

## *Projet IMEDIA*

*Images et Multimédia : Indexation, Recherche et Navigation*

*Rocquencourt*

THÈME 3B

*R* *apport*  
*d'Activité*

2001



## Table des matières

<b>1</b>	<b>Composition de l'équipe</b>	<b>3</b>
<b>2</b>	<b>Présentation et objectifs généraux</b>	<b>3</b>
<b>3</b>	<b>Fondements scientifiques</b>	<b>4</b>
3.1	Indexation d'images et espace de description . . . . .	4
3.2	Requêtes partielles . . . . .	5
3.3	Recherche interactive . . . . .	6
3.4	Indexation Multimédia . . . . .	6
<b>4</b>	<b>Domaines d'applications</b>	<b>7</b>
<b>5</b>	<b>Logiciels</b>	<b>8</b>
5.1	Logiciel IKONA/MAESTRO . . . . .	8
<b>6</b>	<b>Résultats nouveaux</b>	<b>8</b>
6.1	Descripteurs d'images : Bases génériques . . . . .	8
6.1.1	Histogrammes accumulatifs pour la distribution géométrique globale de la couleur . . . . .	8
6.2	Descripteurs d'images : Bases spécifiques . . . . .	10
6.2.1	Indexation et recherche dans de grandes bases d'empreintes digitales . .	10
6.2.2	Détection et localisation précise de visages . . . . .	13
6.2.3	Machines à vecteurs de support pour un détecteur hiérarchique des visages . . . . .	15
6.3	Requêtes partielles . . . . .	18
6.3.1	Requêtes par points d'intérêt . . . . .	18
6.3.2	Requêtes partielles par régions d'intérêt . . . . .	21
6.4	Recherche interactive . . . . .	23
6.4.1	Méthodes de recherche d'images par boucle de pertinence . . . . .	23
6.4.2	Résumé de bases par ses images les plus informatives . . . . .	24
<b>7</b>	<b>Contrats industriels (nationaux, européens et internationaux)</b>	<b>26</b>
7.1	Contrat Thales-Identification . . . . .	26
<b>8</b>	<b>Actions régionales, nationales et internationales</b>	<b>27</b>
8.1	Actions nationales . . . . .	27
8.1.1	Projet du Réseau National de Recherche en Télécommunications (RNRT) RECIS (Recherche et Exploration par le Contenu Image et Son)	27
8.1.2	Projet du Programme pour la Recherche et l'Innovation dans l'Audiovisuel et le MultiMédia (PRIAMM) « MédiaWorks » . . . . .	27
8.1.3	9ème Rencontres INRIA-Industrie . . . . .	27
8.2	Actions européennes . . . . .	29
8.2.1	Projet Européen « STOP » . . . . .	29

8.3	Actions internationales . . . . .	29
8.3.1	Projet STIC INRIA « Indexation et recherche d'images par le contenu : application aux images satellitaires » . . . . .	29
8.3.2	Consortium AIR-D . . . . .	30
8.4	Visites, et invitations de chercheurs . . . . .	30
<b>9</b>	<b>Diffusion de résultats</b>	<b>30</b>
9.1	Animation de la Communauté scientifique . . . . .	30
9.2	Enseignement . . . . .	32
<b>10</b>	<b>Bibliographie</b>	<b>32</b>

## 1 Composition de l'équipe

### Responsable scientifique

Nozha Boujemaa [détachée CR1 INRIA]

### Assistante de projet

Laurence Bourcier [à temps partiel dans le projet]

### Personnel INRIA

François Fleuret [CR2 depuis le 15/10/2001]

Jean-Philippe Tarel [détaché CR2 depuis le 1/11/2001]

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]

### Personnel université

Marie-Aude Aufaure [MdC université de Lyon en congé de recherche depuis le 1/09/2001]

### Ingénieurs experts

François Fleuret [temps partiel jusqu'au 30 juin 01]

Valérie Gouet [depuis depuis le 15/12/2000]

### Doctorants

Sylvain Bernard [Bourse CIFRE, Thales-Identification 1/09/1998 - 31/10/2001]

Sabri Boughorbel [Bourse INRIA Rocq depuis le 1/11/2001]

Julien Fauqueur [Bourse INRIA Rocq depuis le 1/03/2000]

Marin Ferecatu [Bourse INRIA Rocq depuis le 1/10/2001]

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

Rami Langar [février-juin 2000]

Nizar Grira [juillet-septembre 2001]

Sabri Boughorbel [mars-août 2000]

## 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 multimédia. Ces documents sont par essence hétérogènes, et intègrent aussi bien le texte que l'image, le graphique, la vidéo et le son. En particulier, les données visuelles dominent les documents multimédia, aussi bien par leur bande passante que par la complexité de leur contenu. La description de ce contenu par des techniques d'analyse d'images étant moins subjective que la seule description habituelle par des mots clés, elle est primordiale pour l'exploration efficace d'un flot multimédia.

L'objectif de l'équipe IMEDIA est de développer des méthodes d'indexation par le contenu, de recherche interactive et de navigation dans des bases d'images, dans un contexte multimédia. Pour ce faire, 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 ciblé, appelées également bases avec vérité terrain, dans lesquelles les images ont un contenu homogène (visages, images médicales, empreintes digitales, etc.). De ce fait, nos recherches sont à l'intersection de plusieurs disciplines scientifiques, dont les principales sont l'analyse d'images, la reconnaissance d'objets, les bases de données, la recherche d'informations, la théorie de l'information, l'apprentissage et l'interaction homme-machine.

Notre travail de recherche s'articule autour de quatre axes principaux :

1. l'indexation, qui consiste à élaborer et optimiser des descripteurs génériques de l'aspect visuel des images, et des descripteurs spécifiques à une catégorie de bases d'images dédiées à une application particulière ;
2. les requêtes partielles, qui ont pour objectif de développer des méthodes permettant de délimiter et de focaliser spatialement la requête de l'utilisateur sur des zones d'intérêt dans l'image ;
3. la recherche interactive, qui se répartit dans les deux thèmes suivants : « le contrôle de pertinence » et « la navigation intelligente » dans les bases d'images ;
4. l'indexation plurimédia, et en particulier l'indexation bi-modale texte/image, qui a pour objectif de faire coopérer ces deux médias 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 multimédia dans sa globalité.

### 3 Fondements scientifiques

*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, requêtes partielles, recherche interactive et indexation multimédia. Dans ce qui suit, nous présentons une introduction à chacun de ces thèmes.*

#### 3.1 Indexation d'images et espace de description

**Mots clés :** extraction de caractéristiques, analyse d'image, reconnaissance de forme, mise en correspondance, statistiques, base de données image.

**Participants :** Nozha Boujemaa, Sylvain Bernard, Hichem Sahbi, Valérie Gouet, Michel Scholl, Jean-Paul Chièze, Jean-Philippe Tarel.

##### Glossaire :

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

Les données visuelles dominent les documents multimédia, aussi bien par leur volume que par la complexité de leur contenu. L'objectif d'IMEDIA est d'offrir la possibilité d'interroger

les bases d'images (et grâce à des collaborations scientifiques, les bases multimédia) par le contenu, d'une manière ergonomique 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 que nous décrivons ci-après.

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

Notons que pour les bases spécifiques (pour lesquelles il existe une vérité-terrain), on développe des descripteurs dédiés et optimaux pour la tâche considérée (reconnaissance de visages, etc.). Pour les bases génériques, à l'inverse, on extrait des descripteurs universels (couleur, texture, forme, etc.).

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.

Avec l'arrivée de Jean-Philippe Tarel, nous comptons développer l'axe signatures de formes. Une collaboration suivie sur les aspects constructions d'index mutlidimensionnels a été engagé avec Michel Scholl (INRIA/CNAM).

## 3.2 Requêtes partielles

**Mots clés :** Détection, segmentation, classification compétitive, apprentissage statistique, similarité perceptuelle, relation spatiale.

**Participants :** Nozha Boujemaa, Julien Fauqueur, Valerie Gouet.

La plupart des moteurs de recherche d'images par le contenu (académiques ou industriels) proposent le mode de requête par l'exemple, où l'utilisateur désigne une image entière comme requête au système. Les requêtes partielles présentent une alternative intéressante répondant à des besoins réels par rapport à la requête par une image exemple. En effet, l'utilisateur est le plus souvent amené à effectuer une recherche sur une partie de l'image et non sur l'image entière. Le but du système serait alors de retrouver les images de la base qui contiennent des parties similaires. Le problème présente alors une complexité accrue et les questions soulevées sont les suivantes :

- comment décrire localement les images (avec ou sans segmentation) et constituer ces zones d'intérêt d'une manière efficace et pertinente,
- comment gérer les mesures de similarité perceptuelle locale et l'indexation de la base par les signatures partielles,

- comment gérer le problème des relations spatiales dans le contexte des requêtes partielles multiples : avec ou sans contraintes de positions relatives, d'invariances morphologiques, etc.

### 3.3 Recherche interactive

**Mots clés** : similarité perceptuelle, statistique, catégorisation, théorie de l'information.

**Participants** : Nozha Boujemaa, Donald Geman, Bertrand Le Saux, Marin Ferecatu.

L'interrogation des bases d'images est un exemple typique de recherche interactive puisque, paradoxalement, dans la plupart des cas (notamment les applications grand public), l'utilisateur ne sait pas ce qu'il cherche tant qu'il ne l'a pas trouvé. Il s'agit donc d'aider l'utilisateur à voir plus clair dans le flux de données auquel il est confronté, mais aussi d'optimiser les traitements qui doivent s'effectuer en temps-réel. De ce fait, la phase de recherche interactive est primordiale.

