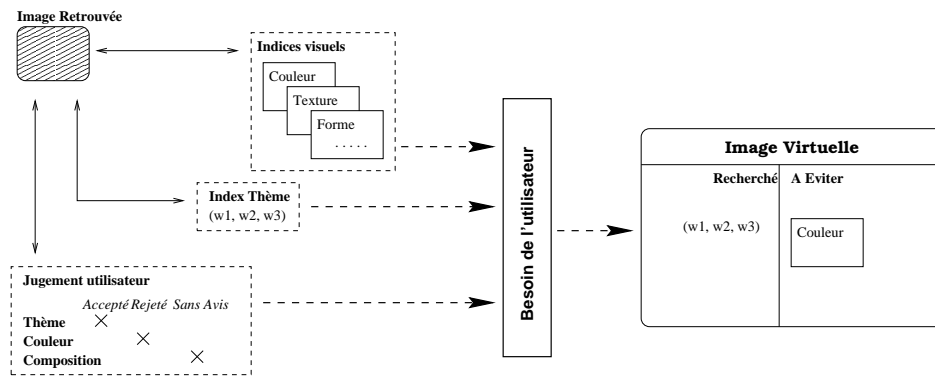


FIG. 4 – Le processus de construction de l'image virtuelle.

### 3.5 Reformulation de la requête

La reformulation des requêtes est une tâche délicate, en partie à cause de la combinaison de conditions thématiques et visuelles dans la requête.

#### 3.5.1 Reformulation thématique

Nous avons implémenté deux stratégies différentes de reformulation. La première est bien adaptée à une recherche de type exploratoire, c'est-à-dire pour connaître progressivement les images du corpus. En effet, seule l'information issue des images présentées et jugées par l'utilisateur est prise en compte au niveau de la reformulation pour construire une nouvelle requête qui «oublie» ainsi la précédente. Cette stratégie suggère que toutes les images pertinentes ont été retrouvées et visualisées, et également que l'utilisateur a jugé au moins l'une de ces images !

La seconde stratégie tient compte de la requête précédente, et s'attache à modifier les poids associés aux différents thèmes en analysant les jugements de l'utilisateur. C'est l'implémentation d'un bouclage de pertinence classique. Les jugements de rejets doivent être traités avec un soin tout particulier, de manière à ne pas introduire trop de clauses restrictives dans l'expression de la requête. Nous proposons un schéma de pondération empirique qui assure des variations souples des poids associés à chaque thème. Le tableau 2 donne les quatre situations possibles, ainsi que la formule de pondération associée ( $nw$  est le nouveau poids calculé à partir du poids précédent  $w \in [1,10]$ ). Les poids sont soit augmentés, soit diminués selon que l'utilisateur continue à vouloir ou à rejeter un thème. Ils sont réinitialisés à une valeur par défaut s'il change d'avis.

<i>Requête actuelle</i>	<i>Requête précédente</i>	
	thème accepté	thème rejeté
Th. accepté	$nw = nw * (1 + \frac{\log(1+w)}{\log 13})$	$nw = \frac{\log 1+w}{\log 13}$
Th. rejeté	$nw = -\frac{\log(1+w)}{\log 13}$	$nw = nw * (1 + \frac{\log(1+w)}{\log 13})$

TAB. 2 – Une stratégie de repondération.

### 3.5.2 Reformulation visuelle



La reformulation visuelle est beaucoup plus complexe. Chaque type d'indice a sa propre méthode pour modifier ses composants, en fonction des jugements de l'utilisateur. Nous soulignons ici quelques principes essentiels.

**La couleur** se prête bien à la reformulation, du moins lorsqu'elle est représentée par un histogramme de couleur. Ici, nous avons calculé pour tout le corpus un ensemble de 128 couleurs représentatives à l'aide d'une méthode de quantification. Ainsi, les couleurs à rechercher ou à éviter peuvent être représentées par un vecteur admettant une valeur normalisée positive ou négative qui indique l'importance relative de la couleur. Ce vecteur est modifié en fonction des jugements de l'utilisateur, de manière à accorder plus au moins d'importance à une couleur donnée.

