

Avant-projet IMEDIA

Images et Multimédia: Indexation, Recherche et Navigation

Rocquencourt

THÈME 3B

R *apport*
d'Activité

1999

Table des matières

1	Composition de l'équipe	3
2	Présentation et objectifs généraux	4
3	Fondements scientifiques	4
3.1	Indexation multimédia	4
3.1.1	Espace de représentation	5
3.1.2	Construction d'index	5
3.2	Recherche Interactive	5
4	Domaines d'applications	6
5	Logiciels	7
6	Résultats nouveaux	7
6.1	Descripteurs d'images	7
6.1.1	Décrire la couleur des images	7
6.1.2	Décrire la texture des images	8
6.1.3	Vers des descripteurs génériques	9
6.2	Reconnaissance d'empreintes digitales	9
6.2.1	La segmentation robuste de l'image	10
6.2.2	Contrôle de qualité	11
6.2.3	Détection des centre(s) et delta(s) de l'empreinte	11
6.3	Détection hiérarchique de visages par apprentissage statistique	11
6.4	Interrogation et consultation bi-modales texte/image d'une base d'images	13
6.5	Requêtes partielles	16
6.5.1	Segmentation systématique	16
6.5.2	Segmentation automatique par DFDM	16
6.5.3	Segmentation par classification compétitive	17
6.6	Contrôle de pertinence	20
7	Contrats industriels (nationaux, européens et internationaux)	21
7.1	Relations industrielles	21
7.1.1	Contrat Alcatel	21
7.1.2	Contrat Thomson-CSF Idmatics	21
7.1.3	Contrat Elucid Technologies	21
7.2	Actions nationales	21
7.2.1	Projet du Réseau National de Recherche en Télécommunications RECIS (Recherche et Exploration par le Contenu Image et Son)	21
7.2.2	Action de Recherche Collaborative SIRIUS	21
7.3	Actions internationales	21
7.3.1	Projet STIC (9902) "Indexation et recherche d'images par le contenu".	21

8	Diffusion de résultats	22
8.1	Animation de la Communauté scientifique	22
8.2	Enseignement	22
9	Bibliographie	23

1 Composition de l'équipe

Responsable scientifique

Chahab Nastar [CR1 INRIA]

Responsable permanent

Nozha Boujemaa [MdC détachée sur un poste de CR1 INRIA]

Assistante de projet

Laurence Bourcier [à temps partiel]

Personnel INRIA

Jean-Paul Chièze [IR1 INRIA, à temps partiel]

Collaborateurs Extérieurs

Donald Geman [Professeur à l'université du Massachusetts et à l'Ecole Polytechnique, Conseiller scientifique]

Chercheurs post-doctorants

Sébastien Gilles [Contrat ARC avec Langue et Dialogue, depuis le 1/2/1999]

Constantin Vertan [Contrat RNRT Recis, depuis le 1/11/1999]

Alexandre Winter [Contrat Alcatel]

Doctorants

Sylvain Bernard [Bourse CIFRE, Thomson-CSF Idmatics depuis le 1/09/1998]

François Fleuret [Bourse INRIA]

Christophe Meilhac [Bourse MENRT, Université Paris V, jusqu'au 1/9/1999]

Bertrand Le Saux [Bourse INRIA Rocq depuis le 1/11/1999]

Nathalie Poirier [Bourse MENRT, Université Paris VI, depuis le 1/10/1999]

Hichem Sahbi [Bourse de la coopération franco-algérienne depuis le 1/10/1999]

Stagiaires

Julien Fauqueur [octobre-décembre 1999]

Nathalie Poirier [mars-août 1999]

Itheri Yahiaoui [mars-août 1999]

Alexandre Jacolin [mars-juin 1999]

Nicolas Israel [mars-juin 1999]

2 Présentation et objectifs généraux

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 (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.