Dans un cadre interactif, les actions de l'utilisateur peuvent avoir deux objectifs :

- *Contrôle de pertinence*. Cette tâche consiste à permettre à l'utilisateur d'interagir avec le système en indiquant parmi les images retrouvées lesquelles sont pertinentes par rapport à celles qu'il recherche et lesquelles ne le sont pas. Le système pourra ainsi apprendre et s'adapter à l'intention de chaque utilisateur, via les exemples positifs et négatifs indiqués. Cette fonctionnalité permet au système un comportement flexible et adaptatif au profil de l'utilisateur.
- *Navigation*. Ce problème consiste à estimer la distribution (habituellement multimodale) des catégories d'images qui intéressent l'utilisateur. Il s'agit d'un cadre plus vaste dans lequel on se pose la question suivante : quelles sont, à tout instant, les images les plus informatives qui permettront à l'utilisateur d'appréhender le mieux possible le contenu de la base ? D'un point de vue pratique, il s'agit d'un problème central pour le schéma de la requête par l'exemple. Il s'agit de fournir à l'utilisateur une première image qui correspond à peu près à ce qu'il recherche (problème de la « page zéro »).

Ce problème peut englober le précédent : ces deux actions peuvent être couplées lors d'une session d'interrogation.

### 3.4 Indexation Multimédia

**Mots clés** : indexation hybride, annotation textuelle, théorie de l'information, estimation de l'incertitude.

**Participants** : Nozha Boujemaa, Marin Ferecatu, Julien Fauqueur, Sabri Boughorbel, Donald Geman, Marie-Aude Aufaure.

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 *métadonnées* (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 multimédia. Plusieurs pistes sont possibles pour faire collaborer en particulier les sources d'informations visuelles et textuelles dans un but d'indexation ou de recherche

L'une des pistes que nous comptons explorer consiste à construire une hiérarchie de concepts à partir d'une base d'images annotées. Cette hiérarchie contiendrait des associations structurales entre données textuelles (annotations) et entre données textuelles et descripteurs visuels. L'utilisateur pourra alors naviguer de façon non linéaire dans une grande base d'images à l'aide de cette hiérarchie de concepts (vue résumée de la base).

## 4 Domaines d'applications

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

- **les applications scientifiques**

Exemples : les bases d'images satellitaires : typologie des terrains ; bases d'images médicales : retrouver les images présentant un caractère pathologique, dans un but éducatif ou diagnostique.

- **l'audiovisuel**

Exemple : rechercher un plan spécifique d'un film ou d'un journal télévisé. Une collaboration est en cours avec la chaîne de télévision TF1. À long terme, cette application intègre l'image, le son et le texte.

- **l'authentification**

Exemples : détecter les contrefaçons de modèles déposés, 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 sont actuellement en cours sur ce thème dans le projet.

- **le design, la publicité**

Exemples : 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.

- **l'art, l'éducation**

Exemples : recherche encyclopédique, recherche d'un tableau ou d'une illustration par un exemple, un croquis, un détail. IMEDIA a été contacté par le ministère de la culture au sujet de ses archives en images.

- **les télécommunications**

Exemple : coder, représenter et rechercher les images par leur contenu sont des enjeux importants dans le contexte MPEG-4 et MPEG-7. 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 future norme.

## 5 Logiciels

### 5.1 Logiciel IKONA/MAESTRO

**Participants** : Toute l'équipe , François Fleuret, Marin Ferecatu [correspondant pour IKONA].

L'évolution de nos travaux de recherche ainsi que la validation des algorithmes que nous développons a nécessité une architecture logicielle différente de celle du moteur « Surfimage ». Ce dernier était une plate-forme expérimentale et a fait l'objet d'un rachat par la société LookthatUp, start-up de l'INRIA. Nous avons choisi de développer un nouvel outil générique d'indexation et de recherche par le contenu dans les bases de données images. Ce nouveau moteur est structuré sous la forme d'une architecture client/serveur : IKONA/MAESTRO. Il supporte l'intégration des requêtes textuelles et est adapté aux requêtes partielles multiples. Les structures de données intègrent dès la conception la dualité images et texte.

Le serveur, MAESTRO, a été développé en C++ sur une architecture UNIX. Il est conçu de manière à pouvoir exploiter des machines multi-processeurs pour gérer une charge de calcul plus importante dans un cadre multi-utilisateurs. MAESTRO inclut les signatures d'images avec les procédures adéquates de comparaison. La modularité de la conception objet nous permet de faciliter l'intégration des développements réalisés par différents contributeurs dans ce cadre.

Le client, IKONA, est écrit en Java, donc portable et facilement exploitable dans le contexte du WWW. La figure 1 montre l'interface du logiciel lors d'une requête.

Ce nouveau système fait déjà l'objet de plusieurs échanges académiques et industriels (cf. 7.1).

Une version interactive (mais aux fonctionnalités réduites) a été mise en place par Jean-Paul Chièze et est disponible sur le web d'IMEDIA : <http://www-rocq.inria.fr/cgi-bin/imedia/ikona>.

## 6 Résultats nouveaux

### 6.1 Descripteurs d'images : Bases génériques

#### 6.1.1 Histogrammes accumulatifs pour la distribution géométrique globale de la couleur

**Mots clés** : descripteurs, histogrammes, distribution géométrique de la couleur, découpage spatial.

**Participants** : Sabri Boughorbel, Nozha Boujemaa.

**Résumé** : *Nous avons mis au point une méthode pour rendre compte de la répartition géométrique de la couleur dans une image. Pour ce faire, nous calculons des distributions sur chaque partie d'un découpage de l'image, et nous proposons deux schémas de cumul de ces distributions en vue d'obtenir une description de la répartition globale des couleurs dans l'image. Les évaluations ont montré une*



FIG. 1 – Interface du logiciel IKONA. À gauche, première page avec des images prises au hasard dans une base généraliste. L'image sur laquelle va porter la requête est entourée. À droite, résultat de la requête.

*amélioration des performances en comparaison avec l'utilisation d'histogrammes de couleurs usuels.*

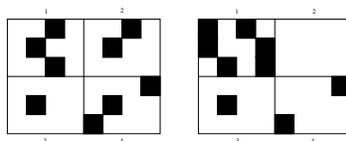


FIG. 2 – Deux images à histogrammes identiques mais avec une répartition géométrique différente des couleurs.

L'histogramme des couleurs permet d'avoir une information globale mais grossière sur la présence des couleurs dans les images [SB91]. Une caractérisation de la répartition locale des couleurs dans une image a été introduite par les histogrammes pondérés [VB00][15] récemment développés à IMEDIA. La puissance et la cohérence de cette caractérisation a été démontrée. La caractérisation de la répartition géométrique des couleurs est un prolongement de ce travail.

Lorsque les bases d'images deviennent de taille importante, le pouvoir de discrimination de l'histogramme standard de couleurs est limité. En effet, des images de contenu visuel différent

[SB91] M. SWAIN, D. BALLARD, « Color indexing », *International Journal of Computer Vision*, 7(1):11-32, 1991., 1991.

[VB00] C. VERTAN, N. BOUJEMAA, « Upgrading Color Distributions for Image Retrieval: can we do better? », *Visual Information and Information Systems*, 2000, p. 178-188.

peuvent avoir des histogrammes similaires. La figure 2 présente deux images différentes qui ont des histogrammes identiques (le nombre de pixels noirs dans les deux images est égal à 9). Par contre, si l'on subdivise l'image en quatre, le nombre de pixels dans la première fenêtre est très différent dans les deux images. L'information géométrique est absente dans l'histogramme de couleurs usuel.

L'idée est donc de découper chaque image en parties, de calculer la distribution des couleurs sur chaque partie, et de les cumuler pour obtenir une description globale [8]. Nous proposons deux façons de cumuler les distributions locales :

- **Cumul multiplicatif :**

$$\tilde{h}(c) = h(c) \prod_{i=1}^N (1 + h_i(c)), \quad \forall c \in \mathcal{C} \quad (1)$$

où  $h_i$  correspond à l'histogramme de la  $i^{\text{ème}}$  partie, et  $N$  est le nombre de parties.

- **Cumul additif :**

$$\tilde{h}(c) = \sum_{i=1}^N f(h_i(c)), \quad \forall c \in \mathcal{C} \quad (2)$$

où  $f$  est une fonction non-linéaire qui permet de mettre l'accent sur les couleurs fortement présentes.

Nous avons choisi des fonctions du type  $f(x) = x^m$  pour  $0 < m < 1$ , car elles sont croissantes, nulles à l'origine, et concaves.

## Expérimentations

Les expérimentations sont réalisées sur le moteur de recherche IKONA. L'évaluation se fait selon le critère précision/rappel sur une base d'images avec vérité-terrain.

Les meilleurs résultats ont été obtenus avec un cumul additif et avec  $m = \frac{1}{3}$ . La Fig. 3 montre l'amélioration des courbes de précision/rappel pour les différentes combinaisons. Une amélioration des performances de l'ordre de 6% est observée pour le cumul additif par rapport à l'histogramme usuel. On obtient une amélioration de 3% par rapport aux histogrammes pondérés (voir Fig. 6.1.1). Le choix de la fonction  $f$  lors du cumul additif s'avérant discriminant, une étude plus approfondie de la métrique sur l'espace des couleurs est nécessaire.

Les histogrammes introduits (pondéré et accumulatif) sont plus performants en terme de précision/rappel que l'histogramme usuel. La simplicité de calcul des descripteurs, des mesures de similarité ainsi que la robustesse des réponses obtenues pour le descripteur accumulatif (notamment pour le cumul additif) nous conduit à mener une étude pour généraliser l'application de cette approche à d'autres types de descripteurs tel que les descripteurs de formes.

## 6.2 Descripteurs d'images : Bases spécifiques

### 6.2.1 Indexation et recherche dans de grandes bases d'empreintes digitales

**Mots clés :** empreintes digitales, segmentation, classification, mise en correspondance de

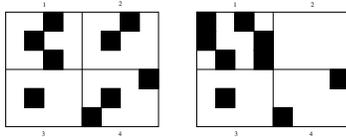


FIG. 3 – Courbes de précision/rappel avec un cumul additif (+), cumul multiplicatif ( $\times$ ) et l’histogramme de couleur (\*).

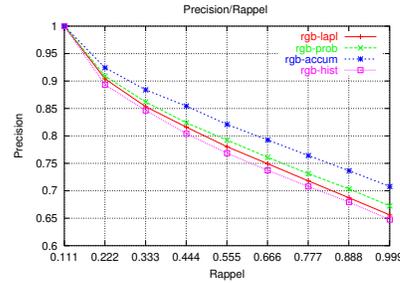


FIG. 4 – Comparaison des courbes de précision/rappel avec cumul additif (\*), et avec des histogrammes pondérés (laplacien, probabiliste et usuel).

points, recalage.