**Les autres indices** ne sont pas aussi aisés à manipuler. En effet, pour des indices plus «élaborés» qu'une simple distribution, des opérations comme la moyenne peuvent n'avoir aucun sens. Considérons deux indices caractérisant deux formes : rien ne permet de penser a priori que la moyenne des deux indices va être un indice de la combinaison des deux formes ! Nous choisissons donc d'adopter une stratégie simple consistant à maintenir une liste d'indices, et à ne conserver qu'un exemplaire des indices qui sont suffisamment similaires. Nous avons introduit ce mécanisme en évoquant la fusion de deux réalisations.

## 4 Expérimentations

Nous présentons dans cette section les résultats de nos expérimentations relatives à l'évaluation des mesures de similarités visuelle et thématique. Nous considérons deux sessions de recherche  $s1$  et  $s2$ . Les images du tableau 3 ont été retrouvées en utilisant une requête composée d'un seul mot-clé. Les images ont été jugées pertinentes par l'utilisateur, ce qui a eu pour conséquence la construction d'une image virtuelle avec la réalisation associée. La figure 4 illustre ce processus. Nous allons utiliser dans les expériences décrites ci-dessous cette réalisation, afin de vérifier qu'elle permet bien de retrouver d'autres images similaires, en se fondant uniquement sur des critères visuels. Les cinq images les plus proches sont présentées avec leur distance normalisée par rapport à la requête.

Session	$s1$	$s2$
Requête	“rock”	“train”
Image		

TAB. 3 – *Les images test pour la recherche.*

## 4.1 Evaluation de la similarité thématique

Les images du tableau 3 ont été retrouvées avec la requête thématique correspondant au mot-clé affiché avec l'image. Nous nous intéressons ici aux inconvénients que l'on peut rencontrer avec ce type de recherche.

WordNet nous a permis de calculer une comparaison thématique simple, mais les ambiguïtés liées à la polysémie demeurent : il a été démontré que la gestion de l'ambiguïté en recherche d'information est un problème délicat [SQ96], et nous avons décidé de ne pas le gérer. Cela conduit naturellement à la dégradation de la précision des résultats, car les images ont été clustérisées sur le critère d'un mot identique, mais ayant deux sens différents.

La qualité de l'indexation peut elle aussi poser problème. Notre corpus est indexé de manière succincte ; les images sont le plus souvent accompagnées de trois ou quatre mots pour décrire leur contenu. En conséquence, les comparaisons thématiques s'effectuent sur une information peu précise, et cela peut gêner le processus de clustérisation. En effet, des images ayant peu ou pas de thèmes en commun vont être fusionnées plus rapidement, c'est-à-dire au cours des premières opérations de clustérisation. C'est un comportement normal, qui illustre de quelle manière l'indexation affecte la clustérisation, ainsi que la taille et le nombre des clusters «thématiquement compacts», qui sont les plus intéressants puisqu'ils permettent une recherche plus précise. Notons cependant que les images extraites des clusters sont ordonnées en fonction des conditions exprimées dans la requête, ce qui améliore sensiblement la précision de la recherche en limitant le bruit.

## 4.2 Evaluation de la similarité visuelle

Les indices visuels vraiment utiles pour la recherche d'image sont ceux qui sont proches de la perception humaine. Cela permet en effet de prendre en compte de manière plus précise les jugements des utilisateurs [PM95]. Nous avons choisi des indices de couleur, de texture, et de forme. Pour permettre une comparaison de la *composition* de l'image, nous calculons ces indices à la fois sur la totalité de l'image (quand cela a un sens) et sur des portions d'images. L'image est divisée en portions contiguës de 32x32 pixels – que nous appellerons «imassettes» –, autorisant ainsi une caractérisation locale plus précise, car un indice calculé sur l'image entière «dilue» l'information.

Il est clair que les indices exploités séparément ne permettent pas de comparer de manière exacte différentes images. Ceci a donné lieu à des travaux [MP97] implémentant une «société de modèles» visant à déterminer automatiquement quelle combinaison d'indices est la plus performante. Nous adoptons pour notre part une combinaison statique.

