INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Projet imedia

Images et Multimédia : Indexation, Recherche et Navigation

Rocquencourt

THÈME 3B

Table des matières

1. Composition de l'équipe

Responsable scientifique
Nozha Boujemaa [Directeur de Recherche INRIA]

Assistante de projet
Laurence Bourcier [partagée avec le projet Eiffel]

Personnel INRIA
Anne Verroust [CR1]
François Fleuret [CR2]
Jean-Philippe Tarel [détaché CR2 depuis le 1/11/2001]
Michel Crucianu [détaché CR2 depuis le 1/09/2002]
Jean-Paul Chièze [IR1, à temps partiel dans le projet]

Conseiller scientifique
Donald Geman [Professeur à l'université Johns Hopkins et à l'ENS Cachan]

Collaborateur extérieur
Michel Scholl [Professeur au CNAM]

Personnel université
Valérie Gouet [MdC CNAM depuis le 01/09/2002]
Marie-Aude Aufaure [MdC université de Lyon en congé de recherche depuis le 1/09/2001]

Chercheur post-doctorant
Andreas Rauber [ERCIM depuis le 3/06/2002]

Ingénieur expert
Valérie Gouet [jusqu'au 31/08/2002]

Doctorants
Sabri Boughorbel [Bourse INRIA Rocq depuis le 1/11/2001]
Marin Ferecatu [Bourse INRIA Rocq depuis le 1/10/2001]
Julien Fauqueur [Bourse INRIA Rocq depuis le 1/03/2000]
Bertrand Le Saux [Bourse INRIA Rocq depuis le 1/11/1999]
Hichem Sahbi [Bourse de la coopération Franco-Algérienne depuis le 1/10/1999]

Stagiaires
Nizar Grira [SupCom (Tunis), avril - décembre 2002]
Moez Tarzi [Université de Versailles, mars - août 2002]
Arnaud Tournier [CNAM, janvier - septembre 2002]
Jérémie Jakubowicz [ENSAE, janvier - juin 2002]

2. Présentation et objectifs généraux

L'une des conséquences de la convivialité accrue et de la baisse des coûts des moyens informatiques est la production et l'échange de flux de plus en plus importants de documents numérisés et multimedia. Ces documents sont par essence hétérogènes, et intègrent aussi bien le texte que l'image, le graphique, la vidéo et le son.

La recherche d'informations ne pourra plus continuer à reposer uniquement sur l'information textuelle mais désormais il est indispensable qu'elle devienne plurimodale couvrant les différents aspects du contenu multimedia. En particulier, le contenu visuel occupe une place prépondérante et représente un vecteur central de transmission de l'information. La description de ce contenu par des techniques d'analyse d'images est moins subjective que la seule description habituelle par des mots clés (quand elle existe). D'autre part, étant indépendante de la langue de recherche, elle devient primordiale pour l'exploration efficace d'un flot multimedia.

A IMEDIA, nous concentrons nos efforts sur l'accès intelligent par le contenu visuel. Pour ce faire nous développons des méthodes de description et d'indexation par le contenu, de recherche interactive et de navigation dans des bases d'images, dans un contexte multimedia.

Les systèmes de recherche d'images par le contenu ont un rôle d'aide à la recherche automatique et à la décision. L'utilisateur final reste le seul maître d'oeuvre capable de prendre une décision au final. Les différents travaux de recherche menés dans ce domaine dans la dernière décennie ont avantageusement permis « la démonstration de faisabilité » de la recherche par le contenu visuel. Néanmoins, l'expérience a montre qu'il existe encore un écart (gap) d'« usage » entre les créateurs de ces techniques et méthodes et leurs potentiels utilisateurs.

A IMEDIA, un de nos objectifs est le rapprochement entre les usages réels et les fonctionnalités issues de nos activités de recherche d'informations par le contenu visuel. Nous nous attacherons donc à concevoir des méthodes et des techniques qui puissent répondre à des scénarios réalistes qui souvent présentent des défis méthodologiques fort intéressants à relever.

Parmi les objectifs d'« usage », nous citons la fonctionnalité précieuse pour l'utilisateur de pouvoir exprimer un intérêt visuel particulier pour une partie de l'image. Ceci lui permet de cibler et d'exprimer finement son intention. Un autre objectif, dans le même sens, est celui d'exprimer des préférences subjectives et de doter le système de propriétés d'apprentissage de ces préférences. Également, nous nous attacherons à développer des méthodes qui gardent un intérêt au niveau du temps de réponse pour l'utilisateur. Il est bien évident que cette appréciation dépend du domaine d'application (spécifique/générique) et du coût de l'erreur.

De ce fait, nos recherches sont à l'intersection de plusieurs disciplines scientifiques, dont les principales sont l'analyse d'images, la reconnaissance des formes, l'apprentissage, l'interaction homme-machine et les bases de données.

Notre travail de recherche s'articule autour des axes principaux suivants :

1. l'indexation des images : cette partie essentielle est fondée sur la modélisation, par des techniques d'analyse d'images, de l'apparence visuelle et permet l'élaboration automatique des signatures d'images ;
2. la classification automatique et l'apprentissage statistique : méthodes génériques et fondamentales pour la résolution des problèmes de reconnaissance des formes qui se posent d'une manière cruciale dans le contexte de l'indexation d'images ;
3. la recherche interactive et la personnalisation dans le but de prendre en compte les préférences de l'utilisateur qui expriment souvent des requêtes subjectives et/ou sémantiques ;
4. l'indexation trans-media, et en particulier l'indexation bi-modale texte/image, qui a pour objectif de faire coopérer ces deux medias pour une indexation et/ou recherche plus efficaces.

Plus généralement, l'équipe IMEDIA déploie ses efforts de recherche, de collaboration, et de transfert, pour répondre au problème complexe de l'accès intelligent aux données multimedia dans sa globalité.

3. Fondements scientifiques

3.1. Introduction

Nous regroupons les problèmes rencontrés dans le domaine de l'indexation et la recherche d'images par le contenu dans les classes thématiques suivantes : indexation d'images, classification, personnalisation et indexation trans-media. Dans ce qui suit, nous présentons une introduction à chacun de ces thèmes.

3.2. Construction et structuration de l'espace de description

Mots clés : *apparence visuelle*, *descripteurs et signatures d'images*, *indexation par le contenu visuel*, *analyse d'image*, *reconnaissance de forme*, *similarité visuelle*, *mise en correspondance*.

Participants : Nozha Boujemaa, Valérie Gouet, Julien Fauqueur, Sabri Bougorbel, Jean-Philippe Tarel, Michel Scholl, Nizar Grira, Anne Verroust, Hichem Sahbi, Jean-Paul Chièze.

Glossaire

> **Indexation par le contenu** *opération qui consiste à extraire d'un document (ici une image) des descripteurs visuels automatiques significatifs, compacts et structurés qui seront utilisés et comparés au moment de la recherche interactive.*

L'objectif d'IMEDIA est d'offrir la possibilité d'interroger les bases d'images par le contenu, d'une manière intelligente et intuitive pour l'utilisateur. Un fois posé en des termes concrets, ce problème donne naissance à un certain nombre de modélisations mathématiques et informatiques.

Pour représenter le contenu d'une image, nous recherchons une représentation compacte (moins de données, plus de sémantique), significative (relativement au contenu de l'image et aux utilisateurs de la base) et rapide à calculer et à comparer. Le choix de l'espace de représentation consiste à choisir des *attributs* significatifs de la base d'images, les *descripteurs* de ces attributs et enfin la représentation de ces descripteurs en machine en termes de *signatures* d'images.

Nous traitons aussi bien des bases d'images « génériques » dans lesquelles les images sont hétérogènes (par exemple la recherche d'images sur l'Internet) que des bases d'images « spécifiques » à un domaine d'application particulier, appelées également bases avec vérité terrain, dans lesquelles les images ont un contenu homogène (visages, images médicales, empreintes digitales, etc.).

Notons que pour les bases spécifiques, on développe des descripteurs dédiés et optimaux pour la cible considérée (reconnaissance de visages, etc.). Pour les bases génériques, à l'inverse, on extrait des descripteurs universels (couleur, texture, forme, etc.).

En plus des signatures génériques et spécifiques, la typologie des signatures s'étend aussi aux signatures locales ou globales selon que la requête s'adresse à toute l'image ou à une de ses parties. Là encore, nous pouvons distinguer les requêtes approximatives et les requêtes précises. Dans ce dernier cas, il faut être capable de fournir diverses descriptions des parties d'images ainsi que les moyens de pouvoir les considérer comme zones de requêtes. En particulier, se pose la question des mesures de similarité aussi bien locales que globales.

Nous nous sommes également posé la question de la description des modèles géométriques 3D avec l'arrivée d'Anne Verroust dans l'équipe, pour compléter ainsi les différentes facettes de description de l'apparence visuelle aussi bien 2D que 3D.

À la fin de la phase de calcul de signatures, la base d'images est représentée par un nuage de points dans un espace à dimension élevée : l'espace des caractéristiques (*« feature space »*).

Une deuxième phase de construction d'index peut s'avérer utile dans le contexte d'un espace de représentation de très grande dimension. Cela revient à pré-structurer le nuage de points des signatures d'images et à les stocker efficacement en machine, dans le but de réduire ultérieurement le coût de la requête (compromis coût du stockage/coût de la requête). Cette deuxième phase présente des problèmes communs avec ceux posés, classiquement, à la communauté « base de données » mais avec un nouveau contexte : celui des données images. La diversité des espaces de signatures que nous construisons fait qu'une étude adaptée est nécessaire pour une structuration spécifique de chacun de ces espaces. Une collaboration suivie sur les aspects constructions d'index multi-dimensionnels a été engagée avec Michel Scholl (INRIA/CNAM).

3.3. Classification automatique

Les méthodes de classification automatique et plus généralement de reconnaissance des formes (« pattern recognition ») ont un apport crucial dans la résolution des problèmes posés par l'indexation et la recherche par le contenu visuel [24] [29]. Nous distinguons les méthodes faisant appel à une phase d'apprentissage des méthodes non-supervisées. En effet, selon notre niveau de connaissance du contenu de la collection d'images nous pouvons ou non disposer d'un corpus d'apprentissage des « objets » ou des contenus visuels recherchés. Pour la détection/recherche d'objets connus à priori, des approches procédant par hiérarchies de classifieurs ont été particulièrement investies. Dans ce cadre, la détection de visages a bénéficié d'un intérêt croissant étant donné la valeur sémantique que porte cette information pour l'indexation de flots visuels. D'autres part, dans une collection dont le contenu n'est pas connu a priori, nous nous intéressons à des techniques de catégorisation automatique qui doivent être capables de générer des classes les plus cohérentes possibles avec un nombre de classes adaptatif au contenu de la base (pour des objectifs de navigation par exemple).

3.3.1. Hiérarchie de classifieurs pour la détection d'objets

Mots clés : *apprentissage statistique*, *optimisation algorithmique*, *classification supervisée*.

Participants : François Fleuret, Donald Geman, Hichem Sahbi.

La détection automatique d'objets constitue une des approches les plus directes pour indexer des bases d'images selon leurs contenus. Les techniques classiques (réseaux de neurones, machines à vecteurs de support, etc.) font de l'induction, c'est à dire de la généralisation à partir d'exemples. Les méthodes récentes d'apprentissage reposent sur des résultats théoriques de statistiques (cadre PAC [1], théorie de la minimisation du risque structurel) qui permettent de borner l'erreur de généralisation sur les vraies données à partir des estimations sur des exemples.
Les travaux que nous menons se focalisent sur l'idée de compromis entre le taux d'erreur (proportion d'objets non détectés ou de détections fausses) et le coût algorithmique. Cette approche nous a amenés à une organisation hiérarchique du processus de détection, qui permet de concentrer le calcul sur les parties complexes d'une image à traiter. Il est apparu que ces approches qui favorisent les calculs rapides ont les même performances en taux d'erreurs que des techniques plus lourdes. Le biais que nous imposons, et qui favorise les classifieurs peu coûteux, agit comme un facteur de régularisation. Il favorise les familles de classifieurs structurellement simples, qui possèdent de bonnes propriétés de généralisation.

3.3.2. Classification non-supervisée

Mots clés : *catégorisation*, *appartenance*, *nombre de classes*, *reconnaissance des formes*, *agglomération compétitive*.

Participants : Nozha Boujemaa, Bertrand Le Saux, Nizar Grira, Valerie Gouet, Julien Fauqueur.

Les méthodes de classification non-supervisée permettent la génération automatique de catégories et relèvent pour nous de méthodes de découverte de connaissances visuelles. Nous rencontrons ce problème à la fois pour :

- répondre à la difficulté de « la page zéro » par la génération de résumé visuel d'une base d'images en opérant sur l'ensemble des signatures de toutes les images ;
- calculer une catégorisation des signatures de la même image générant ainsi une segmentation en régions ;
- structurer et organiser l'espace des signatures (globales ou locales) en groupes permettant une recherche hiérarchique de similarité et ainsi plus rapide en limitant le parcours de tout l'espace au parcours des représentants de ces catégories.