**Participants :** Sylvain Bernard, Nozha Boujema.

**Résumé :** *La thèse de Sylvain Bernard a été soutenue le 9 Novembre 2001 à l’Université de Versailles-St Quentin en Yvelines [1]. Cette thèse s’est déroulée dans le cadre d’un contrat CIFRE avec la société THALES Identification (ex THOMSON-CSF Idmatics). Cette société qui est spécialisée dans la fourniture de systèmes de production de titres d’identité (carte d’identité, passeport, permis de conduire ...) a besoin d’un système propriétaire de reconnaissance des empreintes digitales (AFIS). En effet, un tel système permettrait de vérifier, avant de produire une carte, si le demandeur est déjà présent ou non dans la base des personnes possédant déjà une carte. Les clients de THALES Identification sont généralement des Etats (Pologne, Kenya, Namibie) ; le système doit alors effectuer ses recherches dans de très grandes bases d’empreintes et est soumis à de fortes contraintes de temps.*

Nous rappelons qu’une empreinte est caractérisée par un ensemble de points particuliers appelés minuties qui sont les points de fin ou de bifurcation des lignes de crêtes composant l’empreinte (Fig. 5). C’est la disposition géométrique des minuties les unes par rapport aux autres qui permet de savoir si deux empreintes sont issues du même doigt ou non. Puisque la détection automatique des minuties, qui nécessite une étape de segmentation d’image [5], a été réalisée lors des deux premières années de thèse, nous avons cette année particulièrement concentré nos efforts sur la mise en correspondance des minuties. Ce problème de recalage se heurte aux déformations rigides et non-rigides inhérentes à la saisie. En effet, l’utilisateur ne positionne jamais son doigt au même endroit sur le capteur et l’élasticité de la peau introduit des distorsions. L’algorithme doit de plus être robuste face au bruit et aux occultations. La méthode se doit aussi d’être rapide afin que le système puisse répondre au flux journalier important de requêtes. Pour apporter une réponse à ce problème, nous avons, l’année passée,

testé l'algorithme de la transformée de Hough généralisée [JH97] et constaté son manque de robustesse face au bruit et aux occultations. Nous avons alors enrichi l'algorithme d'une mesure de similarité locale qui permet de tenir compte de la relation spatiale des minuties entre elles. Nous constatons alors un gain en précision important mais l'algorithme n'effectuait que 150 comparaisons par seconde.

Cette année, nous avons introduit des invariants locaux entre minuties voisines ce qui rend le calcul de la mesure de similarité locale beaucoup plus rapide et nous permet de dépasser les 1000 comparaisons par seconde. Dans un second temps, nous avons cherché à nous situer par rapport à la communauté internationale en testant nos algorithmes sur les bases de la compétition internationale FVC2000 (ICPR2000, Barcelone, Septembre 2000). Nous nous positionnons alors en seconde place selon ce benchmark [MMC<sup>+</sup>00]. Dans un troisième temps et après une phase d'optimisation de code poussée, nous avons intégré ces développements dans les produits de la société THALES Identification, ce qui lui a d'ors et déjà permis de remporter plusieurs contrats. Parallèlement, nous avons enrichi les possibilités d'indexation et de recherche de notre logiciel IKONA sur un nouveau type de bases spécifiques : les bases d'empreintes digitales (Fig. 6).

Enfin, les six derniers mois ont été consacrés à la rédaction du mémoire de thèse et à la préparation de la soutenance finale. La soutenance a eu lieu le 9 Novembre 2001 à l'Université de Versailles-St Quentin en Yvelines devant le Jury composé de :

- Président du jury : M. Yves Meyer (ENS Cachan) ;
- Rapporteurs : Mme Françoise Preteux (Institut National des Télécommunications), M. Georges Stamon (Université Paris V) ;
- Examineurs : Mme Nozha Boujemaa (INRIA Rocquencourt), M. Claude Bricot (THALES Identification), M. Claude Timsit (Université de Versailles-St Quentin).

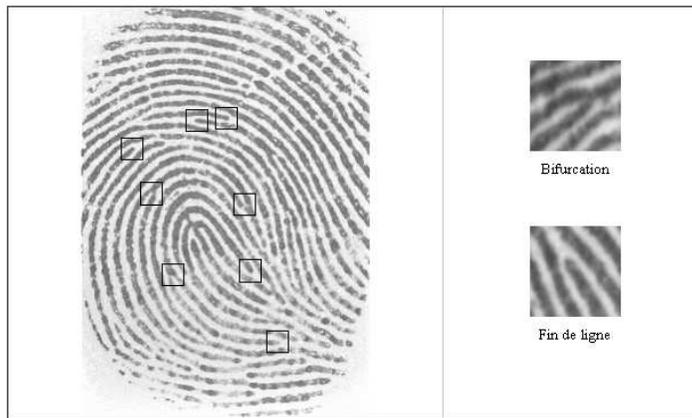


FIG. 5 – Exemples de minuties

[JH97] A. JAIN, L. HONG, « On-line Fingerprint Verification », *PAMI* 19, 4, 1997, p. 302–314.

[MMC<sup>+</sup>00] D. MAIO, D. MALTONI, R. CAPPELLI, J. L. WAYMAN, A. K. JAIN, « FVC2000 : Fingerprint Verification Competition », *in: Proc. of ICPR2000*, Barcelona, 2000.

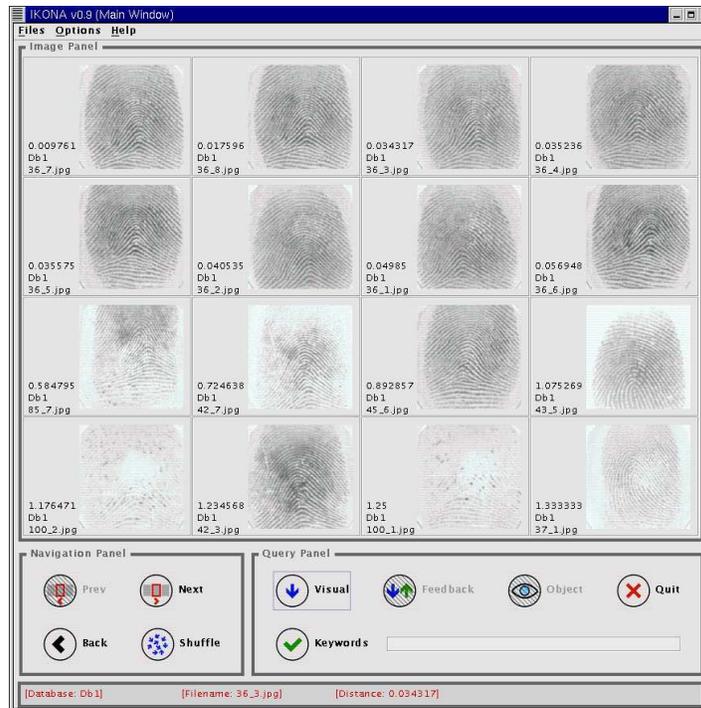


FIG. 6 – Requête sur base d’empreintes avec IKONA

### 6.2.2 Détection et localisation précise de visages

**Mots clés :** apprentissage statistique, détection, optimalité algorithmique, vision artificielle, représentation hiérarchique.

**Participants :** François Fleuret, Donald Geman.

#### Glossaire :

**Détection de visage** estimation automatique du nombre et des emplacements des visages présents dans une image

**Évaluation paresseuse** calcul effectué de manière économique en n’évaluant un terme que si sa valeur influe sur la valeur finale de l’expression

**Pose** position du visage dans le plan image

**Résumé :** *La détection automatique de visages permet d’associer à des images une sémantique de haut niveau sans nécessiter l’intervention d’un opérateur humain. Elle est donc très intéressante dans un cadre de recherche d’images par le contenu. Nous avons continué à développer notre algorithme rapide de détection en affinant l’estimation de la pose des visages détectés, et en l’intégrant à la plateforme IKONA/MAESTRO.*

Notre objectif est de concevoir un algorithme capable de localiser rapidement tous les visages dans une scène, c’est à dire capable d’estimer précisément la position des yeux de

chaque visage (cf. figure 7). Notre approche repose sur une hiérarchie de classifieurs capables de déterminer si un visage dont la pose est plus ou moins contrainte est présent dans une image donnée. Chaque classifieur de cette hiérarchie est construit par apprentissage, à partir d'une petite base d'images de visages.



FIG. 7 – Résultat de la détection sur une image. Les sommets des triangles indiquent les positions des yeux, et une estimation grossière de la position de la bouche.

Notre approche repose sur une représentation hiérarchique des images de visages[3]. Elle a deux avantages fondamentaux. Le premier est théorique : comme les variations globales de la pose du visage dans le plan image sont prises en compte au niveau de la hiérarchie elle même, les classifieurs n'ont pas besoin d'induire l'invariance à la pose à partir de la base d'apprentissage. Ils peuvent donc être très simples et peu sujet au sur-apprentissage. Le deuxième avantage est algorithmique : si l'on combine une hiérarchie de ce type et une évaluation paresseuse qui s'arrête dès qu'un classifieur a répondu négativement, l'algorithme concentre le calcul sur les parties ambiguës et complexes de l'image, et est optimal sous certaines hypothèses. De fait, il est en pratique très rapide (une approche similaire, également très rapide, utilisant une cascade de détecteur a été développée par Viola & Jones<sup>[VJ01]</sup>).

Nos développements récents ont consisté à intégrer dans les classifieurs de la hiérarchie un critère qui compte la présence de certaines zones uniformes (non texturées) en plus de la présence de fragments de bords. Ces zones sont sélectionnées pendant l'apprentissage, de la même manière que les fragments de bords significatifs. Ce second critère permet de rejeter des zones hautement texturées, riches en fragments de bords.

Les réponses issues des classifieurs sont ensuite fusionnées pour obtenir une estimation fine des poses des visages détectés. Le nouvel algorithme que nous utilisons regroupe itérativement les deux alarmes les plus similaires à chaque étape, tant que leur distance est inférieure à un seuil, et rejette ensuite les alarmes de poids trop faibles. Cette procédure élimine les fausses alarmes isolées, et produit des estimations moyennes très précises des poses des visages.