### 4.2.1 Caractéristiques de texture

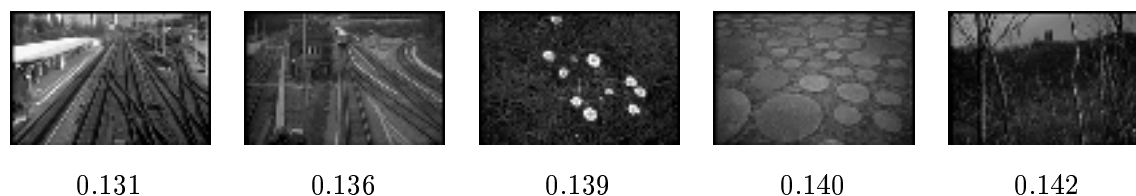
Pour caractériser la texture, les indices introduits par Haralick sont utilisés [HSD73]. Au nombre de quatorze, ils sont calculés selon quatre orientations différentes sur chaque imasette, et pas au

niveau de l'image entière. Ces indices caractérisent surtout des micro-textures, et les calculer au niveau d'une portion trop grande d'image ne présente aucun intérêt.

L'évaluation de ces indices montre qu'ils sont suffisamment puissants pour fournir une caractérisation grossière de l'image lorsque les images sont divisées en imasettes, et quand la composition des images comparées est proche. Le tableau 4 laisse supposer que c'est le large ciel homogène qui a permis de retrouver les images. Pour les scènes plus complexes, les résultats restent assez satisfaisants (tableau 5). Néanmoins, un problème peut survenir lors de la comparaison des images, en raison du caractère très systématique de notre méthode de division de l'image en imasettes, qui gère mal les images similaires numérisées à des échelles différentes. En effet, la grille est figée, et on ne se préoccupe pas de changer la taille des imasettes, ou bien de faire varier leur position pour optimiser le calcul d'un indice. Adopter une telle méthode aurait certes amélioré la précision des indices (qui auraient été calculés sur des zones homogènes, au sens de l'indice), mais cela aurait rendu le processus de comparaison de deux images bien plus complexe.



TAB. 4 – *Test 1 de recherche par la texture.*



TAB. 5 – *Test 2 de recherche par la texture.*

#### 4.2.2 Caractéristiques de couleur

Pour le calcul des indices de couleurs, le corpus tout entier a été analysé de manière à sélectionner 128 couleurs représentatives. Les images sont alors transformées de manière à utiliser uniquement ces 128 couleurs, après quoi des histogrammes de couleurs sont calculés au niveau global (un pour l'image entière) et local (un par imasette). Ils sont codés sous forme de vecteurs de 128 éléments. Les couleurs sont exprimées dans un espace particulier, le modèle  $L^*u^*v^*$ , qui a été conçu pour faciliter les comparaisons de couleurs, en permettant notamment l'utilisation de la distance Euclidienne.

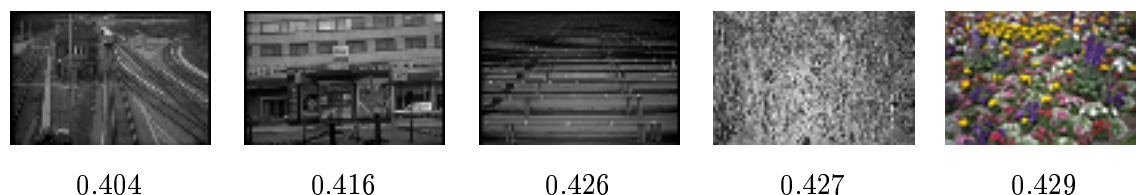
Le tableau 6 montre les résultats obtenus avec les vecteurs de couleurs globaux. On utilise les indices globaux lorsque le jugement de l'utilisateur porte seulement sur la *couleur*. En revanche, s'il a jugé la *composition* de l'image, alors la localisation des couleurs sur l'image est prise en compte en utilisant les vecteurs locaux de couleur ; les résultats d'une recherche de ce type sont présentés sur le



tableau 7). Il apparaît que les vecteurs globaux sont plus performants, ce qui nous incite à chercher à améliorer la caractérisation locale de la couleur, en tentant par exemple une comparaison fondée à la fois sur la couleur et sur la texture ; les premiers essais visant à combiner les deux méthodes semblent prometteurs.