Il s'agit d'un problème d'une difficulté certaine compte tenu de la complexité des espaces de signatures que nous considérons, le bruit et le recouvrement entre les classes perturbant significativement l'estimation

[1] Probably Approximately Correct

fidèle des paramètres des classes. Les questions clés qui caractérisent et dont dépend la qualité du résultat de la catégorisation sont de trois niveaux : la définition du moteur de regroupement, le critère de proximité, la modélisation des données à catégoriser.

Nous avons investi une famille de méthodes de classification par agglomération compétitive permettant de prendre en compte les contraintes de nos conditions initiales : nombre de classes inconnu à déterminer, bruit à modéliser, appartenance non-exclusive permettant de gérer le recouvrement et de différer la prise de décision le plus tard possible.

3.4. Recherche interactive et Personnalisation

Mots clés : *interaction avec l'utilisateur*, *expression des préférences*, *classification subjective*, *gap sémantique*, *boucle de pertinence*, *apprentissage statistique*.

Participants : Marin Ferecatu, Nozha Boujemaa, Julien Fauqueur, Donald Geman.

Dans cette partie nous visons toutes les approches permettant une réduction du « fossé » sémantique à travers toutes les méthodes qui font intervenir l'utilisateur pour exprimer ses préférences, personnaliser les réponses du système selon une classification subjective. La recherche interactive permet également l'expression de l'intention de l'utilisateur et de spécifier sa recherche. La première de ces approches est la recherche d'images par bouclage de pertinence (*relevance feedback* ou RF). Le principe est de fournir au système des exemples positifs (documents pertinents pour la requête en cours) et des exemples négatifs (images que le système doit éviter).

Il s'agit d'un ensemble de méthodes adaptées a la recherche de concepts sémantiques – la démarche du système est de trouver le concept unificateur dans tous les exemples donnés et lui associer une étiquette (nom) en relation avec d'autres mots.

Dans ce sens, la recherche par bouclage de pertinence est vue comme une manière d'approcher le « fossé » sémantique mais en aucun cas ne peut être considérée comme une manière d'approcher le « fossé » numérique. Nous entendons par ce dernier le manque de fidélité (l'imperfection) des descripteurs visuels au contenu visuel.

D'autres approches sont explorées, toujours en se basant sur l'expression de l'utilisateur. On peut citer, par exemple, la composition logique d'une requête visuelle par des parties d'images. Elles permettent d'exprimer l'idée de l'utilisateur de la « sémantique visuelle » recherchée.

3.5. Indexation Trans-media

Mots clés : *indexation/recherche hybride*, *annotation textuelle*, *théorie de l'information* .

Participants : Marin Ferecatu, Francois Fleuret, Valerie Gouet, Nozha Boujemaa, Sabri Boughorbel.

Nous avons décrit, jusqu'à présent, notre problématique dans le cadre de l'exploitation d'indices visuels uniquement. Lorsque des indices supplémentaires sont disponibles, leur utilisation présente un apport certain au résultat de la recherche, compte tenu de la complémentarité des sources d'informations. Parmi ces indices, on peut citer les *métadata* (nom de fichier, date de création, légende, etc.) mais aussi les annotations textuelles lorsqu'elles existent. Notons que celles-ci sont porteuses d'information de haut niveau liée à une sémantique et à une forte connaissance a priori du contexte. Cette généralisation mène à l'indexation multimedia. Plusieurs pistes sont possibles pour faire collaborer en particulier les sources d'informations visuelles et textuelles dans un but d'indexation ou de recherche.

Nous citons, par exemple, l'aide à l'annotation textuelle automatique à partir de certaines signatures visuelles ou alors la propagation d'annotation textuelle sur la base d'interactions entre des ontologies textuelles (associations hiérarchiques de concepts) et celles visuelles. L'utilisateur pourra alors naviguer de façon non linéaire dans une grande base d'images à l'aide de cette hiérarchie de concepts. Cette partie de nos thèmes d'activité représente encore, avec la personnalisation, une autre voie pour avancer dans la réduction du « fossé » sémantique.

4. Domaines d'application

Les domaines d'applications des recherches d'IMEDIA sont nombreux. On peut citer :

- **les applications sécuritaires**
  Exemples : identifier un visage ou des empreintes digitales (biométrie). La biométrie est une application spécifique intéressante tant du point de vue théorique que du point de vue applicatif (reconnaissance, IHM, surveillance). Deux thèses ont été effectuées sur ce thème dans le projet. Sont concernées également par nos activités, les bases d'objets volées, les bases d'images obtenues suite à des perquisitions (lutte contre la pédophilie ou autre). Des collaborations avec le ministère de l'intérieur sont en cours.
- **l'audiovisuel**
  Exemple : rechercher un plan spécifique d'un film, d'un documentaire ou d'un journal télévisé, rechercher une antériorité, présenter un résumé,... Une collaboration est en cours avec la chaîne de télévision TF1 dans le cadre RIAM. L'importance des annotations textuelles dans ce type d'application fait que l'indexation et l'accès plurimedia est crucial.
- **les applications scientifiques**
  Exemples : les bases d'images environnementales : biodiversité végétale et animale ; les bases d'images satellitaires : typologie des terrains ; bases d'images médicales : retrouver les images présentant un caractère pathologique, dans un but éducatif ou de diagnostic.
- **l'art et la culture, l'éducation**
  Exemples : recherche encyclopédique, recherche d'un tableau ou d'une illustration par un exemple, un détail. IMEDIA a été contactée par le ministère de la culture et les musées de France au sujet de leurs archives en images.
  rechercher une texture spécifique pour l'industrie textile, illustrer une publicité par une photo adéquate. IMEDIA a entrepris des travaux en partenariat avec une photothèque qui fournit des images, en particulier, aux agences de publicité
- **les télécommunications**
  Exemple : coder, représenter et rechercher les images par leur contenu sont des enjeux importants dans le contexte MPEG-4 et MPEG-7. Nos travaux relèvent des services associés au télécommunications. IMEDIA n'est pas actif dans ces aspects normatifs mais suit les travaux en cours du groupe MPEG7. Les signatures développées par IMEDIA sont compatibles avec cette norme.

5. Logiciels

5.1. Logiciel IKONA/MAESTRO

Un ensemble de signatures (globales et locales) ainsi que l'architecture clien/serveur ont fait l'objet de dépôts à l'APP cette année.

5.1.1. L'interface utilisateur Ikona en C++

Mots clés : *interface utilisateur*, *recherche image par contenu visuel*, *boucle de pertinence*.

Participant : Marin Ferecatu.

L'interface utilisateur ou « le client » est le logiciel qu'on utilise pour envoyer les requêtes, pour afficher les pages des résultats, et pour gérer les modes d'interaction complexes (boucle de pertinence, mots-clés, etc...). A ce titre, il doit être intuitif, rapide et facile à utiliser.

Une nouvelle interface utilisateur (« client ») a été développée en utilisant la bibliothèque graphique Qt de TrollTech qui remplace l'ancien client écrit en Java.

La nouvelle interface est beaucoup plus rapide (code C++ compilé par rapport au code Java interprété) et plus stable. Comme la bibliothèque Qt est multi-platforme, la nouvelle interface est disponible aussi bien sous Unix que sous Windows ou MacOS.

Ce développement a fait l'objet d'un depot APP. Actuellement, la licence envisagée est GPL - en concordance avec la licence Qt.

Nous avons intégré dans l'interface de nouvelles fonctionalités pour améliorer son ergonomie et sa convivialité.

Les futures directions de développement vont se concentrer sur l'intégration de requêtes hybrides texte et image et sur le mode d'interaction boucle de pertinence hybride texte, images et régions d'images.

5.1.2. Développement de Maestro pour Mediaworks et RECIS

Participants : Jean-Paul Chièze, François Fleuret.

Le moteur de recherche d'images par similarités visuelles Maestro permet dans ses versions récentes de modifier de manière concurrente les bases d'images déjà indexées et utilisables dans les requêtes. Ces fonctionnalités ont été ajoutées dans le cadre de collaborations industrielles, pour lesquelles des logiciels clients spécifiques ont été développés.

Le serveur d'indexation Maestro[25] constitue la plate-forme de test des algorithmes de recherche développés au sein du projet IMEDIA. Il a été initialement conçu afin que des utilisateurs distants puissent faire des recherches sur des bases qui auraient été indexées antérieurement. Les projets Mediaworks et RECIS nécessitaient la possibilité pour un logiciel client d'enrichir les bases déjà existantes en rajoutant de nouvelles images.

Une telle fonctionnalité a été réalisée en étendant le protocole de communication de Maestro afin que l'utilisateur puisse transmettre par TCP/IP des images à indexer, qui sont automatiquement ajoutées.

Cette plate-forme étant multi-utilisateur, il a fallu gérer le problème des modifications concurrentes, qui surviennent lorsque l'un des clients modifie une base d'images sur laquelle travaille un autre client. Le protocole permet à présent de s'approprier une base (si elle n'est utilisée par personne), cela de manière atomique, et de la libérer une fois la modification effectuée.

Dans le projet Mediaworks, le client a la forme d'un agent écrit en Smalltalk par la société Ægis, qui communique directement par TCP/IP avec le serveur Maestro. Cet agent fait partie d'une plate-forme globale qui unifie les interfaces utilisateurs d'une part, et les algorithmes de segmentation et d'indexation d'images de l'autre. Dans le cadre de RECIS, nous avons écrit une interface CGI en OCAML à laquelle les clients peuvent accéder de manière classique via un serveur http. Maestro peut répondre à sont tour à la plateforme RECIS de la même manière, par exemple pour indiquer la fin du calcul des indexes.

6. Résultats nouveaux

6.1. Construction et structuration des espaces de description

6.1.1. Signature de formes et optimisation des mesures de similarité

Mots clés : *reconnaissance des formes*, *descripteur de formes*, *comparaison d'images*, *modélisation statistique*, *distance entre images*.

Participants : Jean-Philippe Tarel, Sabri Boughorbel, Marin Ferecatu.

Afin de progresser dans la construction de signatures de formes, nous avons développé une modélisation statistique de l'indexation par une image requête. Ce modèle permet de poser le problème de la recherche de la mesure de similarité comme une optimisation fonctionnelle. La résolution de cette équation est nécessaire pour réaliser des comparaisons objectives entre les différentes signatures de formes pouvant être construites à partir d'une même image.

Dans le cadre de l'indexation d'images, ont été proposés différents types de signatures caractérisant les couleurs [2] et/ou les formes contenues dans une image. Mais les performances de ces différentes signatures sont rarement comparées. Il est pourtant important de savoir quelle signature de formes est la plus performante. La comparaison de deux types de signatures passe par la construction de bases d'images de référence, dites « vérité terrain », où les images similaires sont groupées à la main. De plus, cette comparaison nécessite de

choisir la mesure de similarité la mieux adaptée à chaque signature. Si cette dernière difficulté n'est pas levée, la comparaison de deux signatures n'est pas objective, car un changement de l'une ou de l'autre des mesures de similarité peut modifier son résultat.

Cela nous a conduits à modéliser le lien entre la performance de la recherche d'images et la variabilité des signatures et, en particulier, à chercher la mesure de similarité optimale. Nous avons choisi de considérer une mesure de similarité comme optimale, si elle maximise la précision moyenne sur l'ensemble de la base. Ce modèle permet de dériver l'équation fonctionnelle approchée que doit optimiser $\delta(x-y)$ contribution partielle de la mesure de similarité sur chaque composante :

$$\tau_\delta(f) = \frac{\frac{d^2\bar{\delta}}{dv^2}(v)}{\sqrt{\bar{\delta^2}(v) - \bar{\delta}(v)^2}}$$

où v est une variable aléatoire de densité f qui modélise la variabilité d'une composante de la signature. $\bar{x}$ est l'espérance de x. Si le cas général n'a pas été résolu, nous avons pu dériver la mesure de similarité optimale pour certaines distributions de la variabilité des signatures (normale, exponentielle, uniforme), en se restreignant à des familles de distances. De plus, ceci a permis une première validation par comparaison entre la précision moyenne dérivée du modèle et celle simulée, comme dans la Fig. 1. Ces premiers résultats sont décrits dans [15].

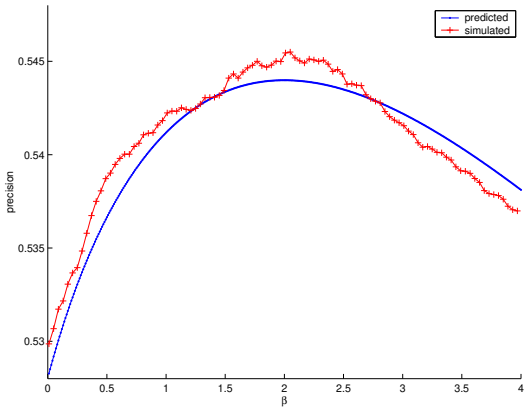


Figure 1. Comparaison entre la précision moyenne prédite et simulée lorsque β varie dans (1), avec $\alpha = 1$, et avec la variabilité f Gaussienne.

Nous avons mené une étude comparative des différentes signatures de formes, en optimisant la mesure de similarité dans une famille à deux paramètres (α, β) contenant les distances de Minkowski :

$$S_{\alpha,\beta}(q,s) = (\sum_i (q_i^\alpha - s_i^\alpha)^\beta)^{\frac{1}{\beta}} \quad (1)$$

Nous avons comparé 17 types de signatures de formes, dont une basée sur la transformée de Radon, et la classique distribution des angles (eoh). Pour l'essentiel, ces signatures sont construites comme les distributions des caractéristiques différentielles de l'image en niveau de gris. L'ordre maximum est deux, et elles sont éventuellement combinées avec une transformation de Hough.