Nous avons intégré cet algorithme dans le système IKONA/MAESTRO sous la forme d'une

---

[VJ01] P. VIOLA, M. J. JONES, « Robust Real-time Object Detection », *rapport de recherche n° CRL2001/01*, COMPAQ Cambridge Research Laboratory, 2001.

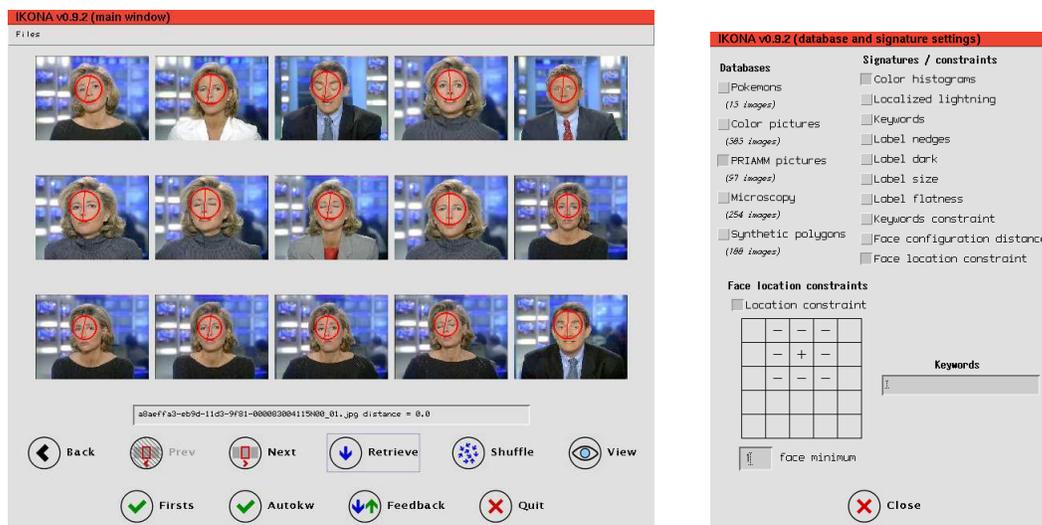


FIG. 8 – Dans cette version de IKONA/MAESTRO, les visages détectés sont indiqués sur les images de la base de données (image de gauche). Une interface (image de droite) permet à l'utilisateur de préciser sur quelles zones il veut retrouver des visages.

interface qui permet à l'utilisateur de spécifier les parties de l'image où il souhaite retrouver des visages (cf. figure 8).

### 6.2.3 Machines à vecteurs de support pour un détecteur hiérarchique des visages

**Mots clés :** Représentation des visages, transformation en ondelettes, apprentissage statistique, machines à vecteurs de supports et détection hiérarchique des visages.

**Participants :** Hichem Sahbi, Nozha Boujemaa, Donald Geman.

**Résumé :** *Nous présentons une approche hiérarchique de détection des visages basée sur les machines à vecteurs de supports (SVMs), permettant une tolérance de l'apparence des visages en termes d'échelle, rotation et position. En allant de la racine vers les cellules feuilles, les détecteurs sont de plus en plus coûteux, dédiés à des poses particulières mais possèdent un taux de fausses alarmes très réduit. Nous introduisons une technique de simplification des classifieurs SVM basée sur le « clustering », la « variation du biais » et la contrainte de « conservation de l'hypothèse visage » pour les fonctions simplifiées, permettant de garantir empiriquement un taux de faux négatifs nul par rapport aux fonctions initiales non simplifiées. Dans notre cas, le processus de détection signifie une recherche en profondeur d'un chemin « racine-feuille » où les détecteurs répondent vrai à « l'hypothèse visage ». Cette approche « Coarse-To-Fine » permet un rejet efficace des images du fond et ainsi une localisation rapide des visages.*

**Modélisation de la pose et description des visages** Soit  $\theta = (p, \phi, s)$  une pose particulière dans un ensemble  $\Lambda$  représentant toutes les poses d'une cellule donnée dans la hiérarchie ( $p, \phi$ , et  $s$  dénotent respectivement la position, la rotation et l'échelle). La construction de cette hiérarchie de détecteurs consiste à apprendre un classifieur SVM  $g_\Lambda$  pour chaque ensemble de poses  $\Lambda$ , avec un coût et une invariance donnée (cf. figure 9). La définition d'une pose  $\theta$  d'un visage est similaire à celle décrite dans [3].

Soient  $l$  observations  $x_i \in R^n$  avec leurs étiquetages associés  $y_i$ . En détection des visages,  $x_i$  sont les  $8 \times 8$  coefficients de basses fréquences de la transformation en ondelette de Daubechies appliquée à chaque imagerie de taille  $64 \times 64$ . Le choix des basses fréquences est motivé par la nature de la tâche à effectuer [14] [13], qui est la détection : nécessité de prendre en compte la forme globale des visages et non pas les détails.

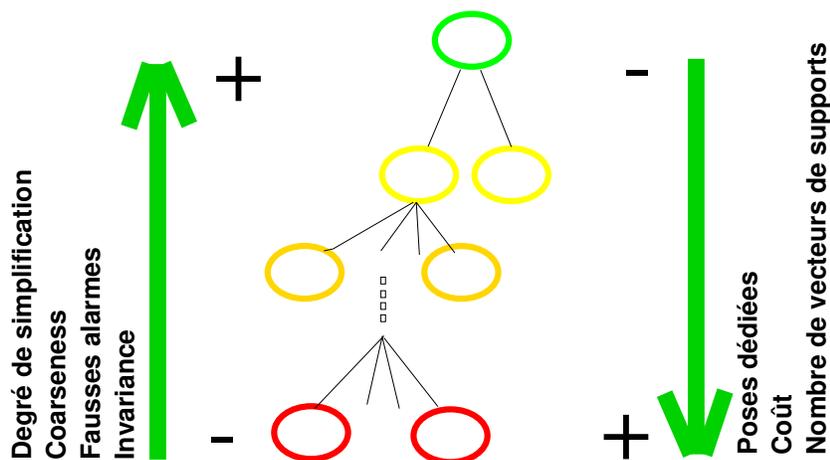


FIG. 9 – Invariance et coût d'évaluation dans le détecteur hiérarchique.

**Un classifieur hiérarchique SVM : notre démarche** On dénote par  $f_\Lambda, g_\Lambda$ , les fonctions de décision SVM dédiées à un ensemble de poses  $\Lambda$  avec un coût d'évaluation différent. Notre approche vise à apprendre pour chaque cellule  $\Lambda$  un classifieur SVM réduit  $g_\Lambda$  avec une complexité dépendante du niveau  $\mathcal{L}(\Lambda)$  dans l'arbre de classification. Ainsi, les fonctions de décision proches de la racine de l'arbre sont plus rapides à évaluer. Plusieurs techniques existent, permettant de réduire la complexité d'un SVM, parmi lesquelles la technique des ensembles réduits [BS97]. Dans notre cas, le processus de réduction est établi par clustering [Dav91a] des vecteurs de support pour chaque fonction  $f_\Lambda$  en fixant une borne supérieure  $C$  caractérisant le nombre maximal de ces vecteurs ( $C$  est le nombre de clusters). Les  $C$  centres de clusters obtenus sont les éléments candidats à devenir vecteurs de supports. Les nouveaux supports sont obtenus

[BS97] C. BURGESS, B. SCHLAKOPF, « Improving the Accuracy and Speed of Support Vector Machines », *Neural Information Processing Systems, Cambridge. MIT Press*, 1997.

[Dav91a] R. N. DAVE, « Characterization and detection of noise in clustering », *Pattern recognition* 12, 11, 1991, p. 657–664.

par maximisation de la marge entre les deux nouvelles populations des vecteurs candidats (les  $C$  centres de clusters). Notons que par construction, les nouvelles fonctions  $g_\Lambda$  garantissent empiriquement la conservation des patterns positifs (c.a.d  $\forall x : f_\Lambda(x) \geq 0 \Rightarrow g_\Lambda(x) \geq 0$ ). L'intérêt majeur de cette hypothèse est la construction des classifieurs  $g_\Lambda$  dans la hiérarchie ayant une capacité de rejet des imageries du fond qui est de plus en plus importante en allant de la racine vers les feuilles avec une conservation des bons patterns (c'est à dire les régions les plus sujette à être classées visage).



FIG. 10 – Quelques résultats de détection de visages par l'approche « coarse-to-fine ».

### Recherche des visages :

Pour décider si un pattern  $x$  est visage, on recherche un chemin de la racine vers une feuille le long duquel tous les classifieurs  $f$  dédiés aux ensembles de poses  $\{\Lambda_1, \dots, \Lambda_m\}$  répondent vrai à l'hypothèse visage (i.e.  $f_{\Lambda_1}(x, \alpha) \geq 0, \dots, f_{\Lambda_m}(x, \alpha) \geq 0$ ). En considérant l'hypothèse de conservation, cette suite de tests peut être accélérée en considérant les fonction  $g$  (i.e.  $g_{\Lambda_1}(x, \gamma) \geq 0, \dots, g_{\Lambda_m}(x, \gamma) \geq 0$ ). En outre, une réponse négative d'une fonction  $g_{\Lambda_i}$  pour une cellule particulière signifie forcément que  $f_{\Lambda_i}$  est négative (par contra-position de l'hypothèse de conservation), et toute la hiérarchie qui découle de  $\Lambda_i$  peut être élaguée. Le nombre d'élagages importants, liés à la fraction importantes des imageries du fond, rend le coût global de notre algorithme très réduit puisque l'on passe d'une durée de 120s (en utilisant une hiérarchie de classifieurs  $f$  : réseau  $f$ ) à 8s (en utilisant un réseau  $g$ ) et ceci sur des images de taille  $350 \times 450$  pixels. En terme de taux de détections, notre algorithme atteint une performance de 88% de détections sur la base de test CMU, avec un taux de fausses alarmes de 710 sur 1561198 patterns classifiés, et ceci correspond à 2 ou 3 fausses alarmes par image (cf. figure 10).