TAB. 6 – *Test de recherche par la couleur (globale).*



TAB. 7 – *Test de recherche par la couleur (local).*

### 4.2.3 Caractéristiques de forme

Nous avons implanté des indices de forme fondés sur des *moments statistiques invariants* [Hu62]. La principale difficulté de cette approche est de déterminer quelle forme doit être caractérisée. Une méthode simple d'extraction consiste à choisir  $n$  couleurs dominantes dans l'image, et à créer ainsi  $n$  versions différentes de l'image, chacune faisant apparaître uniquement, et à l'aide de pixels noirs, les pixels correspondant à l'une des  $n$  couleurs. On obtient ainsi  $n$  formes à caractériser par les indices. Les expérimentations montrent cependant que cette méthode n'est pas adaptée aux scènes complexes, dans lesquelles les «images binaires» ainsi calculées ne font pas apparaître de formes suffisamment compactes. Un pré-traitement de l'image pourrait améliorer les performances de la méthode.

## 5 Conclusion et perspectives

Nous avons présenté une méthode intégrée permettant une recherche à la fois thématique et visuelle d'images, dans le contexte d'une base partiellement indexée d'images. Notre stratégie repose sur une double caractérisation des images : l'analyse d'image permet le calcul d'indices décrivant les propriétés visuelles des images, tandis que des mots-clés éventuellement associés aux images en décrivent la thématique. Le corpus est organisé en deux hiérarchies, par un processus de clustérisation. L'intégration thématico-visuelle est réalisée au niveau corpus par le mécanisme de *réalisation*, et au niveau du système par un processus de recherche adapté. Celui-ci, quoique fortement fondé

sur une recherche thématique, va exploiter les caractéristiques visuelles des images pour établir des comparaisons complémentaires. Les utilisateurs peuvent juger les images qui leur sont présentées selon des aspects visuels et thématiques, permettant ainsi au système de reformuler les requêtes de manière plus précise lors de l'étape de bouclage de pertinence.

Pour aller plus loin, nous pensons que l'analyse des sessions de recherche peut aider à déterminer des hypothèses d'indexation intéressantes : des images non indexées ont pu être retrouvées pendant la session, et le mécanisme de bouclage de pertinence indique si l'image est pertinente par rapport à la requête. Sous certaines conditions, nous pensons qu'un label d'indexation thématique temporaire peut être associé à l'image. La confiance que le système accorde à ce label va se modifier au fil des sessions, aboutissant à un espace de document dynamique, tel qu'il avait été introduit dans [SM83].

## Références

- [AZP96] P. Aigrain, H. Zhang, and D. Petkovic. Content-Based Representation and Retrieval of Visual Media: A State-of-the-Art Review. *Multimedia Tools and Applications*, 3:179–202, 1996.
- [CK93] T. Chang and C.J. Kuo. Texture analysis and classification with tree-structured wavelet transform. *IEEE Transactions on image processing*, 2(4):429–441, October 1993.
- [CKPT92] D. Cutting, D. Karger, J. Pedersen, and J. Tukey. Scatter/Gather: A Cluster-based Approach to Browsing Large Document Collections. In *Proc. of the int. ACM SIGIR conf. on research and development in Information Retrieval*, pages 318–329, 1992.
- [CM97] Y. Chiaramella and M. Mechkour. Indexing an image test collection. Technical report, FERMI-BRA Project, 1997. (Ch. du deliverable D10).
- [CNT95] Y.Q. Chen, M.S. Nixon, and D.W. Thomas. Statistical geometrical features for texture classification. *Pattern Recognition*, 28(4):537–552, 1995.
- [DCB<sup>+</sup>89] B. Draper, R. Collins, J. Brolio, A. Hanson, and E. Riseman. The Schema System. *International Journal of Computer Vision*, 2:209–250, 1989.
- [DJA79] L.S. Davis, S.A. Johns, and J.K. Aggarwal. Texture analysis using generalized co-occurrence matrices. *IEEE Transactions on pattern analysis and machine intelligence*, PAMI-1(3):251–259, July 1979.
- [DP96] A. Del Bimbo and P. Pala. Image indexing using shape based visual features. In *ICPR'96 13th int. IAPR conf. on Pattern Recognition*, volume 3, Vienna, Austria, August 1996. IEEE.
- [EW89] A. El-Hamdouchi and P. Willett. Comparison of hierarchic agglomerative clustering methods for document retrieval. *The Computer Journal*, 32(3):220–227, 1989.