Les résultats sont les suivants :

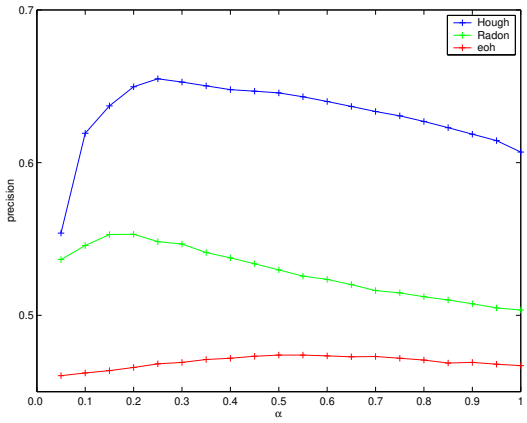


Figure 2. Comparaison entre signatures de formes lorsque α varie. Les signatures sont basées sur la classique distribution des angles (eoh), la transformée de Radon, la transformée de Hough de type « Vector-Gradient ».

- L'introduction de α a permis d'améliorer plusieurs signatures, déjà existantes, basées sur la couleur et la forme.
- La signature capturant le mieux les formes est la transformée de Hough de type « Vector-Gradient » (où l'angle du gradient est utilisé pour affiner la transformée [28]), comme c'est visible dans la Fig. 2. Cette signature n'est pas invariante à une translation ou une rotation.
- Enfin, la signature la plus performante, invariante à une transformation Euclidienne, est celle obtenue par l'application d'une transformée de Fourier sur la transformée de Hough de type « Vector-Gradient ».

En conclusion, une fois identifiée la loi statistique de la variabilité d'une signature, la résolution de l'équation d'optimalité dérivée de notre modélisation de l'indexation permet d'identifier la ou les mesures de similarité qui lui sont le mieux adaptées. Cela permet la comparaison objective des performances des signatures de formes, l'étude de leurs limites intrinsèques, et aide à la compréhension des raisons de leurs performances.

6.1.2. Vers une utilisation optimale des points d'intérêt couleur pour la recherche précise de parties d'images

Mots clés : *recherche précise*, *Points d'intérêt couleur*, *Invariants différentiels*, *Contraintes géométriques*.

Participants : Valérie Gouet, Nozha Boujemaa.

L'année dernière, nous avions proposé une signature locale de l'image à partir de points d'intérêt couleur (HCP), permettant de faire une recherche précise sur des parties ou objets de l'image. L'originalité de l'approche était que la description proposée exploitait l'information couleur pour caractériser de manière compacte et robuste les points d'intérêt extraits automatiquement de l'image. La signature ainsi développée a été intégrée à la plate-forme d'indexation et de recherche IKONA de l'équipe. Le système d'interrogation obtenu permet ainsi à l'utilisateur de sélectionner à la souris la zone de l'image sur laquelle la requête doit porter et le moteur de recherche retourne par ordre décroissant de similarité les images contenant les parties ayant le contenu le plus similaire à celui de la requête.

Cette année, nos travaux ont porté sur l'amélioration de cette classe de descripteurs pour une utilisation optimale des points d'intérêt couleur dans des cas concrets d'indexation d'images où la qualité des images n'est pas connue.

Nous présentons ici l'étude que nous avons réalisée sur l'impact réel des différentes améliorations envisageables pour la caractérisation couleur de points d'intérêt. Plusieurs aspects ont été considérés, tels que l'utilisation de grands ordres dans la caractérisation des points, l'impact du codage de l'image sur ces caractérisations, ou encore l'apport des contraintes géométriques. Cette étude nous a permis de mieux cerner les difficultés de mise en oeuvre d'une telle caractérisation pour la recherche d'images, et donc de définir une description optimale des points d'intérêt pour ce domaine. C'est également une étape préparatoire et nécessaire pour la mise en place d'une structure d'index efficace, aspect que nous avons également commencé à étudier et qui est traité dans la section 6.1.4. Enfin, nous présentons comme illustration une application particulière d'indexation et de recherche d'images exploitant l'approche HCP ainsi que les spécificités qui en découlent.

La caractérisation HCP est basée sur une extraction automatique de points d'intérêt (selon l'opérateur Harris couleur) caractérisés par un ensemble de grandeurs différentielles d'ordre 1 invariantes aux transformations euclidiennes de l'image[30]. Nous avons pu constater dans la littérature que cette classe de descripteurs particulièrement performants pouvait être mise en oeuvre différemment d'une étude à l'autre, impliquant notamment le calcul des invariants à des ordres différents. Or jusqu'à ce jour, aucune des solutions rencontrées n'a été comparée à l'autre, ni évaluée dans des cas concrets d'utilisation. Nous avons ici essayé de répondre à de multiples questions concernant leur utilisation concrète pour l'appariement d'images dans le contexte particulier de l'indexation et la recherche dans des bases d'images. En effet, il s'avère que dans la pratique, le processus d'acquisition des images peut ne pas être toujours contrôlé. Les images peuvent en effet avoir été acquises dans des conditions difficiles ou encore avoir subi des compressions plus ou moins destructives. Plusieurs aspects ont donc été envisagés :

- Quel est le réel impact des invariants d'ordres 2 et 3 sur la recherche d'images ?
- Comment se comportent-ils face au codage de l'image (ici c'est le codage JPEG qui est considéré) et face aux changements d'illumination ?
- Quel est le réel impact de l'utilisation de contraintes géométriques dans la caractérisation des points ?

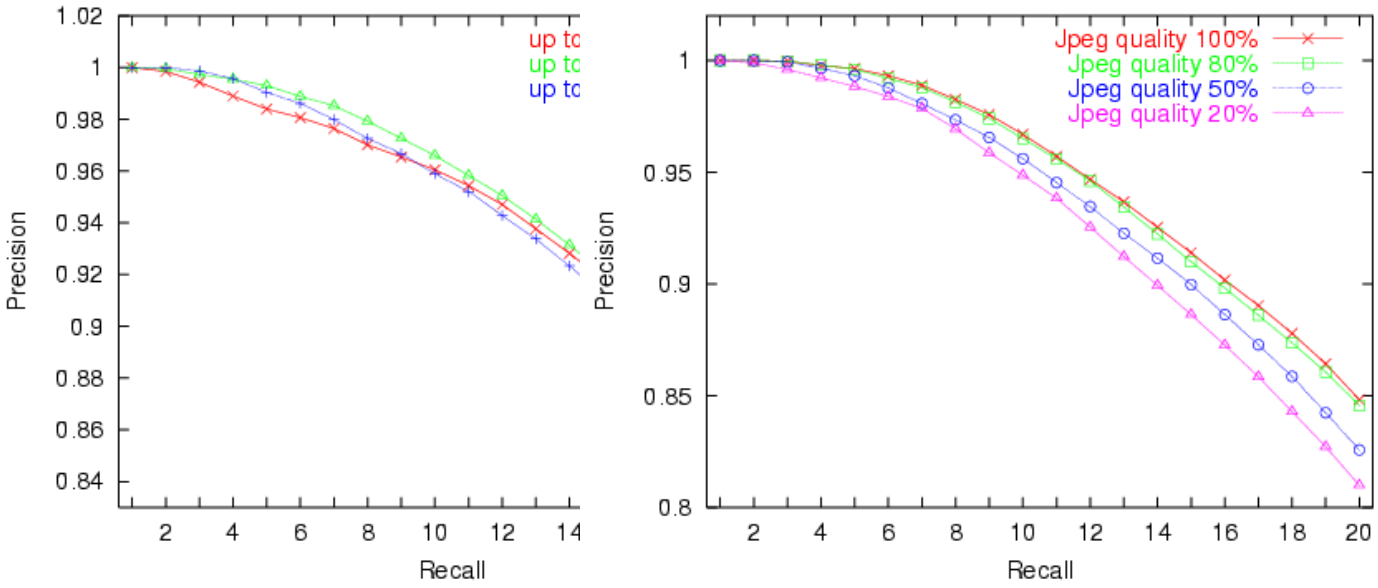
Les expériences sur plusieurs bases d'images vérité-terrain ont conduit aux observations suivantes :

En premier lieu, la grande sensibilité des invariants différentiels du troisième ordre a été confirmée, comme l'illustrent les diagrammes précision/rappel de la figure 3. Ici la base vérité-terrain utilisée est la base d'objets couleur Columbia *coil-100* représentant des objets sous une multitude de points de vue. Les précisions les plus faibles obtenues concernent les points de vue les plus éloignés de ceux des requêtes. On constate donc en particulier avec le premier diagramme la grande sensibilité des invariants d'ordre 3 lorsque le point de vue varie. Le deuxième diagramme montre que cette sensibilité s'accentue encore avec le taux de compression.

Il ne faut cependant pas rejeter systématiquement leur utilisation, car de telles grandeurs aussi précises peuvent s'avérer pertinentes dans le cadre d'applications scientifiques impliquant des images de grande qualité et différant de faibles transformations, comme par exemple les images satellitaires.

Les diagrammes de la figure 3 montrent également que les grandeurs d'ordre 2 conduisent à des résultats à peine meilleurs que la description impliquant seulement l'ordre 1. Cependant, nous avons constaté que l'utilisation de cette classe d'invariants devient réellement pertinente lorsque l'on considère les changements d'illumination. Ainsi une comparaison entre les différentes méthodes de normalisation existantes[31] pour ce type de transformation a été réalisée. Elle a montré que la normalisation des invariants atténue bien mieux les différences d'illumination que les méthodes qui consistent à normaliser l'image. Les invariants d'ordre 1 utilisés seuls ne pouvant pas être normalisés, ce résultat justifie pleinement l'utilisation des invariants couleur jusqu'à l'ordre 2. Cette première partie de l'étude a donné lieu à une publication [9].

La détermination de l'ordre optimal dans le calcul des invariants est cruciale pour le choix de la structure d'index à utiliser, puisque celle-ci dépend fortement de la dimension de l'espace de recherche. Ces résultats nous servent donc de point de départ pour les travaux que nous menons actuellement en ce qui concerne la structuration de l'espace de recherche (ces travaux sont présentés à la section 6.1.4).



(a) Qualité JPEG 100% (b) Ordres 1+2+3

Figure 3. Diagrammes précision/rappel obtenus en fonction des différents ordres d'invariants (a) et de différents taux de compression JPEG *(b).*

La dernière observation porte sur l'impact de l'utilisation de contraintes géométriques de voisinage dans la caractérisation des points d'intérêt. Lors d'une étude publiée dans [20], nous avons pu constater que l'information géométrique ainsi ajoutée n'améliore que très peu ou même dégrade les résultats lorsque la recherche d'images est réalisée sur l'image entière. Ceci s'explique par le fait qu'une recherche globale dans une base généraliste étant par définition *approximative*, l'information photo-métrique employée seule suffit. En revanche, l'information géométrique prend du sens quand le mode d'interrogation est la requête partielle. Dans ce mode, la recherche est faite de manière *précise* sur une zone particulière de l'image pour laquelle l'information photo-métrique seule ne garantit pas la cohésion ni ne décrit précisément la structure géométrique.

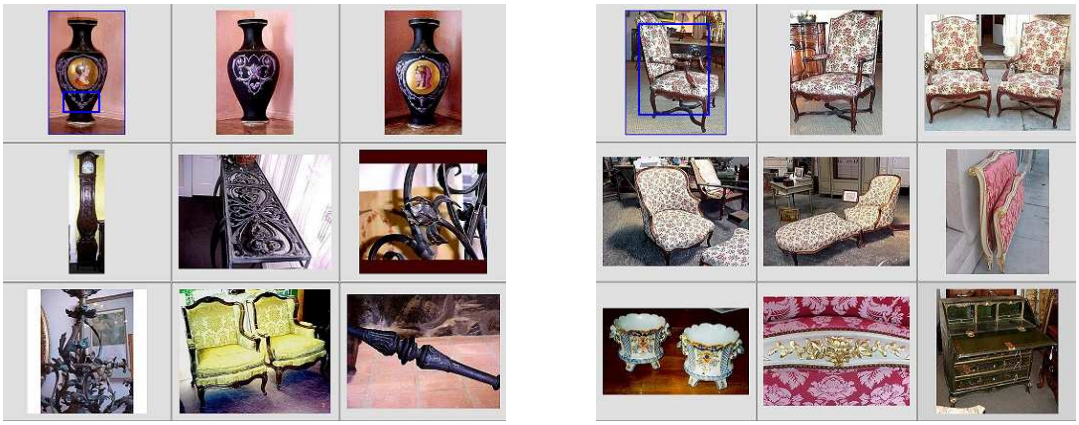
Ces contraintes viennent s'ajouter à la caractérisation avec un poids que l'on choisit généralement beaucoup plus faible que celui associé à l'information photo-métrique (excepté pour des applications très spécifiques comme la détection et l'identification de logos, scénario qui est présenté à la section 6.3.1).