### 6.3 Requêtes partielles

Lorsque la recherche visuelle porte sur une zone de l'image, les requêtes globales sont inefficaces car les similarités sont mesurées sur l'image entière. Les requêtes partielles sont l'alternative qui consiste à décrire et retrouver les « parties visuellement saillantes » de l'image. Deux approches locales et **complémentaires** sont développées à IMEDIA. Elles diffèrent par la définition de ces parties saillantes :

- Requêtes partielles par régions d'intérêt (zones visuellement homogènes, par exemple champ de lavande, peau, etc.)
- Requêtes partielles par points caractéristiques couleur, présents dans les zones de l'image à forte variation locale (par exemples vêtements à motifs, objets particuliers présentés sous différents points de vue, etc.). Cette approche est particulièrement utile quand la segmentation des images est difficile pour des raisons de complexité du contenu et/ou de qualité de prise de vue des images

#### 6.3.1 Requêtes par points d'intérêt

**Mots clés :** Requêtes partielles, Images couleur, Descripteurs locaux, Points d'intérêt, Invariants différentiels.

**Participants :** Valérie Gouet, Nozha Boujemaa.

**Résumé :** *Lorsque la recherche est envisagée sur une partie de l'image définie par l'utilisateur, les descriptions globales sont inadéquates, il est alors nécessaire de considérer une description locale de l'image. Ce type de description passe par l'extraction de primitives de l'image. Une solution consiste à extraire de celle-ci un ensemble de points d'intérêt, exhibant ainsi les zones de l'image les plus riches en information photométrique. Chaque point est ensuite décrit par des attributs photométriques et géométriques. L'ensemble de ces attributs fournit une caractérisation locale de l'image, robuste aux occultations et qui peut être rendue robuste à diverses transformations telles que les changements d'illumination et de point de vue selon les descripteurs de points envisagés.*

*Nous avons proposé une description compacte des points basée sur les invariants différentiels de Hilbert généralisés à la couleur, dans le but de faire de la recherche de parties d'images ou d'objets contenus dans les images.*

*Ces travaux sont réalisés dans le cadre d'un contrat avec la Police Judiciaire pour le programme européen « STOP » concernant la protection des personnes.*

Le choix des points d'intérêt comme descripteur local de l'image repose sur plusieurs considérations : c'est d'abord la primitive qui est la moins sensible au contenu géométrique de l'image. Ensuite, la description étant locale, les points sont robustes aux transformations de l'image telles que les changements de point de vue, mais aussi les occultations. Enfin, les techniques d'indexation et de recherche d'images par le contenu exploitent l'information photométrique contenue dans l'image. Par définition, un point d'intérêt est situé là où l'information photométrique est la plus significative. En les utilisant, nous sommes donc en droit d'attendre une caractérisation riche et compacte à la fois.

**Choix de l'extracteur de points :** Lorsqu'il s'agit de faire de l'indexation et de la recherche d'images, les points d'intérêt doivent posséder une excellente *répétabilité*. C'est à dire qu'ils doivent pouvoir être extraits avec la même précision d'une image à l'autre. Plusieurs extracteurs de points existent en niveau de gris, mais c'est le détecteur de Harris dans sa version couleur qui possède la meilleure répétabilité. Ce sont donc les points Harris couleur qui nous servent de base pour l'indexation des images.

**Caractérisation couleur à partir des invariants de Hilbert :** La méthode qui a fait ses preuves en niveau de gris pour caractériser un point est basée sur les invariants différentiels de Hilbert jusqu'au troisième ordre, et plus particulièrement sur une combinaison de ces grandeurs pour assurer l'invariance en translation et en rotation ainsi qu'une certaine stabilité numérique. Une généralisation à la couleur a été proposée pour des applications de stéréovision [GMP98]. Elle implique les invariants du premier ordre seulement ainsi que deux invariants spécifiques à la couleur. On obtient au point  $\vec{x}$  une caractérisation fonction de sa couleur RVB et des gradients sur chaque plan :

$$\vec{v}_{col}(\vec{x}) = \left( R \quad \|\nabla R\|^2 \quad V \quad \|\nabla V\|^2 \quad B \quad \|\nabla B\|^2 \quad \nabla R \cdot \nabla V \quad \nabla R \cdot \nabla B \right)^T \quad (3)$$

Nous avons choisi d'appliquer cette approche de caractérisation couleur locale pour l'indexation et la recherche d'images, car elle possède deux avantages majeurs :

- Elle minimise de façon significative le principal inconvénient des invariants en niveau de gris, à savoir leur sensibilité au bruit lié à l'utilisation des dérivées du second et du troisième ordre. Cette remarque est très pertinente en recherche d'images où la qualité des images collectées n'est pas toujours maîtrisée ;
- C'est la représentation la plus compacte, puisque toute l'information photométrique couleur du voisinage local du point est résumée dans les 8 composantes de  $\vec{v}_{col}$ . La description gagne en qualité, sans perdre en termes de coût de stockage et de recherche (par rapport aux 9 invariants en niveau de gris).

**Algorithme d'indexation et de recherche :** Pour l'image de la requête comme pour les images de la base, la description obtenue est un ensemble de points décrits dans un espace à 8 dimensions. Dans ce contexte, interroger la base d'images revient à rechercher dans cet espace les points les plus proches des points de la requête, comme l'illustre le schéma de la figure 11. Le résultat de la recherche est alors le sous-ensemble d'images qui comportent le plus grand nombre de points retournés.

Un tel processus de recherche exploite uniquement l'information photométrique contenue dans les images. La recherche peut être raffinée en tenant compte de contraintes géométriques liant les points de l'image requête, comme par exemple leur répartition spatiale dans l'image.

Une évaluation automatique de cette approche a été réalisée dans [12] sur des parties d'images ainsi que sur des objets 3D, pour comparaison avec les approches de description globale. La méthode a été implémentée au sein de la plateforme IKONA. Un exemple d'utilisation

---

[GMP98] V. GOUET, P. MONTESINOS, D. PELÉ, « A Fast Matching Method for Color Uncalibrated Images using Differential Invariants », *in : Proceedings of the British Machine Vision Conference, I*, p. 367-376, Southampton, UK, septembre 1998.

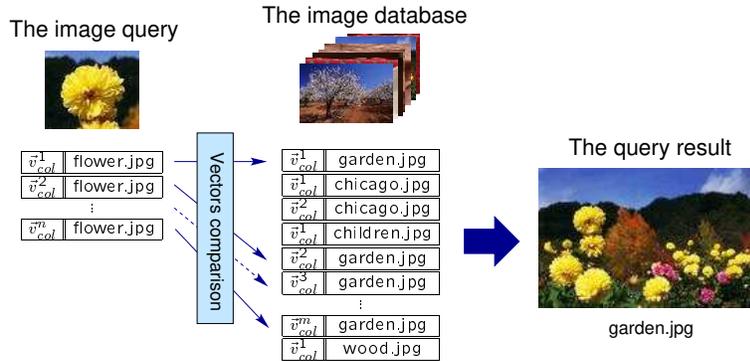


FIG. 11 – Indexation et recherche d'images à partir de points d'intérêt.

est illustré par la figure 12. Ici l'utilisateur a défini manuellement une zone rectangulaire représentant la requête et contenant l'objet qu'il souhaite rechercher, à savoir un tournesol. Les 9 premières images les plus similaires sont présentées.

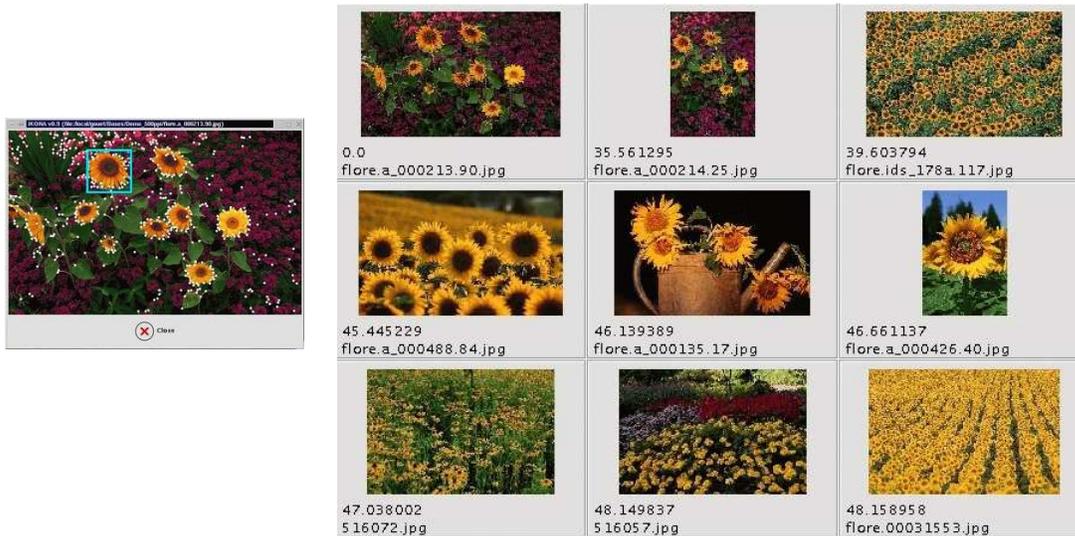


FIG. 12 – Un exemple de recherche de partie d'image à partir des points d'intérêt couleur.

Comme en niveau de gris, il est possible pour la couleur d'employer les invariants de Hilbert à des ordres supérieurs [Gro00]. Nous étudions actuellement la pertinence de ces invariants supplémentaires dans différents contextes susceptibles d'être rencontrés en recherche d'images, comme le changement de point de vue, le bruit de codage (JPEG dans notre étude) ou encore les changements d'illumination. En parallèle, une recherche est menée pour optimiser les temps de recherche dans la base de points, par l'utilisation d'index multi-dimensionnels afin de permettre à ce type de signature de rester efficace sur de grandes bases d'images.

[Gro00] P. GROS, « Color Illumination Models for Image Matching and Indexing », in : *Proceedings of 15<sup>th</sup> International Conference on Pattern Recognition*, p. 580–583, Barcelona, Spain, septembre 2000.