- [FCF96] G.D. Finlayson, S.S. Chatterjee, and B.V. Funt. Color angular indexing. In Bernard Buxton and Roberto Cipolla, editors, *Computer Vision – ECCV’96*, pages 16–27. Springer, Cambridge UK, 1996. (Lecture Notes in Computer Science, volume 1065).
- [Fel98] C. Fellbaum, editor. *WORDNET: An Electronic Lexical Database*. MIT Press, 1998.
- [GOC<sup>+</sup>92] C. Goble, M. O’Docherty, P. Crowther, M. Ireton, J. Oakley, and C. Xydeas. The Manchester Multimedia Information System. In *Lecture Notes in Computer Science no 580*, pages 39–55. Springer, 1992.
- [GWJ91] A. Gupta, T. Weymouth, and R. Jain. Semantic queries with pictures: the VIMYS model. In *Proceedings of the 17th int. conf. on Very Large Data Bases*, pages 69–79, Barcelona, September 1991.
- [Hal90] G. Halin. Machine Learning and Vectorial Matching for an Image Retrieval Model: EXPRIM and the System RIVAGE. In J.-L. Vidick, editor, *ACM 13th Int. Conf. on Research and Development in Information Retrieval*, pages 99–114, Brussels (Belgium), september 1990. Presses Universitaires de Bruxelles.
- [HS80] M. Hassner and J. Sklansky. The use of markov random fields as models of texture. *Computer Graphics and Image Processing*, 12:357–370, 1980.
- [HS92] R.M. Haralick and L.G. Shapiro. *Computer and robot vision*. Addison-Wesley, 1992.
- [HS94] G. Healey and D. Slater. Using illumination invariant color histogram descriptors for recognition. In *IEEE Proc. of Conf. on Computer Vision and Pattern Recognition*, pages 355–360, Seattle, WA, 1994.
- [HSD73] R.M. Haralick, K. Shanmugam, and I. Dinstein. Textural features for image classification. *IEEE Trans. on Systems, Man, and Cybernetics*, SMC–3(6):610–621, November 1973.
- [Hu62] M. K. Hu. Visual pattern recognition by moment invariants. *IRE Trans. Inform. Theory*, IT-8:179–187, 1962.
- [HY82] H. Haneko and E. Yodogowa. A markov random field application to texture classification. In *Proc. Pattern Recognition and Image Processing*, pages 221–225, Las Vegas, Nevada, June 14–17 1982.
- [Kat92] T. Kato. Database architecture for content-based image retrieval. In *Storage and Retrieval for Image and Video Databases*, pages 112–123, 1992.
- [KKH94] Y. Kiyoki, T. Kitagawa, and T. Hayama. A metadatabase system for semantic image search by a mathematical model of meaning. *SIGMOD RECORD*, 23(4):34–41, December 1994.
- [Lon98] S. Loncaric. A survey of shape analysis techniques. *Pattern Recognition*, 31(8):983–1001, 1998.