6.1.2.1. Application : la recherche de motifs ou d'objets dans des bases d'objets d'art

La signature HCP a initialement été développée dans le cadre d'un contrat avec la Police Judiciaire pour le programme européen « STOP » concernant la protection des personnes. Les requêtes étant focalisées sur des parties d'images représentant des objets potentiellement déformables, comme des vêtements, nous nous étions limités à l'utilisation de contraintes géométriques assez générales impliquant uniquement la proportion de points correctement appariés dans le voisinage des appariements étudiés. Des contraintes géométriques plus fortes peuvent être considérées lorsque les objets sont rigides, ce qui est le cas dans les scénarios présentés ci-après.

La figure 4 montre des requêtes illustrant deux scénarios différents relatifs à la recherche d'objets d'art. L'image 4-(a) correspond à une requête faite sur un motif particulier, comme le ferait par exemple un collectionneur recherchant des objets de même style. L'image 4-(b) montre quant à elle la recherche d'un objet très précis (une chaise à fleur rose de style Régence), et illustre un scénario de recherche d'objets dans une base répertoriant par exemple des objets volés.

Dans ces deux scénarios, les parties d'images recherchées sont considérées comme étant rigides et les transformations géométriques possibles sont la translation, la rotation, le changement d'échelle et le changement de point de vue. En plus de la proportion de points appariés dans le voisinage des points, il a donc été possible d'exploiter des contraintes géométriques angulaires telles que la conservation de différences de mesures d'angles de gradients entre points voisins. Dans les deux exemples donnés, on retrouve la partie d'image recherchée dans des conditions relativement différentes (point de vue, arrière-plan et éclairage). Ces résultats ont été publiés dans [4].



(a) Recherche d'un motif particulier (b) Recherche d'un objet particulier

Figure 4. Deux scénarios illustrant la recherche précise de motifs et d'objets dans une base d'objets d'art (images fournies par French Accents).

La description d'une image à partir de points d'intérêt suppose que les zones de l'image susceptibles d'être sélectionnées par l'utilisateur lors de la requête sont assez hétérogènes pour détecter un nombre de points d'intérêt suffisamment grand. Au contraire, d'autres approches de descriptions locales existent à partir de primitives différentes, telles que les régions. Dans l'équipe notamment, une approche basée sur une segmentation de l'image en régions caractérisées par une distribution des couleurs a été développée. Cette approche est duale à celle relevant des points d'intérêt, puisque dans ce cas la requête est réalisée sur des zones de l'image - les régions - relativement homogènes. A l'heure actuelle, les approches « point » et « région » opèrent de manière complètement indépendante. Il apparaît pourtant comme évident qu'il soit avantageux de les combiner pour obtenir un système de requêtes partielles plus pertinent et répondant mieux aux demandes de l'utilisateur. Nous travaillons maintenant sur ces aspects.

6.1.3. Indexation 3D

Mots clés : *modèles géométriques*, *formes 3D*, *squelettes*.

Participante : Anne Verroust.

Les structures squelettiques présentées dans [33] ont servi d'axe support dans des algorithmes de déformation axiale ou de métamorphose et une approche similaire locale a été utilisée pour modéliser des déformations avec changements de topologie [16]. La fonction de Morse utilisée dans [33][36] est une approximation de la distance géodésique à un point de la surface appelé point source. Les squelettes sont adaptés pour résoudre certains problèmes de reconnaissance ou d'indexation de formes. Le travail présenté par M. Hilaga, Y. Shinagawa, T. Kohmura et T. L. Kunii en 2001 utilisait un squelette dérivé de celui de [33] pour proposer une méthode de classification de formes. Il nous semble que pour obtenir une méthode d'indexation 3D plus générale, il serait intéressant de coupler la classification de formes via des squelettes à des méthodes issues de l'analyse d'image et c'est pour cette raison qu'un travail sur ce thème est envisagé avec l'arrivée d'Anne Verroust dans le projet.

6.1.4. Structuration de l'espace de description

Mots clés : *index multi-dimensionnels*, *ST-Tree*, *VA-File*.

Participants : Nizar Grira, Arnaud Tournier, Valérie Gouet, Michel Scholl, Nozha Boujemaa.

Afin d'améliorer les temps de recherche lors de requêtes par points d'intérêt, nous avons également travaillé sur la mise en place de structures d'index multidimensionnels permettant une recherche efficace des plus proches voisins des points de la requête. De nombreuses structures issues du domaine des bases de données existent mais ne semblent par directement adaptées à notre problème. En effet, l'espace de recherche dans lequel sont définis les points d'intérêt est, d'après les conclusions de l'étude présentée à la section 6.1.2, un espace de dimension moyenne, mais pouvant contenir un très grand nombre de points ($N \times n$ points si N est le nombre d'images et n le nombre de points par image). Deux structures d'index ont été évaluées pour la recherche dans cet espace particulier.

Deux structures d'index, qui ont fait leur preuve pour des espaces de description standards, ont donc été évaluées pour la recherche dans cet espace particulier : une structure arborescente de type SR-Tree[32] qui partitionne l'espace et une méthode de filtrage des données de type VA-File[37]. Dans notre cas, la recherche est une recherche de type « sphere query » réalisée pour chaque point (calcul d'appartenance à une hypersphère).

Plusieurs bases de points ont été considérées pour l'évaluation : une base de points d'intérêt réels extraits d'une base d'images généraliste et une base de points synthétiques distribués selon une loi uniforme. L'évaluation des performances a été faite en comptabilisant les temps CPU de réponse moyens et le nombre moyen d'entrées-sorties nécessaires à la réalisation d'une requête-point, en fonction du rayon de la sphère (ϵ) et de la dimension (d) de l'espace de recherche. Les résultats obtenus montrent en particulier que pour une description standard des points couleur (soit $d = 8$), les temps de réponse dépendent de la nature des données. En effet, pour la distribution réelle (groupée), le SR-Tree est plus performant (cf. figure 5) alors que pour des données distribuées uniformément, le VA-File obtient des temps de réponse inférieurs jusqu'à une certaine valeur de ϵ. En revanche, lorsque l'on augmente la dimension de l'espace ($d = 15$ et $d = 40$), c'est la technique du VA-File qui obtient les meilleurs résultats. En effet, le nombre d'approximations réalisées par cette structure augmente avec la dimension de l'espace, ce qui réduit le nombre de comparaison avec les données réelles sur disque.

D'après les conclusions de la section 6.1.2 sur la dimension optimale à considérer pour l'espace de recherche, il ressort que c'est la structure du SR-Tree qui semble la plus appropriée à la recherche par points d'intérêt. Nous avons ensuite essayé d'adapter cette technique à cette classe de requêtes particulière : au lieu de réaliser autant de recherches de plus proches voisins qu'il y a de points dans la requête-image (et donc autant de parcours de l'arbre), nous avons amélioré l'algorithme de parcours de sorte que chaque noeud de l'arbre soit parcouru au plus une fois pour l'ensemble des points de la requête. Cette optimisation suppose que plusieurs points de la requête sont représentés par le même noeud de l'arbre et donc assez proches dans l'espace de description. Les performances obtenues sont également présentées à la figure 5 (courbe « Multiple-Range-Query »). L'utilisation actuelle du SR-Tree amélioré permet maintenant d'exploiter la description HCP sur des bases de plusieurs dizaines de milliers d'images avec des temps de réponse acceptables, ce qui n'était pas envisageable avec un parcours séquentiel des données.

Ces travaux ont été réalisés dans le cadre d'un stage de DEA (Arnaud Tournier [23]) et d'un stage de fin d'étude (Nizar Grira [21]).

L'étude sur la structuration de l'espace de recherche nous a permis de constater l'importance du paramètre ϵ (représentant le rayon de la sphère de recherche des plus proches voisins), puisqu'il influence grandement les temps ainsi que la qualité des réponses retournées par le système de recherche. Dans le domaine des Bases de Données, ce paramètre est en général fixé à l'avance ou bien déterminé par l'utilisateur (exemple : *« je recherche les personnes de moins de 40 ans »*). En ce qui nous concerne, cela ne peut évidemment pas être le cas. Une première solution consiste à l'estimer à partir d'une ou plusieurs bases représentatives, avec le risque qu'il ne soit pas finement adapté à des bases d'images quelconques. La deuxième approche, que nous avons

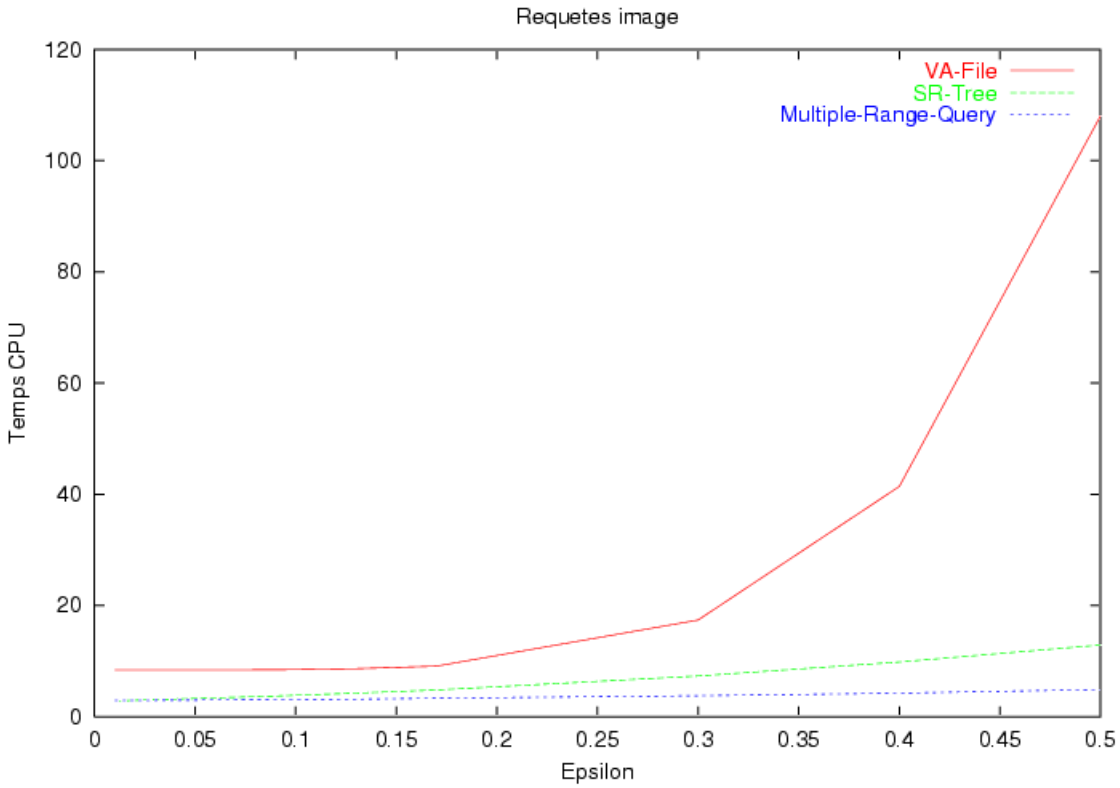


Figure 5. Comparaison des temps de réponse en fonction de ϵ pour les structures VA-File, SR-Tree et SR-Tree optimisé (courbe « Multiple-Range-Query »).

choisie d'explorer, consiste à estimer automatiquement ϵ pour la base d'images considérée. Pour ce faire, la piste que nous comptons approfondir est celle du regroupement par classification.

6.2. Classification automatique et apprentissage statistique

Notre travail de recherche en apprentissage statistique s'organise selon deux thèmes principaux. Le premier concerne l'utilisation de combinaisons hiérarchiques de classifieurs simples pour la détection d'objets, l'étude de l'optimalité de ces combinaisons, et l'étude de ces classifieurs même. En particulier, nous travaillons également sur les propriétés d'invariance de machines à vecteurs de support basées sur certains noyaux peu utilisés habituellement, par exemple le noyau triangulaire. Le second thème est la classification non-supervisée. On travaille en particulier sur des méthodes qui permettent le calcul automatique du nombre de classes par agglomération compétitive. Nous essayons de prendre en compte la variabilté de densités et de formes des classes pour une meilleure estimation de leurs paramètres.

6.2.1. Localisation précise de visages

Mots clés : *détection de visages*, *apprentissage statistique*, *estimation de la pose*.

Participants : François Fleuret, Donald Geman, Moez Tarzi, Jérémie Jakubowicz.

De nombreuses applications pratiques de la détection de visages demandent une estimation fine des positions de visages détectés. La plupart des algorithmes existants ne donnent que des estimations grossières de la position et de la taille de chaque visage. Nous proposons ici une approche robuste qui estime finement la position des yeux.

Les classifieurs que nous utilisons reposent sur un comptage de fragments de bords dans des voisinages de l'image. Leurs invariances individuelles entraînent plusieurs détections pour chaque visage présent dans l'image. À chacune d'entre elles est automatiquement associée une estimation grossière de la pose (x, y, θ, s), où (x, y) est la position du centre des yeux, θ l'inclinaison et s la distance entre les yeux.