### 6.3.2 Requêtes partielles par régions d'intérêt

**Mots clés :** Requêtes partielles, régions d'intérêt, segmentation, classification, distribution locale de couleur, nuances fines de couleur.

**Participants :** Julien Fauqueur, Nozha Boujemaa.

**Résumé :** *Les requêtes partielles par régions d'intérêt permettent à l'utilisateur de sélectionner une région exemple dans une image et de retrouver toutes les images de la base comportant une région similaire, indépendamment du reste de l'image.*

*Une nouvelle technique de segmentation a été développée afin d'extraire automatiquement des régions grossières, discriminantes pour la recherche par similarité visuelle. Nous présentons aussi un descripteur couleur fidèle à l'apparence visuelle dominante des régions.*

**Détection de régions grossières :** Une nouvelle technique de segmentation a été développée pour la détection automatique de régions dans des bases d'images génériques. Elle repose sur la classification non supervisée de distributions locales de couleurs et la prise en compte d'informations topologiques globales. En comparaison à la segmentation couleur utilisée précédemment, cette segmentation est mieux adaptée aux requêtes partielles. En effet, les régions détectées présentent une « diversité homogène visuelle » qui les rend plus caractéristiques les unes vis-à-vis des autres dans la base. La figure 13 illustre des exemples de régions détectées.

**Description adaptative des régions :** Les descripteurs couleur existant pour les images ou les régions [SC96], [CTB99], [MM99] sont basés sur une quantification grossière de l'espace couleur en quelques centaines de couleurs. Cette approche convient aux images, mais les régions requièrent une représentation plus fine, car, par définition, elles sont plus homogènes visuellement.

La description d'une région se base sur une moyenne de 17 couleurs extraites adaptativement, pour chaque région, parmi les 5 millions de l'espace couleur. Ce nouveau descripteur caractérise la Distribution Adaptative des Nuances Fines de Couleurs (ou *DANFC*). L'index résultant est plus compact et caractérise l'apparence visuelle des régions de façon beaucoup plus fine que l'histogramme couleur classique. Des exemples des couleurs utilisées pour la description spécifique à chaque région sont présentés dans la figure 13.

**Tests et résultats :** Le mode de requêtes partielles a été intégré cette année dans IKONA/MAESTRO.

Son déroulement est le suivant : dans une première étape « off-line », chaque image de la base est automatiquement segmentée en régions d'intérêt qui sont individuellement indexées ;

- 
- [SC96] J. SMITH, S. CHANG, « VisualSEEk: A Fully Automated Content-Based Image Query System », *in: ACM Multimedia*, p. 87-98, 1996.
- [CTB99] C. CARSON, M. THOMAS, S. BELONGIE, « Blobworld: A system for region-based image indexing and retrieval », 1999.
- [MM99] W.-Y. MA, B. S. MANJUNATH, « NeTra: A Toolbox for Navigating Large Image Databases », *Multimedia Systems* 7, 3, 1999, p. 184-198.



FIG. 13 – Illustration du principe de détection grossière et description fine de régions extraites de 2 images : image originale, image des régions détectées puis image reconstituée à partir des couleurs de chaque région utilisées pour leur description. Les petites régions ont été supprimées.

la recherche par région peut alors s'effectuer : l'utilisateur sélectionne une région requête dans l'interface IKONA, et MAESTRO détermine et retourne les images comportant des régions similaires. IKONA surligne les régions similaires dans les images retournées.

IKONA présente les images de la base sous forme de vignettes dont les régions extraites sont cliquables. Les régions retournées sont cerclées de blanc dans leur image d'origine. Les travaux ont été testés sur une base générique d'images.

Même dans les scènes naturelles complexes, les régions détectées présentent une « diversité homogène » visuelle cohérente. L'information contenue dans la description des régions capture les nuances de couleur de celles-ci et améliore la précision des résultats par rapport à l'histogramme classique (voir les exemples de requêtes région dans la figure 14). Nous envisageons l'intégration d'information de forme dans la description des régions et le développement d'un mode de requêtes par régions multiples.

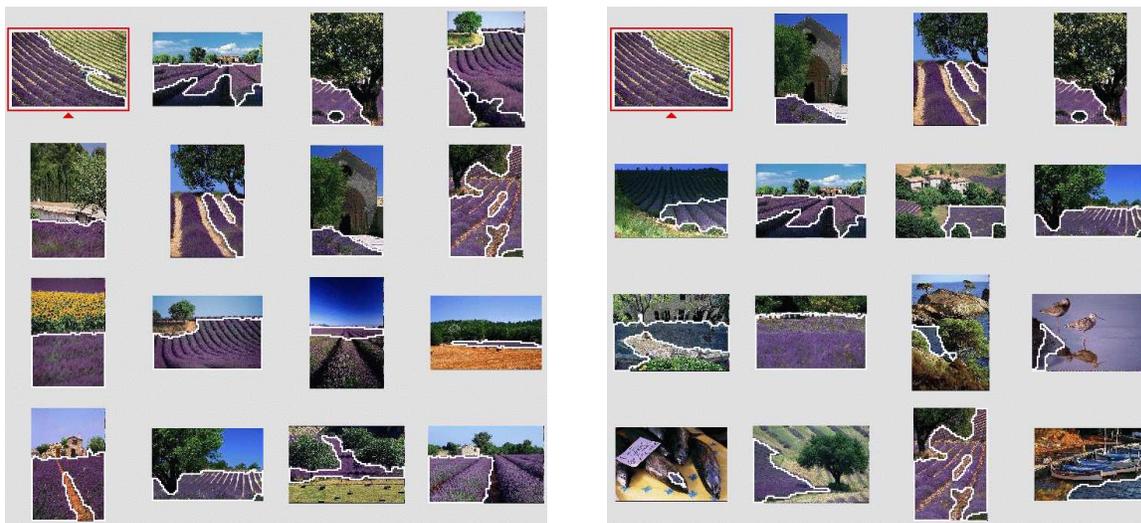


FIG. 14 – Requête à partir de la région de lavande dans l'image encadrée de rouge. Images gauche : résultats avec le nouveau descripteur *DANFC*. Images droite : résultats avec l'histogramme couleur classique Luv quantifié en 216 bins.

## 6.4 Recherche interactive

### 6.4.1 Méthodes de recherche d'images par boucle de pertinence

**Mots clés :** recherche image par contenu visuel, boucle de pertinence, apprentissage statistique, signatures visuelles.

**Participants :** Marin Ferecatu, Nozha Boujemaa.

**Résumé :** *Les méthodes de recherche de type boucle de pertinence sont mieux adaptées aux utilisateurs qui cherchent des classes d'images dans les grandes bases d'images et qui ne savent pas à l'avance la structure exacte de leur requête, mais qui ont une idée relativement précise sur ce qu'ils cherchent.*

La mise au point de signatures visuelles performantes, implantées efficacement, est un préalable à la recherche sur les modes d'interactions de type « boucle de pertinence ».

De par sa conception, notre logiciel IKONA a une structure modulaire et peut être facilement enrichi avec de nouveaux algorithmes et méthodes de recherche.

IKONA fonctionne principalement en deux étapes : l'indexation des images (l'extraction des informations visuelles et l'organisation de ces informations pour être rapidement accessibles) et la boucle d'interaction avec l'utilisateur (l'utilisateur spécifie — ou affine — la requête et le système lui présente les résultats).

La principale méthode d'interaction est la recherche directe par contenu visuel. Chaque image est représentée comme un ensemble de vecteurs (des caractéristiques ou signatures) et une fonction distance est utilisée pour faire la recherche.

A l'heure actuelle, IKONA contient plusieurs **signatures** de qualité qui couvrent les trois principaux aspects de la description des images : couleur, texture et forme.

Une implantation efficace d'un ensemble de signatures a été réalisé et évalué pour choisir la meilleure combinaison, avant d'effectuer le contrôle de pertinence.

Les signatures ont été choisies pour avoir de bonnes performances sans trop augmenter la dimension de l'espace de représentation résultant. Dans la figure 15 on peut voir un diagramme précision/rappel qui compare les performances des signatures choisies (histogramme pondéré, histogramme d'orientation, et signature de fourier) par rapport à leur superposition (**standard**).

**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).

A partir des spécifications dynamiques de l'utilisateur, IKONA essaie de trouver un modèle statistique pour décrire de manière optimale la classe d'images.

Nous avons implanté et évalué deux approches : une méthode d'estimation de type fenêtre de Parzen pour approcher la densité de probabilité sur l'espace de toutes les images (probabilité traduisant la relevance ou non-relevance de l'image) <sup>[MN99]</sup>, et une estimation paramétrique de

---

[MN99] C. MEILHAC, C. NASTAR, « Relevance Feedback and Category Search in Image Databases », *in* : *Proc. of IEEE ICMCS*, Florence, Italy, 1999.

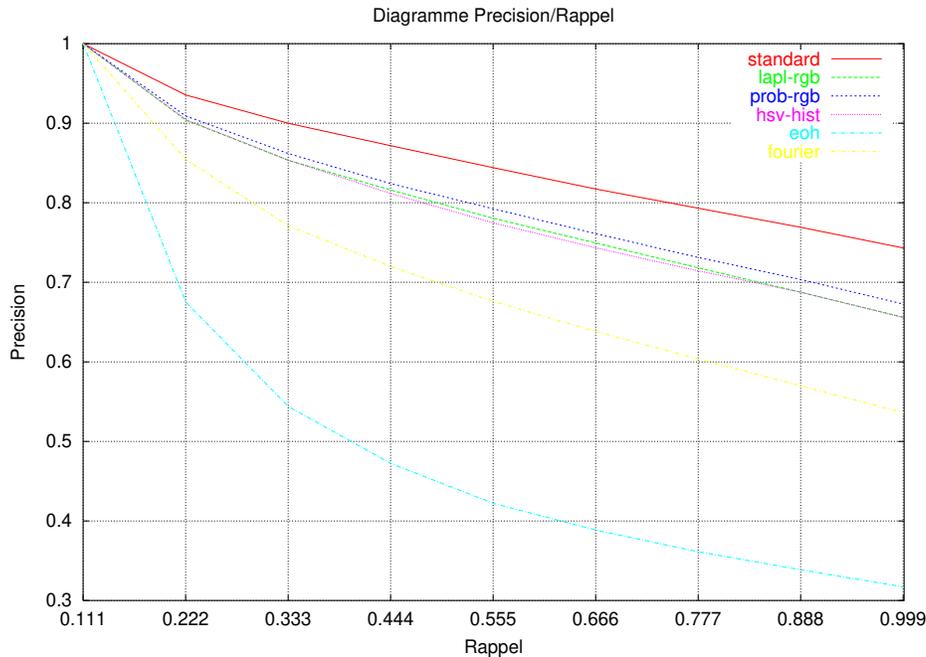


FIG. 15 – Diagramme précision/rappel pour le choix standard des signatures en IKONA

type Gaussienne. Cette dernière suppose que les composantes sont indépendantes et adapte de manière itérative les paramètres de la Gaussienne de telle sorte que la distribution approche de façon satisfaisante les exemples positifs et évite le mieux possible les exemples négatifs [NMM98].