- [MCL97] M. De Marsicoi, L. Cinque, and S. Levialdi. Indexing pictorial documents by their content: a survey of current techniques. *Image and vision computing*, 15(2):119–141, february 1997.
- [Mec95a] M. Mechkour. EMIR2: An Extended Model for Image Representation and Retrieval. In *Database and Expert Systems Applications (DEXA)*, pages 395–404, London, United Kingdom, 1995. Springer. Also in *Lecture Notes in Computer Science*, vol. 978.
- [Mec95b] M. Mechkour. *Un modèle étendu de représentation et de correspondance d'images pour la recherche d'informations*. PhD thesis, Université Joseph Fourier, Grenoble I, 1995.
- [Meg95] C. Meghini. An image retrieval model based on classical logic. In *Proc. of the int. ACM SIGIR conf. on research and development in Information Retrieval*, pages 300–308, Seattle, WA, 1995.
- [MKF97] B. Mehtre, M. Kankanhalli, and W. Foon Lee. Shape measures for content based image retrieval: a comparison. *Information processing and management*, 33(3):319–337, 1997.
- [MM96] B. Manjunath and W. Ma. Texture features for browsing and retrieval of image data. *IEEE Transactions on pattern analysis and machine intelligence*, 18(8):837–842, august 1996.
- [Mou93] N. Mouaddib. EMIR. Technical report, CRIN, Nancy, France., 1993.
- [MP97] T.P. Minka and R.W. Picard. Interactive Learning with a "Society of Models". *Pattern Recognition*, 30(4):565–581, 1997.
- [NBE<sup>+</sup>93] W. Niblack, R. Barber, W. Equitz, M. Flickner, E. Glasman, D. Petkovic, P. Yanker, C. Faloutsos, and G. Taubin. The QBIC project: querying images by content using color, texture and shape. In Wayne Niblack, editor, *Storage and Retrieval for Image and Video Databases*, pages 173–181, San Jose, CA, 1993. SPIE.
- [OS95] V. E. Ogle and M. Stonebraker. CHABOT: Retrieval from a relational database of images. *IEEE Computer*, 28(9):40–48, September 1995.
- [PH95] D. Panjwani and G. Healey. Markov random field models for unsupervised segmentation of textured color images. *IEEE Transactions on pattern analysis and machine intelligence*, 17(10):939–954, 1995.
- [PM95] R.W. Picard and T.P. Minka. Vision Texture for Annotation. *Multimedia Systems*, 3:3–14, 1995.
- [PR92] R.J. Prokop and A.P. Reeves. A survey of moment-based techniques for unoccluded object representation and recognition. *CVGIP: Graphical Models and Image Processing*, 54(5):438–460, September 1992.
- [Pra91] W.K. Pratt. *Digital Image Processing*. John Wiley & Sons, New York, second edition, 1991.
- [Rij79] C.J. van Rijsbergen. *Information Retrieval*. Butterworths, second edition edition, 1979.

- [RS92] F. Rabitti and P. Savino. Querying semantic image database. In *Storage and Retrieval for Image and Video Databases*, pages 69–78, 1992.
- [RW95] E. Rivlin and I. Weiss. Local invariants for recognition. *IEEE Transactions on pattern analysis and machine intelligence*, 17(3):226–238, March 1995.
- [SB91] M.J. Swain and D.H. Ballard. Color indexing. *International Journal of Computer Vision*, 7(1):11–32, 1991.
- [SC97] J. R. Smith and S.-F. Chang. Querying by color regions using the VisualSEEK content-based visual query system. In Mark T. Maybury, editor, *Intelligent Multimedia Information Retrieval*, pages 23–41. AAAI Press, Menlo Park, 1997.
- [SM83] G. Salton and M. J. McGill. *Introduction to Modern Information Retrieval*. McGraw-Hill, 1983.
- [SM97] C. Schmid and R. Mohr. Local Grayvalue Invariants for Image Retrieval. *IEEE Transactions on pattern analysis and machine intelligence*, 19(5), 1997.
- [Sma94] M. Smail. Case-Base Reasoning Meets Information Retrieval. In *RIAO 94: Intelligent Multimedia Information Retrieval Systems and Management*, page 133, 1994.
- [SO95] M. Stricker and M. Orengo. Similarity of color images. In Wayne Niblack and R. C. Jain, editors, *Storage and Retrieval for Image and Video Databases*, pages 381–392, San Jose, CA, 1995. SPIE volume 2420.
- [SQ96] A. Smeaton and I. Quigley. Experiments on using semantic distances between words in image caption retrieval. In *Proc. of the int. ACM SIGIR conf. on research and development in Information Retrieval*, pages 174–180, 1996.
- [WC92] C.M. Wu and Y.C. Chen. Statistical feature matrix for texture analysis. *CVGIP: Graphical Models and Image Processing*, 54(5):407–419, 1992.
- [WDM<sup>+</sup>95] J.K. Wu, A. Desai-Narasimhalu, B.M. Mehtre, C.P. Lam, and Y.J. Gao. CORE: a content-based retrieval engine for multimedia information systems. *Multimedia Systems*, 3:25–41, 1995.
- [Wil88] P. Willet. Recent trends in Hierarchic Document Clustering. *Information processing and management*, 24(5):577–597, 1988.
- [WLS95] T. Whalen, E.S. Lee, and F. Safayeni. The Retrieval of Images from Image Databases. *Behaviour & Information Technology*, 14(1):3–13, 1995.