Nous avons amélioré l'estimation de ces poses en introduisant un processus d'aggrégation hiérarchique reposant sur la métrique :

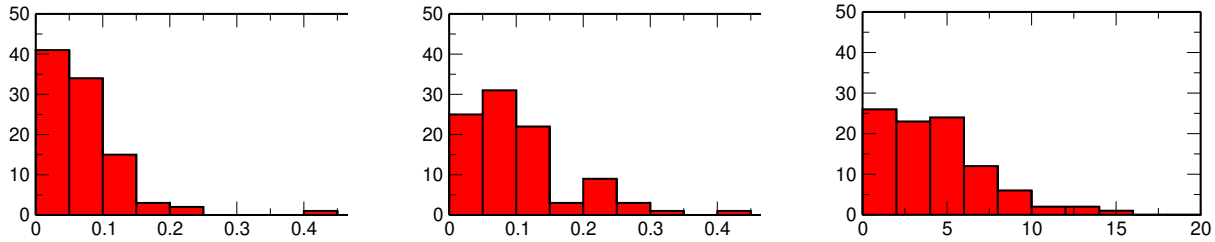


Figure 6. Distribution des erreurs relatives de l'estimation de la position du centre des yeux (gauche) et de la distance entre les yeux (centre) et de l'erreur absolue en degrés de l'estimation de l'inclinaison (droite).

Cette estimation de la pose a été utilisée dans deux applications concrètes. D'une part dans le projet Mediaworks où la taille de chaque visage relativement à la taille de l'image est utilisée pour estimer la valeur du cadrage (cf. section 6.3.2). Ensuite dans le cadre du stage de DEA de Moez Tarzi qui a travaillé sur l'optimisation de la transmission d'images de vidéo conférences. La détection a permis de séparer le flux vidéo en deux flux, l'un contenant les images des visages des participants, et l'autre les images du reste de la scène. Une telle stratégie permet d'assurer le débit, et donc la qualité, alloué à l'encodage des visages, qui constituent la zone la plus importante pour un utilisateur.

Enfin, Jérémie Jakubowicz a validé au cours de son stage de fin d'étude la généralité de l'approche hiérarchique en mesurant empiriquement les performances de perceptrons dédiés à des problèmes de classification d'images synthétiques de plus en plus contraintes. Comme prévu, à complexité du classifieur constante (dans ce cas le nombre de neurones cachés), le taux d'erreur diminue.

6.2.2. Machines à vecteurs de support pour une détection hiérarchique des visages

Mots clés : *machines à vecteurs de support*, *détection hiérarchique des visages*, *calcul efficace*, *machines à base de noyaux et optimisation*.

Participants : Hichem Sahbi, Donald Geman.

Glossaire

Simplification des SVM : Réduire le coût d'évaluation d'une machine à vecteurs de support.

Dans le cadre de nos travaux sur la détection et la reconnaissance des visages [12] [13], nous proposons un algorithme efficace de localisation des visages basé sur une hiérarchie de machines à vecteurs de support (SVM) qui sert de plate-forme de recherche rapide des occurrences des visages dans une scène. La hiérarchie de SVM est organisée de telle sorte qu'en allant de la racine vers les feuilles, les SVM sont de plus en plus coûteux et leur taux de fausses alarmes est relativement réduit. Une analyse multi-échelle est effectuée afin d'extraire les sous-images d'une scène, et les classifier à l'aide d'une stratégie de recherche en profondeur dans la hiérarchie avec un coût moyen extrêmement faible. Ce coût est défini par un modèle qui lie le taux d'erreur des SVM et leur coût d'évaluation. Dans ce cas, seules les sous-images contenant des structures rares et semblables aux visages (bords verticaux ou horizontaux, régions fortement texturées, etc) nécessitent un traitement intense.

La pose d'un visage est définie par trois paramètres (position, orientation et échelle) pris dans un domaine dit générique. Ce domaine est subdivisé récursivement en sous-domaines emboîtés de poses de plus en plus restreintes où une population de visages, associée à chaque sous-domaine, est utilisée afin d'apprendre une machine à vecteurs de support. Par la suite cette hiérarchie de SVM sera référencée sous le nom de *réseau-f*. Une sous-image est classifiée « visage » par le réseau-f, s'il existe une chaîne complète de la racine vers une feuille où tous les classifieurs SVM répondent positivement à l'hypothèse « visage ».
En allant de la racine vers les feuilles du réseau-f, les domaines de poses sont de plus en plus restreints, ce qui rend le nombre de vecteurs de support (le coût d'évaluation) et le taux d'erreur décroissant des SVM sous-jacentes. Le coût prohibitif des SVM dans les niveaux supérieurs du réseau-f rend le coût global de classification énorme. Le *réseau-g* est une plate-forme similaire au réseau-f dans laquelle le coût d'évaluation des SVM est contraint à être croissant en fonction des niveaux de la hiérarchie. Ceci permet un rejet rapide des sous-images contenant des structures simples tout en garantissant *l'hypothèse de conservation*, i.e. préserver les structures visages à tous les niveaux de la hiérarchie [14].

On associe pour chaque SVM f_c dans le réseau-f, une machine g_c dans le réseau-g ayant un coût fixé a priori par un modèle (voir paragraphe suivant). Afin de construire g_c, nous proposons une variante de la technique de « l'ensemble réduit » (reduced set technique) [26] qui permet de réduire le coût d'évaluation d'un SVM f_c en générant une machine g_c ayant un ensemble réduit de vecteurs de support et qui approxime le plus possible les performances de généralisation du classifieur f_c. Cette contrainte est définie par une fonction objective dite de « corrélation » qui est généralement non-convexe et difficile à minimiser. Ainsi, afin de résoudre ce problème d'optimisation, nous proposons une méthode basée sur le clustering permettant

d'initialiser efficacement les coordonnées de l'ensemble réduit des vecteurs de support. Une étude comparative des différentes techniques d'initialisation de ces coordonnées a conclu que le clustering fournit largement la meilleure approximation de l'ensemble réduit et permet donc une meilleure convergence de la méthode du gradient conjugué utilisée comme outil de minimisation. Notons que chaque machine g_c du réseau-g est conçue afin de garantir l'hypothèse de conservation, et ceci en adaptant son seuil de rejet et rendre ainsi le taux de faux négatifs nul empiriquement. Ceci permet au classifieur g_c de rejeter beaucoup de structures « non-visage » avec un coût extrêmement réduit toute en préservant les structures « visage ».

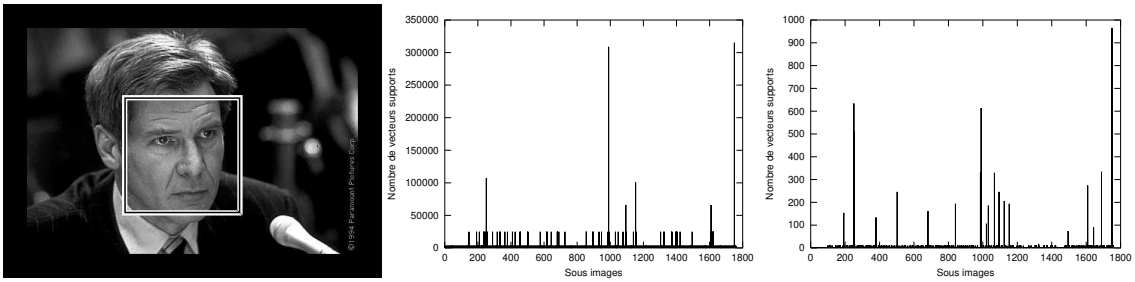


Figure 7. Cette figure montre le coût de classification des sous-images traitées à l'aide du réseau-f (au milieu) et du réseau-g (à droite).

Enfin, le coût de chaque SVM dans le réseau-g est déterminé par la résolution d'un problème de minimisation par contraintes qui caractérise le coût moyen de l'évaluation du réseau-g et son taux d'erreur. Cette analyse est effectuée sous l'hypothèse d'existence d'une fonction convexe liant le coût d'un SVM et son taux d'erreur qui est largement dominé par le taux de fausses alarmes car la majorité des sous-images traitées contient des structures « non-visages ». L'utilisation du réseau-g permet ainsi un traitement efficace des scènes où seules les sous-images contenant des structures visages nécessitent le recours aux SVM appartenant aux feuilles du réseau-g. Ces SVM sont plus coûteuses que celles proches de la racine, néanmoins elles sont très rarement utilisées, ce qui rend le coût moyen de classification extrêmement faible.

6.2.3. Invariance au changement d'échelle des SVM utilisant le noyau triangulaire

Mots clés : *machines à vecteurs de support*, *invariance*, *apprentissage statistique*.

Participants : François Fleuret, Hichem Sahbi.

Nous étudions les performances expérimentales et les propriétés théoriques du noyau triangulaire utilisé dans le cadre des machines à vecteurs de support (*support vector machines*). Nous avons montré des performances supérieures à celles obtenues avec le noyau classique (gaussien) sur des tâches de reconnaissance de caractères et de détection de visages. Ces performances sont liées à l'invariance de l'apprentissage aux variations d'échelles que ce noyau assure.

Les machines à vecteurs de support (*support vector machines*) [27] sont parmi les méthodes d'apprentissage statistique les plus utilisées aujourd'hui, aussi bien pour leurs performances que pour leur facilité d'implémentation.

Leur utilisation pour un problème concret impose de choisir un *noyau*. Cette fonction mathématique projette implicitement les données dans un espace où elles auront des propriétés de séparabilité linéaire. Le choix du noyau, et leur étude, ont toujours constitué l'un des problèmes majeurs de ce domaine.

Nos travaux sur la détection de visage[35] ont démontré les très bonnes performances du noyau triangulaire comparativement aux noyaux classiques. Intuitivement, ce noyau possède une invariance de forme lorsqu'on change d'échelle (figure 8).

Si nous notons f^γ une SVM basée sur le noyau triangulaire et construite par apprentissage sur une population qui a été dilatée d'un facteur γ, nous prouvons analytiquement que pour tout γ et tout x nous avons $f^\gamma(\gamma x) = f^1(x)$. Ainsi, le noyau triangulaire nous assure bien que le processus d'apprentissage est invariante aux changements d'échelle. Nous avons rédigé un rapport technique [22].

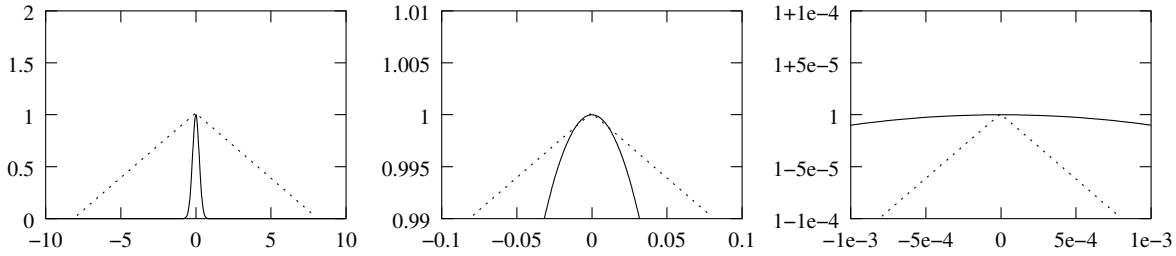


Figure 8. Noyau gaussieu (ligne continue) et noyau triangulaire (ligne pointillée) à différentes échelles (de gauche à droite respectivement $\times 10^0$, $\times 10^2$ et $\times 10^4$). Intuitivement, alors que le noyau triangulaire reste identique à toutes les échelles, le noyau gaussien se comporte comme une dirac ou une pondération uniforme en fonction de l'échelle à laquelle est il est vu.

6.2.4. Classification non-supervisée par agglomération adaptative

Mots clés : *classification robuste*, *classification adaptive*, *combinaison des signatures*, *résumé d'une base d'images*, *images informatives*.

Participants : Bertrand Le Saux, Nozha Boujemaa, Nizar Grira.

Nous étudions des algorithmes de classification non-supervisée hautement adaptifs à la structure des données, et notamment capables de trouver le nombre de classes et de détecter des classes de tailles et de densités variées. Nous proposons un algorithme efficace qui est utilisé pour catégoriser des bases d'images et produire des résumés visuels de séquences vidéo.

L'algorithme de catégorisation a été perfectionné depuis l'an dernier pour arriver dans un état stable, baptisé ARC (Adaptive Robust Clustering) : il est capable de s'adapter à la taille et à la population des classes, et les signatures sont normalisées au cours de la catégorisation.

1. Dans le cas particulier des bases d'images, les catégories naturelles présentent des compacités variées (certaines sont concentrées, d'autres sont étendues et lâches). Dans l'agglomeration compétitive classique, la force de la compétition est réglée par le facteur $\alpha(k)$. On fait en sorte que l'agglomération soit adaptive à chaque catégorie.
   Une distance moyenne pour chaque catégorie est introduite :

$$d^2_{moy}(s) = \frac{\sum_{i=1}^{N}(u_{si})^2 d^2(x_i, \beta_s)}{\sum_{i=1}^{N}(u_{si})^2} \quad for\ 1 \leq s \leq C \tag{2}$$

   et une distance moyenne pour la base entière est calculée :

$$d^2_{moy} = \frac{\sum_{j=1}^{C}\sum_{i=1}^{N}(u_{ji})^2 d^2(x_i, \beta_j)}{\sum_{j=1}^{C}\sum_{i=1}^{N}(u_{ji})^2} \tag{3}$$

   On peut alors pondérer $\alpha(k)$:

$$\alpha_s(k) = \frac{d^2_{moy}}{d^2_{moy}(s)}\alpha(k) \quad for\ 1 \leq s \leq C \tag{4}$$

   le rapport $\frac{d^2_{moy}}{d^2_{moy}(s)}$ est inférieur a 1 pour les catégories étendues, donc l'effet de la compétition sera moins fort pour elles.