L'intérêt de mettre au point de nouvelles méthodes est d'autant plus fort lorsque l'on s'intéresse à l'intégration de deux médias dans cette interaction : le texte et l'image. Ce type d'intégration est le sujet de thèse de Marin Ferecatu (thèse commencée le 1/10/2001 au sein de l'équipe IMEDIA). Une partie de ces premiers résultats a été publiée ([2, 10, 16]).

#### 6.4.2 Résumé de bases par ses images les plus informatives

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

**Participants :** Bertrand Le Saux, Nozha Boujemaa.

**Résumé :** *Une des phases majeures de l'aide à la navigation intelligente est de fournir un résumé le plus représentatif de la base d'images à l'utilisateur. Il s'agit de catégoriser la base puis de choisir le meilleur représentant par catégorie pour construire un résumé qui sera présenté à l'utilisateur.*

[NMM98] C. NASTAR, M. MITSCHKE, C. MEILHAC, « Efficient query refinement for image retrieval », in : *Proc. of CVPR*, Santa Barbara, USA, 1998.

Une image est représentée par un index (vecteur) dans un espace de signatures. Des signatures complémentaires sont utilisées pour rendre compte le mieux possible des différences entre les images : couleur, texture, forme. Trois problèmes rendent la catégorisation difficile : l'espace des signatures est de très grande dimension, le nombre de catégories est a priori inconnu, et les classes naturelles présentent des formes diverses et variées avec un recouvrement important.

Une méthode originale de catégorisation a été mise au point : ARC (pour Adapative Robust Agglomeration). Tout d'abord, l'espace des signatures est réduit par une analyse en composantes principales afin de ne garder que les principales composantes de chaque signature. Puis la base est catégorisée en minimisant un critère d'agglomération compétitive (CA) qui permet de trouver le nombre de classes optimal. Nous avons considéré comme point de départ la fonction objective suivante <sup>[FK97]</sup> :

$$J = \sum_{k=1}^C \sum_{i=1}^N (u_{ki})^2 d^2(x_i, \beta_k) - \alpha \sum_{k=1}^C \left[ \sum_{i=1}^N (u_{ki}) \right]^2 \quad (4)$$

Les catégories d'images sont plus ou moins denses et plus ou moins étalées, et la fonction (4) ne permet pas de retrouver les classes peu denses car elles sont éliminées lors de la compétition. Nous avons mis au point un mécanisme de compétition adaptative : le terme qui tend à réduire le nombre de classes est pondéré par un facteur dépendant de la densité particulière de chacune des classes. Cette méthode est soumise à publication.

Nous introduisons, également, une classe bruit <sup>[Dav91b]</sup> pour collecter les points ambigus ou aberrants, définie comme à égale distance de toutes les autres classes :

$$d^2(x_i, \beta_1) = \delta^2 \quad (5)$$

L'introduction de cette classe permet une meilleure définition des classes d'objets. Les différentes catégories d'images ont des formes variées, ce qui nous amène à utiliser la distance de Mahalanobis pour distinguer les autres classes :

$$d_f(x_i, \beta_k) = |C_{k,f}|^{1/p_f} (x_{i,f} - \beta_{k,f})^T C_{k,f}^{-1} (x_{i,f} - \beta_{k,f}) \quad (6)$$

Les performances de notre algorithme ont été comparées avec l'algorithme SOON<sup>[FBL01]</sup>[11] et avec l'algorithme de compétition de base, non-adaptatif (4). Pour chaque catégorie, un représentant est choisi suivant les étapes suivantes : la valeur moyenne de chaque signature est calculée pour chaque image ; la moyenne de toutes les images définit un prototype virtuel ; l'image-prototype de la catégorie est l'image la plus proche du prototype virtuel.

Les deux algorithmes sont testés sur une base avec vérité terrain, la base Columbia <http://www.cs.columbia.edu/CAVE/research/softlib/coil-20.html> (objets photographiés sous

- 
- [FK97] H. FRIGUI, R. KRISHNAPURAM, « Clustering by competitive agglomeration », *Pattern Recognition* 30, 1997, p. 1109-1119.
- [Dav91b] R. N. DAVE, « Characterization and detection of noise in clustering », *Pattern Recognition Letters* 12, 1991, p. 657-664.
- [FBL01] H. FRIGUI, N. BOUJEMAA, S.-A. LIM, « Unsupervised Clustering and Feature Discrimination with Application to Image Database Categorization », in: *NAFIPS*, Vancouver, Canada, 2001.



FIG. 16 – Prototypes des classes obtenus avec notre algorithme ARC.



FIG. 17 – Prototypes des classes obtenus avec SOON.

différents angles de prise de vue) qui comprend 20 classes avec 72 représentants par classe. Les résumés de la base sont donnés sur les figures 16, 17 et 18. Avec les méthodes SOON et CA, les catégories naturelles les plus dispersées sont séparées en plusieurs classes, donc certains prototypes sont redondants. La méthode ARC offre un meilleur résumé en terme de nombre et de qualité de construction des classes. De plus, comme l'algorithme CA n'a pas de cluster bruit, les résultats sont très bruités. La figure (21) présente une des classes obtenues avec cette méthode, un tiers des images est mal classifié. Les deux autres méthodes permettent d'obtenir des classes plus homogènes. Cependant, avec la méthode SOON, l'inconvénient est qu'un grand nombre d'images sont considérées comme ambiguës et rejetées dans le cluster bruit. La figure (20) montre que, pour le même objet que précédemment, seulement un cinquième des images est retrouvé. Notre méthode ARC ne met que les images réellement ambiguës dans le cluster bruit, ce qui permet de retrouver la quasi totalité des images correspondant à un même objet (ici 71 sur 72) avec un petit nombre d'images mal catégorisées, comme le montre la figure (19).

## 7 Contrats industriels (nationaux, européens et internationaux)

### 7.1 Contrat Thales-Identification

**Participants :** Sylvain Bernard, Nozha Boujemaa.

Ce contrat est un contrat d'accompagnement de la thèse CIFRE de Sylvain Bernard. Il a démarré en octobre 1998 et continue jusqu'en octobre 2001. La thèse a été soutenue le 9



FIG. 18 – Prototypes des classes obtenus avec CA non adaptatif.

novembre 2001.

## 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, Julien Fauqueur, Jean-Paul Chièze.

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.

#### 8.1.3 9ème Rencontres INRIA-Industrie

Cette rencontre a porté sur le thème « Recherche d'informations dans les documents électroniques ». Le but étant d'identifier des partenariats pour faire émerger une action de développement. IMEDIA a participé avec deux présentations orales et des démonstrations. Pour en savoir plus : <http://www.inria.fr/valorisation/Web-semantic/index.fr.html>



FIG. 19 – cluster de l'objet 'emballage' obtenu avec ARC



FIG. 20 – cluster de l'objet 'emballage' obtenu avec SOON



FIG. 21 – cluster de l'objet 'emballage' obtenu avec CA

## 8.2 Actions européennes

### 8.2.1 Projet Européen « STOP »

**Participants :** Valérie Gouet, Nozha Boujemaa.

Il s'agit d'un contrat avec la Police Judiciaire dans le cadre du programme européen « STOP » pour la protection des personnes. Il concerne particulièrement les problèmes de pédophilie. La police espagnole est également impliquée. La start-up LTU-Technologies, issue d'IMEDIA, est partenaire de contrat européen. Le budget est de 300kf pour IMEDIA.

## 8.3 Actions internationales

### 8.3.1 Projet STIC INRIA « Indexation et recherche d'images par le contenu : application aux images satellitaires »

**Participant :** Nozha Boujemaa.

C'est un projet de collaboration qui implique deux sites de l'INRIA : Rocquencourt (IMEDIA), Rennes (VISTA) avec l'université de Tunis II. Il finance sur un an, deux stages avec plusieurs séjours de courte durée dans les deux sens. Ainsi, Ronan Fablet, Patrick Gros (tous deux de VISTA) et Nozha Boujemaa ont participé aux enseignements « Multimédia » du DEA STIC de Sup'Com (école d'ingénieurs en informatique et télécommunications). Ziad Belhadj

ainsi que Amel Benazza, enseignants-chercheurs du département « signal et communication » ont effectué des visites à IMEDIA. Nozha Boujemaa est responsable, du côté français, de ce projet.

### 8.3.2 Consortium AIR-D

IMEDIA participe au consortium AIR&D, récemment lancé par l'INRIA avec Philips et Thomson-Multimédia, autour des nouvelles technologies de l'**intelligence ambiante**.

## 8.4 Visites, et invitations de chercheurs

Hichem Frigui (University of Memphis, États-Unis), Lambert Shomacker (NICI, Pays-Bas) et Christophe Garcia (université de Crète à Heraklion) nous ont visités pour de courtes durées, et ont animé des séminaires à IMEDIA.