2. La normalisation des différentes signatures est effectuée au cours de la catégorisation. Les distances intra-signature sont triées, et un poids plus fort est donné aux distances partielles les plus faibles. De cette manière, si une des signatures est particulièrement pertinente pour la catégorisation, l'algorithme lui accordera un poids plus important.
3. Un des derniers problèmes qui subsiste est la difficulté que rencontre l'algorithme dans le cas où une classe a une population élevée : alors quelques catégories peuvent co-exister au sein de la même classe. Dans le cadre du stage de Nizar Grira, nous avons travaillé à une méthode permettant de fusionner des catégories selon un critère de proximité géométrique.

Notre méthode a été comparée avec une des méthodes de catégorisation les plus récentes et performantes appliquées à la catégorisation de bases d'images : l'algorithme SOON [34]. Avec SOON, certaines catégories naturelles sont séparées en plusieurs clusters, ce qui donne des résumés redondants. De plus SOON considère plus d'un quart de la base comme du bruit, et les classes obtenues, même si elles sont parfaitement homogènes, ne fournissent plus une bonne représentation de la base. Au contraire, ARC ne considère comme bruit que les images vraiment ambigues qui gêneraient le processus de catégorisation. Les classes les moins bien définies comportent une faible part de bruit, mais ce problème sera résolu en permettant à l'utilisateur de préciser les classes qu'il recherche.

La méthode ARC a été présentée dans deux conférences [18] et [17].

Elle a été appliquée dans le cadre d'une collaboration avec TF1 pour *la génération automatique de résumé* de séquences de journal télévisé et la détection des scènes de plateau (figures 9 et 10).

Figure 9. Catégories obtenues avec la méthode ARC, sur des images extraites d'une séquence vidéo de journal télévisé. Toutes les images de plateau sont collectées dans le cluster en bas à gauche.

6.3. Indexation trans-media : Aide à l'annotation textuelle d'images

Nous avons démarrer un premier scenario d'indexation hybride texte/image. Il s'agit de proposer une annotation textuelle automatique à partir de l'analyse du contenu visuel d'une image. Nous présentons dans ce qui suit deux cas élaborés dans le cadre du projet RIAM Médiaworks.

6.3.1. Détection, identification et suggestion automatiques de logos

Mots clés : *détection et étiquetage d'objets particuliers*.

Participants : Valérie Gouet, Nozha Boujemaa.

Le scénario consiste à rechercher dans une image et à identifier d'éventuels logos, à partir d'un thésaurus de logos caractéristiques (noté BLC). Ce type de recherche correspond à un scénario particulier de requêtes partielles par points d'intérêt. Ici nous supposons que les transformations géométriques possibles sont réduites à la translation et au changement d'échelle, et que les logos sont des objets rigides, plans et potentiellement

Figure 10. Une des catégories obtenues. Comme la catégorisation est basée sur la similarité visuelle, le résultat est indépendant de la valeur de plan : les plans d'ensemble sont retrouvés avec les gros plans

transparents. Il est donc possible d'envisager des invariants photométriques plus contraints et précis que les invariants différentiels de Hilbert, comme le jet local calculé à plusieurs échelles (pour différents supports de gaussienne). Des contraintes géométriques fortes peuvent également être envisagées pour rendre la reconnaissance plus robuste. Ici nous avons estimé les paramètres de translation et de changement d'échelle à partir des meilleurs points appariés (à partir de l'information photométrique) pour chaque image de la base BCL impliquée. Contrairement à l'implémentation générale qui a été présentée, il a été donné un poids plus important à l'information géométrique que celui affecté à l'information photométrique, afin de rendre la caractérisation robuste jusqu'à un certain degré de transparence des logos.

Ce scénario a été développé dans le cadre du projet MediaWorks à partir de logos et d'images extraits de vidéos fournies par TF1, l'objectif étant de reconnaître automatiquement la provenance des images et de caractériser leurs droits. La solution proposée prend en entrée une image et indique en sortie si elle contient un ou plusieurs logos, auquel cas elle les situe dans l'image et les identifie en donnant le label associé aux logos les plus similaires dans la base BLC, avec un taux de confiance fonction de la mesure de similarité. Deux exemples types sont présentés à la figure 11.

Image test	Logos détectés et localisés	Labels correspondants (taux de confiance)	logos BLC correspondants
		Saab (85%) L'équipe (91%) TF1 (84%)	
		Saab (76%) Saab (68%) Saab (54%) Saab (85%)	

Figure 11. Détection, identification automatiques de logos.

Cette approche d'aide à l'indexation textuelle par analyse du contenu visuel de l'image peut être généralisée à des catégories d'images différentes. Reprenons par exemple le scénario présenté à la section 6.1.2. Pour

chaque image de la base d'objets d'art considérée est disponible un ensemble de mots-clés, décrivant par exemple le style, l'histoire, la provenance ou encore le motif du tissu de l'objet. On peut ainsi imaginer une personne détenant un objet particulier et l'ayant photographié, mais n'ayant aucune information sur celui-ci. Il suffirait alors qu'elle interroge la base à partir de cette photographie pour obtenir en réponse les objets les plus similaires et donc la ou les descriptions qui leurs sont associées.

6.3.2. Estimation automatique de la valeur de plans

Mots clés : *détection de visage*, *estimation de la pose*.

Participant : François Fleuret.

La spécification du cadrage utilisés dans les documents vidéo constitue une information importante, qui est aujourd'hui ajoutée manuellement par des documentalistes. L'utilisation de la détection de visages permet de déterminer cette information automatiquement.

Le type de cadrage utilisé dans un reportage ou une interview (gros plan, plan moyen ou plan éloigné) est très utile aux documentaliste lors de leurs recherches dans de grosses banques de sujets. Cette information est actuellement rajoutée manuellement dans des fiches textuelles par des opérateurs qui visualisent le document vidéo en entier.

À partir des spécifications données par les documentalistes, nous avons défini un ensemble de critères qui permettent, étant donnée une liste de visages détectés automatiquement, de préciser automatiquement lesquels sont des gros plan ou des plans moyens.

A partir d'une image seule, le système pourra donc accompagner la liste des détections de labels 'GP' pour gros plan et 'PM' pour plan moyen. La figure 12 montre un exemple d'estimation de valeurs de plans. Le résultat est superposé sur les images et sera fournie comme suggestion automatique d'annotations textuelles au documentaliste. Il s'agit d'un premier exemple d'indexation hybride image/texte possible.

Figure 12. Exemple de détection visualisée à l'aide du client IKONA. Les gros plans sont indiqués automatiquement.

6.4. Recherche interactive et personnalisation

6.4.1. La recherche subjective par bouclage de pertinence

Mots clés : *bouclage de pertinence*, *modelisation statistique*, *profil utilisateur*.

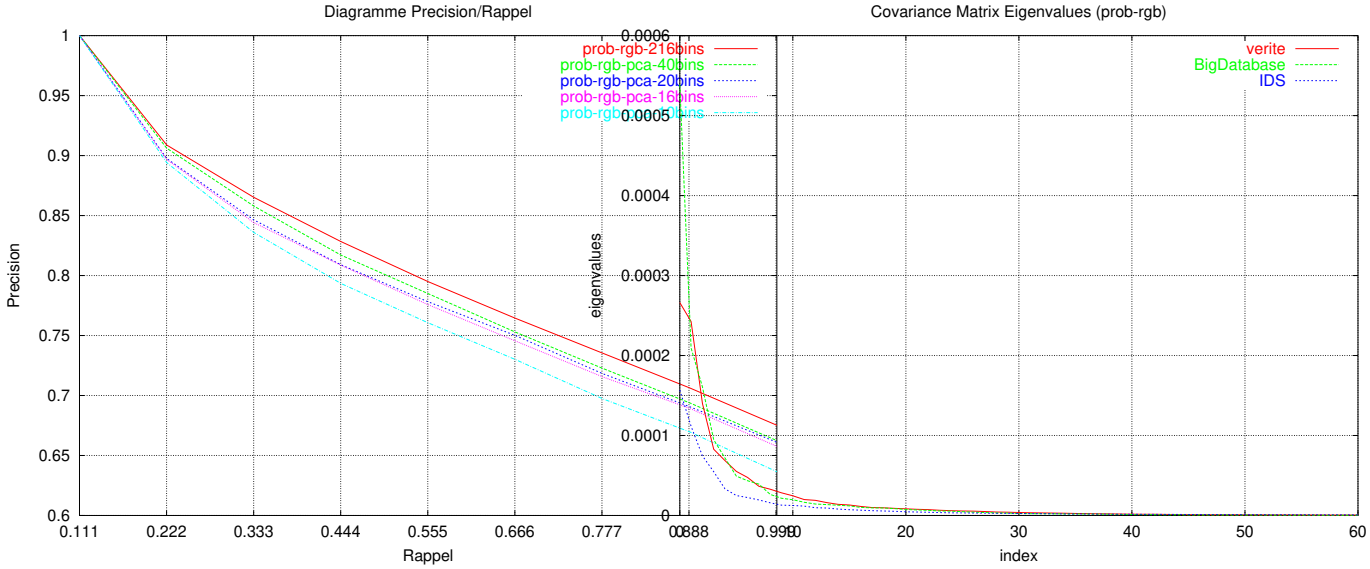
Participants : Ferecatu Marin, Boujemaa Nozha.

Les méthodes de type boucle de pertinence permettent aux utilisateurs de spécifier des exemples positifs (le système cherche des images visuellement similaires aux exemples positifs) et des exemples négatifs (le système évite, autant que possible, de retourner des images proches de ceux-ci). Les autres images sont considérées comme indifférentes (non spécifiées).

Nous avons mis en oeuvre une méthode performante de recherche d'images par boucle de pertinence, qui est inspirée par des méthodes d'estimation non-paramétriques des fonctions de densité de probabilité.

Dans une première étape, nous utilisons l'analyse en composantes principales (ACP) avec toutes les signatures pour réduire la dimension plus de dix fois, pour une perte globale de performance de moins de 2% dans les courbes précision-rappel.

Après l'ACP nous obtenons un vecteur de caractéristiques de l'image qui contient presque la même quantité d'information mais dans un espace de dimension très réduite.



(a) (b)

Figure 13. (a) diagrammes précision/rappel pour la signature probabilité-RVB après l'ACP (b) le graphe des valeurs propres de la matrice de covariance pour des quelques bases d'images.

Dans la figure 13-a on peut voir les diagrammes précision-rappel pour la signature probabilité-RVB et des divers choix du nombre de dimensions gardées après l'ACP. Pour un rappel égal à 1 et une ACP de 20 bins on peut voir une baisse d'approximativement 2% dans la précision, mais la dimensionalité est plus de dix fois moins grande.

La figure 13-b présente les valeurs propres de la matrice de covariance de l'index probabilité-RVB pour diverses bases d'images. Un bon critère pour choisir le nombre k de dimensions à garder après l'ACP est, par exemple, que l'énergie des dimensions éliminées représente moins de 1% de l'énergie totale sur toutes les dimensions :

$$\frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{N} \lambda_i} \leq 0.01$$

où N est le nombre de dimensions avant l'ACP et λ_i sont les valeur propres.

L'ACP est vue comme une pré-condition pour la mise en place des méthodes RF rapides qui doivent fournir des réponses en temps réel.

Notre méthode RF est optimisée en espace mémoire et temps d'exécution ce qui la rend bien adaptée pour la recherche dans des grandes bases d'images.

Chaque image positive/négative (vecteur) est considérée comme la source d'une fonction (potentiel) de pertinence/non-pertinence autour d'elle même (F_{RX}, F_{NX}). Pour toute image dans la base (I) nous calculons le critère de pertinence :

$$J(I) = \sum_{X \in \mathbf{R}} F_{RX}(I) - \sum_{X \in \mathbf{N}} F_{NX}(I)$$

ou $\mathbf{R}$ et $\mathbf{N}$ sont les ensembles d'images pertinentes et non-pertinentes.

Pour les fonctions F, il y en a quelques conditions pour un bon choix. Elles doivent :

- être rapides à calculer ;
- contenir quelques paramètres pour pouvoir permettre une bonne adaptation au nombre d'images déjà choisies ;
- être dominantes localement (telle que pour n'importe quelle configuration des exemples négatifs, un exemple positif n'est jamais annulé) ;

Nous utilisons avec de bons résultats des fonctions de type hyperbolique $1/kx$, où le paramètre k est choisi tel que la fonction a une petite couverture (1-5%) du domaine des images. Également, le paramètre k change proportionnellement avec le nombre d'exemples, et il est plus petit pour les exemples négatifs que pour les exemples positifs.

Même si elles ne sont pas des fonctions densité de probabilité, ces fonctions sont beaucoup plus rapides à calculer et plus stables numériquement que les plus classiques fonctions gaussiennes.

La Figure 14 présente la première page des résultats envoyé par IKONA, après trois itérations RF avec la méthode décrite.

D'habitude, on a besoin de plus d'exemples négatifs que d'exemples positifs. Si les exemples positifs fournissent des éventuels centres pour les « concepts » à chercher, les exemples négatifs sont utilisés pour définir les limites des ces centres, ce qui permet une meilleure délimitation.

Une partie de ces résultats a déjà été publiée dans [4].

Dans la suite de ce travail nous allons nous concentrer sur la généralisation de la méthode pour prendre en compte les régions des images et aussi sur l'utilisation d'éventuels mots clés/texte dans la boucle de pertinence.

6.4.2. Retour de pertinence pour l'affinage de la définition des catégories

Mots clés : *affinage de catégorie d'image*, *contrôle de pertinence*, *machines à vecteurs de support*, *reclassification subjective*.

Participants : Bertrand Le Saux, Nozha Boujemaa.

Les catégories déterminées par le système sont proposées à l'utilisateur, qui dispose d'outils pour les affiner en sélectionnant simplement quelques images exemples.

Le résumé obtenu par la catégorisation est présenté à l'utilisateur pour lui permettre de naviguer dans la base d'images. Il a ainsi une vue d'ensemble de la base, et peut accéder aux catégories qui l'intéressent en cliquant sur le représentant. Lors de la catégorisation, un compromis est fait entre la recherche la plus exhaustive des classes significatives, et la présence d'images mal-classifiées dans les catégories. Pour affiner les classes retrouvées, et s'adapter plus précisément à ce que recherche l'utilisateur, un contrôle de pertinence est introduit.

Pour cela, un classifieur de type Machine à Vecteurs de Support est utilisé. L'utilisateur sélectionne des exemples positifs et négatifs d'images d'après leur aspect visuel. Les signatures correspondantes servent à entraîner le classifieur, qui calcule la frontière dans l'espace des signatures d'images entre images positives et images négatives. Dans un deuxième temps, les images de la catégorie sont classifiées par rapport à cette frontière, ce qui permet de rejeter certaines images et de ne conserver que les images recherchées (figure 15).

À partir des résultats de la classification, l'utilisateur peut ajouter les images rejetées à la catégorie bruit, ou bien créer un nouveau groupe indépendant. Il a également la possibilité de fusionner des groupes existants, et

Figure 14. Recherche des portraits par boucle de pertinence

de sauvegarder ses changements pour des sessions ultérieures. Ainsi l'utilisateur peut adapter le résultat de la catégorisation à ses propres critères.

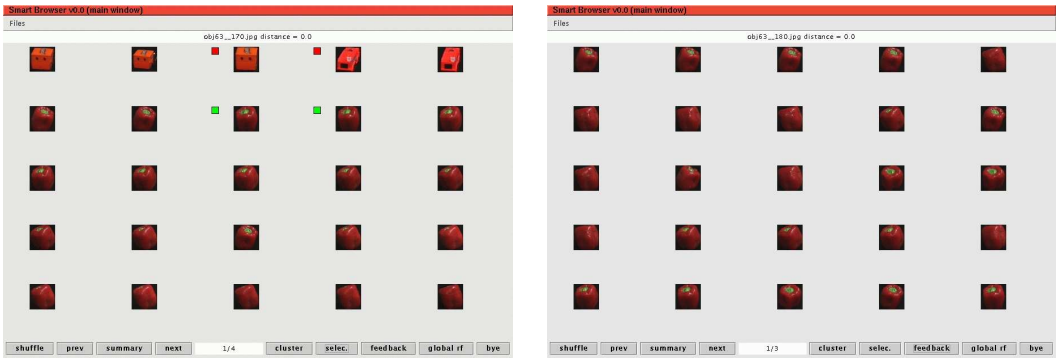


Figure 15. droite : Sélection des exemples positifs et négatifs par l'utilisateur pour un affinage des classes. gauche : Résultat de l'affinage de la classification par SVM.

6.4.3. Recherche d'images par composition logique de catégories de régions

Mots clés : *composition d'image*, *sémantique visuelle*, *catégories de régions*, *segmentation d'image*, *requête régions multiples*, *« range query »*, *interface utilisateur*, .

Participants : Julien Fauqueur, Nozha Boujemaa.

Nous avons développé un cadre de recherche d'images original et nouveau permettant des requêtes complexes par régions multiples basées sur la composition de catégories de régions. Cette approche est basée sur la formation automatique des catégories de régions par regroupement de leurs caractéristiques visuelles. Elles permettent à l'utilisateur de formuler sa requête en spécifiant les différents types de régions devant apparaître dans les images retrouvées mais aussi les types indésirables. Le système proposé peut alors retrouver les images à partir de requêtes complexes sur la composition des catégories de régions telles que : « trouver les images contenant des régions de ce type et de ce type, mais pas de région de ce type ». Le système supporte les « range query » en considérant les régions de la même catégorie mais aussi celles des catégories voisines.

L'approche d'indexation et de recherche s'avère simple et très rapide, même pour des requêtes très complexes sur de grandes bases d'images. Les tests ont été effectués sur une base de 9995 images de la base d'images Corel.

6.4.3.1. Définition des catégories de régions et de leur voisinage :

Nous souhaitons former au moment de l'indexation des groupes de régions similaires afin de pré-calculer efficacement la similarité entre les régions de la base. Les régions des images de la base sont extraites par segmentation par classification des distributions locales de couleurs quantifiées [7]. Nous procédons par classification des descripteurs visuels des régions en utilisant l'algorithme d'Agglomération Compétitive. Chaque groupe est appelé « catégorie de régions » et deux catégories proches (i.e. dont les prototypes sont proches) sont appelées « catégories voisines ». Au moment de la recherche d'images, les régions similaires seront définies comme celles d'une même catégorie mais aussi d'une catégorie voisine selon le rayon de « range query » choisi. 91 catégories sont ainsi obtenues à partir des 50.220 régions de la base. Pour chaque catégorie, une région représentative est définie afin d'illustrer chaque catégorie dans l'interface graphique.

6.4.3.2. Recherche d'images par composition de catégories de régions :

L'utilisateur formule sa requête en sélectionnant parmi les 91 catégories celles auxquelles les régions des images doivent appartenir (appelées « catégories requête positives ») et celles auxquelles elles ne doivent pas appartenir (« catégories requête négatives »). Dans le cas d'une seule catégorie requête positive, on cherche les images dont les index comportent le label de la catégorie requête *ou* le label d'une catégorie voisine (« range query »). L'index de chaque image est constitué de la liste des labels des catégories auxquelles ses

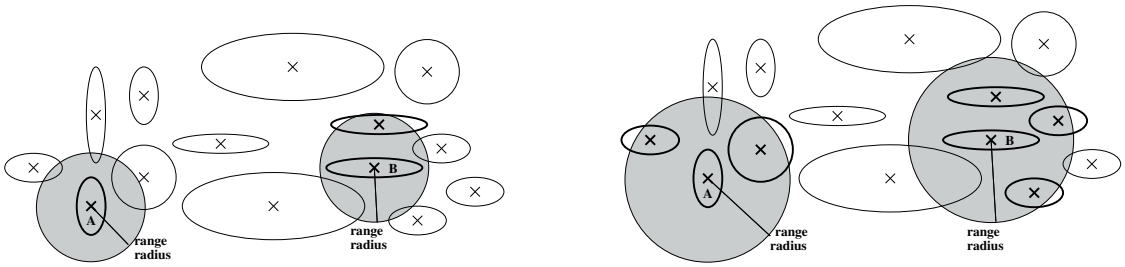


Figure 16. ***Rayon de « range query » et catégories voisines****. A et B sont deux catégories requêtes. Selon le rayon de « range query », plus ou moins de catégories voisines sont prises en compte.*

régions constituantes appartiennent. La stratégie de recherche s'étend facilement au cas de plusieurs catégories requête positives et plusieurs catégories requête négatives (requête par composition de catégories de régions par « range query »). Le système peut alors retrouver rapidement des images à partir de requêtes complexes par composition.

6.4.3.3. Résultats et interface utilisateur :

L'interface de requête (fig. 17) présente les vignettes des 91 régions représentatives que l'utilisateur peut sélectionner afin de spécifier les catégories requête positives et les négatives. Le rayon de « range query » qui définit l'étendue du voisinage des catégories peut être ajusté pour affiner les types de régions recherchées (fig. 17). Ce paradigme de recherche est résolument différent de la recherche de régions dans l'exemple précédemment développé. En effet, ici, l'utilisateur spécifie la composition des images qu'il recherche à partir de types de régions représentatifs de la base.

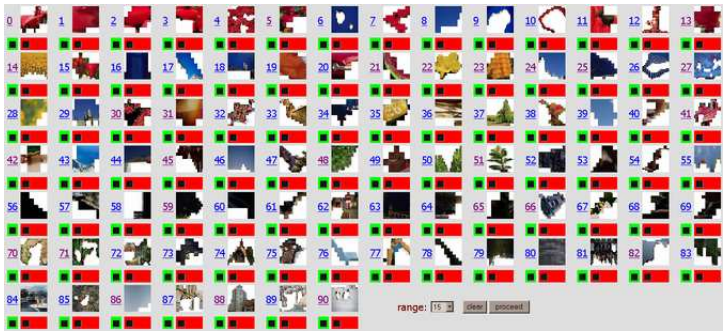


Figure 17. L'interface de requête permet de sélectionner parmi les 91 catégories celles qui constituent la requête. Le menu déroulant permet d'ajuster l'étendue du voisinage des catégories.

Un exemple de requête est illustré dans la figure 18. La requête est traduite automatiquement par le système sous forme d'une expression logique illustrant les représentants sélectionnés ainsi que les représentants des catégories voisines prises en compte pour le rayon de « range query » donné.

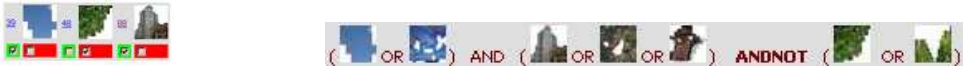


Figure 18. Illustration d'une requête. Gauche : les catégories 39 (bleu) et 88 (gris) sont les requêtes positives et 39 (vert) la négative. Droite : expression logique de la requête avec les représentants des catégories requêtes et des catégories voisines.

Les temps de recherche sont au maximum de 0.03 seconde pour les requêtes les plus complexes sur la base de 9995 images.

Figure 19. Images retournées à partir de la requête par composition.

6.4.3.4. Conclusions :

L'originalité de cette approche est basée sur le regroupement des régions similaires de la base. Elle présente les trois avantages suivants :

1. les images sont retrouvées à partir d'une composition de types de régions
2. les catégories de régions et leurs voisines permettent des « range queries »
3. l'implantation de l'indexation couleur réduit considérablement les temps de recherche

Cette approche, au même titre que le bouclage de pertinence, permet de réduire l'écart entre l'attente de l'utilisateur et les images retrouvées. La contrainte de composition dans la recherche d'images traduit une certaine « sémantique visuelle » latente dans les images.

Note : ces travaux ont été publiés dans un rapport de recherche [19]. Les travaux précédents de recherche par régions (extraction et description de régions) sont parus cette année ([6][5][7]).

8. Actions régionales, nationales et internationales

8.1. Actions nationales

8.1.1. Projet du Réseau National de Recherche en Télécommunications (RNRT) RECIS (Recherche et Exploration par le Contenu Image et Son)

Participants : Nozha Boujemaa, Jean-Paul Chieze.

C'est un projet de la catégorie exploratoire qui implique le projet IMEDIA, France Télécom, l'INSA de Lyon et Nouvelles Frontières. Le consortium a obtenu un financement de 3,8 MF sur 3 ans dont 1,2 MF sont affectés pour l'équipe IMEDIA de l'INRIA - notification juin 99.

8.1.2. Projet du Programme pour la Recherche et l'Innovation dans l'Audiovisuel et le MultiMédia (PRIAMM) « MédiaWorks »

Participants : Nozha Boujemaa, François Fleuret, Valerie Gouet, Bertrand Le Saux, Julien Fauqueur, Hichem Sahbi.

Il s'agit de concevoir et de développer une plate-forme générique pour l'indexation et la recherche de documents audiovisuels par le contenu. Ce contrat met le projet IMEDIA en collaboration avec une équipe de linguistes du LIMSI (CNRS), un fournisseur de contenu TF1, et les sociétés AEGIS et EML. L'originalité de ce projet est la réalisation d'un mécanisme d'indexation et de recherche hybride pluri-média image et texte. Le budget global est de 10MF dont 2,2MF pour l'INRIA (IMEDIA et VISTA) sur 3 ans dont 1,7MF pour IMEDIA.

8.2. Actions européennes

IMEDIA participe activement à la proposition XVIS (NeO) et prépare dans le cadre du consortium AIR&D un projet européen AceMedia.

8.3. Actions internationales

8.3.1. Projet LIAMA « Advanced medical imaging methods for Hominid Morphology studies »

Participant : Anne Verroust.

C'est un projet de coopération franco-chinoise pluridisciplinaire : il implique des paléontologues, des radiologistes et des informaticiens. Les responsables de ce projet sont Marc Jaeger du CIRAD-AMIS pour la partie française et Hanqing Lu du CAS-IA-NLPR pour la partie chinoise. Il s'agit d'introduire des outils de visualisation et des méthodes de caractérisation géométrique de données tridimensionnelles pour améliorer la restauration des ossements d'hominidés, leur visualisation interactive et leur caractérisation morphologique.