## 9 Diffusion de résultats

### 9.1 Animation de la Communauté scientifique

Nozha Boujemaa :

- Membre d'un « Working Group » NSF/Delos(ERCIM) portant sur l'indexation et la recherche par le contenu avec 5 participants US et 5 participants européens. Le « kick-off meeting » sera en janvier 2002. Les participants européens sont : Alberto Del Bimbo et Giuseppe Amato (Italie), Arnold Smeulders (Pays-Bas), Ioannis Pitas (Grèce), Joseph Kittler (Grande-Bretagne) Patrick Bouthemy et Nozha Boujemaa (France) ;
- Membre d'un « Working Group » ERCIM dont le thème est « Image Understanding » et le « kick-off meeting » est en février 2002 ;
- Membres de cinq jurys de thèse en 2001 : Wolfgang Muller (Computer Vision Group - Université de Geneve, Suisse), Mohamed Borchani (Laboratoire SIP - Université de Paris 5), Jamal Malki (L3i - Université de La Rochelle), Sylvain Bernard (Laboratoire Prism - Université de Versailles), Gilles Roux (LI - Université de Tours) ;
- Program Chair de MMCBIR'2001 Les 24 et 25 septembre 2001 s'est tenu à l'INRIA-Rocquencourt le workshop « Indexation et Recherche par le contenu dans les documents multimédias » (Multimedia Content-based Indexing and Retrieval) MMCBIR 2001. Ce workshop a été co-organisé par l'INRIA, le centre IBM Research T.J. Watson et l'Université de Berkeley, avec le soutien financier de la NSF « Digital Libraries Initiative - Phase 2 » (DLI2) ;

Ce workshop a réuni, sur invitation, des scientifiques de renommé du monde européen et américain de la recherche académique et industrielle. Nozha Boujemaa a été chargé de la coordination du coté européen.

Pour plus d'informations : <http://www.inria.fr/actualites/colloques/2001/mmcbir/index.en.html> [6]

- Membre de comité de programme des conférences : IR'2001 (septembre - Finland), MM-DE'2001 (juillet - France), Media'Net 2002 (juin - Co-organisé par l'INRIA en Tunisie), KDMCD 2002 (Mai - Taiwan) ;

- Membre du comité d'organisation de ACM Multimedia 2002 (Décembre - France), chargée de la coordination des workshops ;
- Présidente de la session : « Content-based Querying » au IEEE Workshop on Content-based Access of Image and Video Libraries (CBAIVL - Hawaii, decembre 2001) [9] ;
- Coordination scientifique, avec Michel Mauny, d'un atelier Franco-Marocain dont l'INRIA (via les relations internationales - Faranak Grange) avait la charge d'organiser. Il porte sur le double thème « Génie logiciel et Multimédia » à Rabat : [http://www-direction.inria.fr/international/AFRIQUE\\_Proche\\_Orient/reseau/AtelierMaroc.htm](http://www-direction.inria.fr/international/AFRIQUE_Proche_Orient/reseau/AtelierMaroc.htm) ;
- Membre du Conseil d'administration de l'AFRIF (Association Française de Reconnaissance et Interprétation des Formes) ;
- Membre du « noyau dur » du groupe thématique du GDR PRC ISIS sur la couleur.

Activité d'expertise scientifique de Nozha Boujemaa :

- Expertise de projets soumis au financement de la NSF dans le thème « Information Technology Research » (ITR proposals) plus particulièrement centré sur « Information Management ». Il s'agit de projets pouvant atteindre une durée de 5 ans pour un budget annuel atteignant 1 million de dollars. Ce panel s'est déroulé à Los Angeles en Février 2001 ;
- Expertise de 2 projets pour l'appel ToKeN aux Pays-Bas pour le compte du NW0 (l'équivalent du CNRS) ;
- Audite technologique pour la Société Générale Asset Management (SGAM) suite à une demande d'investissement d'une start-up dans le secteur de l'indexation et la recherche par le contenu ;
- Expertise de projets soumis au financement du réseau national RIAMM (RECHERCHE ET INNOVATION EN AUDIOVISUEL ET MULTIMEDIA) (CNC/Ministère de l'Industrie) ;
- Expertise d'articles scientifiques pour les revues : IEEE PAMI, CVIU, IVC, IEEE SMC, etc. ainsi que pour plusieurs conférences.

Séminaires invités de Nozha Boujemaa :

- Invitation pour l'animation d'une table ronde autour du thème « Major Technical and Business Challenges for Content-based Retrieval » qui s'est déroulée au « IEEE Workshop on Content-based Access of Image and Video Libraries » (CBAIVL-2001) en conjonction avec IEEE CVPR-2001 à Hawaii. Le panel était également composé des personnes suivantes : Shih-Fu Chang et John Kender (Columbia Univ.), Thomas S. Huang (University of Illinois, Urbana-Champaign), Yong Rui (Microsoft), Linda Shapiro (Univ. of Washington) ;
- Séminaire invité au JRC (laboratoire de recherche de la CEE) dans sa composante « Reliable Information Technologies Unit ». Une visite du laboratoire avec une discussion sur l'apport de l'indexation d'images aux questions de sécurité en générale et aux problèmes de surveillance de centrale nucléaires en particulier. Marin Ferecatu a accompagné Nozha Boujemaa lors de cette visite au cours de laquelle IKONA a été installé au JRC pour une phase d'évaluation ;
- Présentation orale invitée [7], en février à Berlin, pour la conference Suchbilder (rechercher des images) organisée par une communauté cinématographique/artistique al-

lemande (producteur de cinéma, institut d'art contemporain avec le département des études culturelles et artistiques de l'université Humboldt à Berlin. Les aspects culturels, technologiques et usages dans le cinéma ont été abordés. Julien fauqueur a pris en charge les démonstrations d'IMEDIA. Pour en savoir plus : <http://www.suchbilder.de/index-e.html> ;

- Invitation pour une présentation orale au séminaire conjointement organisé par l'IRISA-thèque et le CNRT TIM-Bretagne sur l'indexation multimédia : <http://tim.irisa.fr/veille/indexation/>.

## 9.2 Enseignement

**Nozha Boujemaa** : Chargé de cours à l'UTC dans le cadre de la filière « Ingénierie des Industries Culturelles » -

**Donald Geman** : Chargé de cours au DEA de l'ENS Cachan.

**François Fleuret** : Cours de C++ undergraduate à l'université de Chicago, cours de vision en graduate.

**Valérie Gouet** : Cours de Synthèse d'images et de Vision 3D à l'option informatique de l'Ecole des Mines d'Alès.

**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.

**Hichem Sahbi** : Chargé de TD introduction à la programmation orienté objet appliqué à Java, pour les étudiants en DEUG MIAS à l'université Versailles St Quentin en Yvelines.

## 10 Bibliographie

### Thèses et habilitations à diriger des recherches

- [1] S. BERNARD, *Indexation et recherche dans de grandes bases d'empreintes digitales*, thèse de doctorat, Université de Versailles Saint-Quentin, 2001.

### Articles et chapitres de livre

- [2] N. BOUJEMAA, M. FERECATU, « Mining multimedia documents by visual content », *Numéro special Ambient Intelligence de ERCIM News*, octobre 2001.
- [3] F. FLEURET, D. GEMAN, « Coarse-to-fine visual selection », *International Journal of Computer Vision*, 2001, à paraître.

### Communications à des congrès, colloques, etc.

- [4] S. BERNARD, N. BOUJEMAA, D. VITALE, C. BRICOT, « Fingerprint Classification Using Kohonen Topologic Map », in : *International Conference on Image Processing ICIP'2001*, Thessaloniki, Greece, Oct. 2001.

- 
- [5] S. BERNARD, N. BOUJEMAA, D. VITALE, C. BRICOT, « Fingerprint Segmentation using the Phase of Gabor Wavelets », *in : à paraître dans "Proc. of the 5th Asian Conference on Computer Vision ACCV'2002"*, Melbourne, Australia, Jan. 2002.
  - [6] N. BOUJEMAA, J. FAUQUEUR, M. FERECATU, F. FLEURET, V. GOUET, B. LE SAUX, H. SAHBI, « Interactive Specific and Generic Image Retrieval », *in : MNCBIR'2001*, Rocquencourt, France, Sept. 2001.
  - [7] N. BOUJEMAA, J. FAUQUEUR, « Images Indexing, Retrieval and Browsing », *in : SuchBilder'2001*, Berlin, Allemagne, Fev. 2001.
  - [8] N. BOUJEMAA, C. VERTAN, S. BOUGHORBEL, « Color Soft Signature for Image Retrieval by Content », *in : Proc. of EUSFLAT 2001*, p. 394-401, Leicester, Great Britain, 2001. communication invitée.
  - [9] N. BOUJEMAA, « MEDIA point-of-view on Major Technical and Business Challenges for Content-based Retrieval », *in : Proc. of CBAIVL 2001*, Hawaii, USA, Dec. 2001. panel.
  - [10] M. FERECATU, N. BOUJEMAA, F. FLEURET, « IKONA—A Modular Architecture for Content Based Image Retrieval Systems », *in : Proc. of SCI2001, vol. XIII*, Orlando, FL, USA, 2001.
  - [11] H. FRIGUI, N. BOUJEMAA, S.-A. LIM, « Unsupervised Clustering and Feature Discrimination with Application to Image Database Categorization », *in : NAFIPS*, Vancouver, Canada, 2001.
  - [12] V. GOUET, N. BOUJEMAA, « Object-based queries using color points of interest », *in : Proc. of CBAIVL 2001*, Hawaii, USA, 2001.
  - [13] H. SAHBI, N. BOUJEMAA, « Accurate Face Detection Based On Coarse Segmentation and Fine Skin Color Adaption », *in : Proc. of ICISP 2001*, Agadir, Morocco, 2001.
  - [14] H. SAHBI, N. BOUJEMAA, « Robust Matching By Dynamic Space Warping For Accurate Face Recognition », *in : International Conference on Image Processing ICIP'2001*, Thessaloniki, Greece, Oct. 2001.
  - [15] C. VERTAN, N. BOUJEMAA, « Integrated Color Texture Signature for Image Retrieval », *in : Proc. of ICISP 2001*, Agadir, Morocco, 2001.
  - [16] C. VERTAN, M. CIUC, N. BOUJEMAA, M. FERECATU, « Local Extremeness Weighting of Color Distributions for Image Retrieval », *in : Proc. of SCI2001, vol. XIII*, Orlando, FL, USA, 2001.