Anne Verroust est un des participants de ce projet et a effectué une visite à Pékin dans ce cadre fin octobre 2002.

9. Diffusion des résultats

9.1. Animation de la Communauté scientifique

Nozha Boujema :

- Membre d'un groupe de travail NSF/Delos qui a pour objectif la rédaction d'un « white paper » prospectif sur le theme de l'indexation et la recherche par le contenu et ses applications pour les archives d'héritage historique et culturel. Ce rapport a été remis à la NSF et a la Commission Européenne via Delos en décembre 2002.
- Membre du steering committee du NoE Xvis et responsable de la coordination entre les équipes INRIA participantes.
- Membre de l'action spécifique CNRS « Indexation Multimedia ».
- Co-éditeur d'un numéro spécial de la revue TSI sur l'indexation d'images fixes et animées.
- Membre des comités de programme de : WWW Multimedia, ACM Multimedia (chargée de la coordination des workshops), IEEE Fuzzy Systems, ORASIS,CBMI.
- Chair de la session « On Multimedia Content Based Retrieval For Cultural Heritage Applications » au Workshop MIR 2002 en parallele avec ACM Multimedia 2002
- Chargée de l'organisation de session spéciale « Fuzzy for CBIR » pour IEEE Fuzzy Systems.
- Invitée pour presentation en sessions plénières pour les conférences ICISP' et ISCIII'03.
- Orateur invitée pour une session spéciale organisée par Alberto Del Bimbo au workshop MIR (Multimedia Information Retreival) en conjonction avec ACM Multimedia 2002.
- relectrice pour les revues IEEE Trans. on PAMI, IEEE Trans. on IP, IEEE video and IEEE Transactions on Knowledge and Data Engineering, IEEE Vision, Image and Signal Processing.
- Experte pour le réseau national RIAM.
- Experte pour la Commission Européenne pour l'évaluation des projets soumis dans le cadre des FET (Future and Emerging Technologies).
- Membre des commissions de détachement/délégation de l'INRIA Rocquencourt, membre du GRECOS (prospective scientifique) Rocquencourt.
- Responsable des Relations Internationales de l'unité de Rocquencourt depuis octobre 2002.
- Rapporteur de la thèse de D. Vitale sur la reconnaissance des personnes par les empreintes digitales (universite de St-Etienne).

Anne Verroust :

- présidente de l'AFIG (Association Française d'informatique Graphique) ;
- responsable du pôle « informatique graphique » au GdR ALP (Algorithmique, Langage et Programmation) ;
- membre de la commission de spécialistes de l'université de Lille 1 ($27^{ème}$ section).

François Fleuret :

- Membre de la commission de bourses post-doctorales de l'INRIA.
- Orateur invité au colloquium PRCV 2002 à l'université de Prague.
- Participation à des comités de lecture : PAMI, JMLR.

Valérie Gouet :

- Jeudi 21 Mars 2002, Cordelia Schmid, Valérie Gouet, Patrick Gros et Sébastien Gilles, Séminaire IN'Tech INRIA Rhône-Alpes : Recherche par le contenu de documents multi-medias.

- Vendredi 22 février 2002, Participation au groupe de travail ERCIM « Image Understanding », Amsterdam.
- Participation à des comités de lecture : MDDE, ACM Multimedia.

Jean-Philippe Tarel :

- Membre du comité de programme du *IEEE Workshop on Omnidirectional Vision (Omnivis'2002)*, Copenhagen, Danemark, Juin 2002.
- Expert pour jury de TFE de l'ENTPE.

9.2. Enseignement

Nozha Boujemaa : Chargé de cours à l'UTC dans le cadre de la filière « Ingénierie des Industries Culturelles »
Coordinatrice et chargée de cours Option Multimedia de la 3ème année de la filière d'ingénieurs SupCom ainsi que du DEA STIC également a SupCom - Tunis
Participation à un séminaire professionnel pour les documentalistes : Cours INRIA-IST sur la « Recherche d'informations sur les reseaux », Le Bono 2002
Donald Geman : Chargé de cours au DEA de l'ENS Cachan.
Anne Verroust : Chargée de cours à l'ENSTA (cours d'algorithmique géométrique aux élèves de troisième année dans le module image et vision)
Valérie Gouet :
cours « Vision par Ordinateur » en Option Informatique de l'École des Mines d'Alès.
cours « Indexation d'images par descripteurs locaux » en dernière année de Supcom Tunis.
cours au CNAM (« Introduction à l'image » et « atelier photoshop » DEA ESTC option CAM) TP au CNAM (« Algorithmique et programmation » NOHA cycle A).
Julien Fauqueur : Chargé de TD et TP d'algorithmique en DEUG MASS à l'université Paris IX - Dauphine.
Bertrand Le Saux : Chargé de TD et TP d'algorithmique en DEUG MASS à l'université Paris IX - Dauphine.

10. Bibliographie

Articles et chapitres de livre

[1] N. BOUJEMAA, ET AL.. *Recherche d'informations sur les réseaux.* ADBS éditions, 2002, chapitre Recherche interactive dans les documents multimédia.

Communications à des congrès, colloques, etc.

[2] S. BOUGHORBEL, N. BOUJEMAA, C. VERTAN. *Histogram-Based Color Signatures for Image Indexing.* in « Proc. of IPMU 2002 », Annecy, France, 1-5 Juillet, 2002.

[3] N. BOUJEMAA, ET AL.. *White paper : REPORT OF THE DELOS-NSF working group on Digital Imagery for Significant Cultural and Historical Materials.* 2002.

[4] N. BOUJEMAA, M. FERECATU, V. GOUET. *Approximate search vs. precise search by visual content in cultural heritage databases.* in « 4th Intl. Workshop on Multimedia Information Retrieval (MIR'2002) in conjunction with ACM Multimedia 2002 », Juan-les-Pins, France, décembre, 2002.

[5] J. FAUQUEUR, N. BOUJEMAA. *Coarse Detection and Fine Color Description for Region-Based Image Queries.* in « Proc. of IEEE International Conference on Pattern Recognition (ICPR'02) », 2002.

[6] J. FAUQUEUR, N. BOUJEMAA. *Image Retrieval by Regions : Coarse Segmentation and Fine Color Description.* in « Proc. of International Conference on Visual Information System (VIS'02), Hsin-Chu, Taiwan », 2002.

[7] J. FAUQUEUR, N. BOUJEMAA. *Region-based Retrieval : Coarse Segmentation with Fine Signature.* in « Proc. of IEEE International Conference on Image Processing (ICIP'02) », 2002.

[8] F. FLEURET, D. GEMAN. *Fast Face Detection with Precise Pose Estimation.* in « Proceedings of ICPR2002 », 2002.

[9] V. GOUET, N. BOUJEMAA. *On the robustness of color points of interest for image retrieval.* in « IEEE International Conference on Image Processing (ICIP'2002) », Rochester, New York, USA, septembre, 2002.

[10] F. ROSSI, B. CONAN-GUEZ, F. FLEURET. *Functional Data Analysis With Multi Layer Perceptrons.* in « Proceedings of IJCNN 2002 (WCCI 2002) », volume 3, IEEE/NNS/INNS, pages 2843-2848, Honolulu, Hawai, USA, May, 2002.

[11] F. ROSSI, B. CONAN-GUEZ, F. FLEURET. *Theoretical Properties of Functional Multi Layer Perceptrons.* in « Proceedings of ESANN 2002 », pages 7-12, Bruges, Belgium, April, 2002.

[12] H. SAHBI, N. BOUJEMAA. *Coarse-to-fine face detection based on skin color adaptation.* in « Proceedings of ECCV's 2002 Workshop on Biometric Authentication », pages 112-120, 2002.

[13] H. SAHBI, N. BOUJEMAA. *Robust face recognition using Dynamic Space Warping.* in « Proceedings of ECCV's Workshop on Biometric Authentication », pages 121-132, 2002.

[14] H. SAHBI, D. GEMAN, N. BOUJEMAA. *Face Detection Using Coarse-to-Fine Support Vector Classifiers.* in « Proceedings of ICIP », pages 925-928, 2002.

[15] J.-P. TAREL, S. BOUGHORBEL. *On the Choice of Similarity Measures for Image Retrieval by Example.* in « Actes ACM MultiMedia Conference », Juan-Les-Pins, France, 2002, http://www-rocq.inria.fr/~tarel/acm02.html.

[16] A. VERROUST, M. FINIASZ. *A control of smooth deformations with topological change on a polyhedral mesh based on curves and loops.* in « International Conference on Shape Modelling and Applications (SMI 2002) », Banff, Alberta, Canada, mai, 2002.

[17] B. LE SAUX, N. BOUJEMAA. *Unsupervised Robust Clustering for Image Database Categorization.* in « International Conference on Pattern Recognition », august, 2002.

[18] B. LE SAUX, N. BOUJEMAA. *Unsupervized Categorization for Image Database Overview.* in « International Conference on Visual Information System (VISUAL'2002) », march, 2002.

Rapports de recherche et publications internes

[19] J. FAUQUEUR, N. BOUJEMAA. *Image retrieval by range query composition of region categories.* rapport technique, numéro RR-4686, INRIA, Rocquencourt, France, décembre, 2002, http://www.inria.fr/rrrt/rr-4686.html.

[20] V. GOUET, N. BOUJEMAA. *About optimal use of color points of interest for content-based image retrieval.* rapport technique, numéro RR-4439, INRIA, Rocquencourt, France, avril, 2002, http://www.inria.fr/rrrt/rr-4439.html.

[21] N. GRIRA. *Indexation d'images par descripteurs locaux.* Rapport de stage de fin d'étude, Ecole Supérieure des Télécommunications de Tunis, Tunisie, 2002.

[22] H. SAHBI, F. FLEURET. *Scale invariance of Support Vector Machines based on the Triangular Kernel.* rapport technique, numéro RR-4601, INRIA, october, 2002, http://www.inria.fr/rrrt/rr-4601.html.

[23] A. TOURNIER. *Etude de structures d'index multidimensionnels pour la recherche d'images par points d'intérêt.* Rapport de DEA SIR, Université de Paris VI, France, 2002.

Bibliographie générale

[24] N. BOUJEMAA. *"Sur la classification non-exclusive en analyse d'images".* habilitation à diriger des recherches, Université de Versailles-Saint-Quentin, 2000.

[25] N. BOUJEMAA, F. FAUQUEUR, M. FERECATU, F. FLEURET, V. GOUET, B. LE SAUX, H. SAHBI. *Interactive Specific and Generic Image Retrieval.* in « Proceedings of MMCBIR 2001 », 2001.

[26] C. BURGES, B. SCHOLKOPF. *Improving the Accuracy and Speed of Support Vector Machines.* in « Neural Information Processing Systems, Cambridge. MIT Press », 1997, pages 375-381.

[27] N. CHRISTIANI, J. SHAWE-TAYLOR. *An Introduction to Support Vector Machines and other kernel-based learning methods.* Cambridge University Press, 2000.

[28] R. CUCCHIARA, F. FILICORI. *The Vector-Gradient Hough Transform.* in « IEEE Transactions on Pattern Analysis and Image Intelligence », numéro 7, volume 20, 1998.

[29] F. FLEURET. *Détection hiérarchique de visages par apprentissage statistique.* thèse de doctorat, Université Paris-VI, Paris, 2000.

[30] V. GOUET, N. BOUJEMAA. *Object-based queries using color points of interest.* in « IEEE Workshop on Content-Based Access of Image and Video Libraries (CBAIVL/CVPR 2001) », pages 30-36, Kauai, Hawaii, USA, 2001.

[31] V. GOUET, P. MONTESINOS. *Normalisation des images en couleur face aux changements d'illumination.* in « 13ème Congrès Francophone AFRIF-AFIA de Reconnaissance des Formes et Intelligence Artificielle », volume II, pages 415-424, Angers, France, 2002.

[32] N. KATAYAMA, S. SATOH. *The SR-tree : an index structure for high-dimensional nearest neighbor queries.* pages 369-380, 1997, http://citeseer.nj.nec.com/katayama97srtree.html.

[33] F. LAZARUS, A. VERROUST. *Level Set Diagrams of polyhedral objets.* in « SMA'99 (Fifth ACM Symposium on Solid Modeling and Applications) », ACM Press, Ann Arbor, juin, 1999.

[34] M. B. H. RHOUMA, H. FRIGUI. *Self-Organization of Pulse-Coupled Oscillators with Application to Clustering.* in « IEEE Transactions on Pattern Analysis and Machine Intelligence », numéro 2, février, 2001.

[35] H. SAHBI, D. GEMAN, N. BOUJEMAA. *Face detection using Coarse-to-Fine Support Vector Classifiers.* in « Proceedings of ICIP2002 », 2002.

[36] A. VERROUST, F. LAZARUS. *Extracting Skeletal Curves from 3D Scattered Data.* in « The Visual Computer », numéro 1, volume 16, 2000, pages 15-25.

[37] R. WEBER. *Similarity Search in High-Dimensional Vector Spaces.* Thèse de doctorat, Swiss Federal Institute of Technology, Zurich, décembre, 2000.