Globalement, notre travail de recherche s'articule autour de deux axes principaux :

1. l'indexation multimédia, qui concerne à la fois l'étude des descripteurs d'images génériques, ceux spécifiques à une application, mais aussi l'indexation bi-modale texte/image ;
2. la recherche interactive, qui regroupe nos recherches sur le contrôle de pertinence et les requêtes partielles, ainsi que la navigation intelligente dans les bases d'images.

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.

Le logiciel **Surfimage** matérialise les recherches effectuées au sein d'IMEDIA. Il permet notamment d'effectuer des interrogations à la fois dans des bases hétérogènes et dans des bases avec vérité terrain, en adoptant une stratégie spécifique.

3 Fondements scientifiques

3.1 Indexation multimédia

Mots clés : base de données images, reconnaissance de forme, analyse d'image, extraction de caractéristiques, statistiques.

Participants : Chahab Nastar, Nozha Boujemaa, Donald Geman, Sébastien Gilles, Alexandre Winter, Constantin Vertan, Hichem Sahbi, Nathalie Poirier.

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 multimédias, 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 multimédias) par le contenu, d'une manière ergonomique et intuitive pour l'utilisateur [10]. Ce problème posé en des termes très concrets donne naissance à un certain nombre de modélisations mathématiques et informatiques que nous décrivons ci-après.

3.1.1 Espace de représentation

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, puis 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 (avec vérité-terrain), on développe des descripteurs dédiés et optimaux pour la tâche considérée (excentricité du contour d'un visage, position d'une tumeur...). Pour les bases génériques, à l'inverse, on extrait des descripteurs universels (couleur, texture, forme ...).

3.1.2 Construction d'index

La deuxième phase d'indexation consiste à stocker efficacement les signatures en machine. Cela revient à structurer le nuage de points des signatures d'images dans l'espace des caractéristiques, dans le but de réduire ultérieurement le coût de la requête (compromis coût du stockage/coût de la requête).

La construction d'index n'est pas un problème classique en raison des dimensions importantes des descripteurs qui empêchent de recourir aux structures arborescentes.

3.2 Recherche Interactive

Mots clés : similarité perceptuelle, mise en correspondance, théorie de l'information, estimation de l'incertitude.

Participants : Chahab Nastar, Nozha Boujemaa, Donald Geman, Jean-Paul Chièze, Bertrand Le Saux.

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

y 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 trois objectifs :

- *Recherche d'une image-cible*. La performance du système est alors quantifiable en termes du nombre d'interactions nécessaire pour atteindre cette image.
- *Catégorisation*. Ce problème englobe le précédent mais consiste également à estimer la distribution (habituellement multimodale) des catégories d'images qui intéressent l'utilisateur.
- *Navigation*. Il s'agit d'un problème 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 ?

4 Domaines d'applications

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

– les applications scientifiques

Exemple : dans le cas d'images médicales, retrouver les images présentant un caractère pathologique, dans un objectif éducatif ou diagnostique.

– les applications grand public

Exemple : butinage Web, commerce électronique.

– 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). Une thèse est actuellement en cours sur ce thème dans le projet.

– l'art, l'éducation

Exemples : recherche d'un tableau ou d'une illustration par un exemple, un croquis, un dessin ; recherche encyclopédique.

– les télécommunications

Exemple : coder et représenter les images par leur contenu sémantique, dans le contexte MPEG-4 et MPEG-7. IMEDIA n'est pas actif dans ces aspects normatifs mais suit les travaux en cours pour, le cas échéant, intégrer les standards dans sa méthodologie.

– le design, la publicité

Exemples : rechercher une texture spécifique pour l'industrie textile, illustrer une publicité par une photo adéquate.

– l’audiovisuel

Exemple : rechercher un plan spécifique d’un film. À long terme, cette application intègre l’image, le son et le texte.

5 Logiciels

Participant : Chahab Nastar [correspondant pour Surfimage].

Développé depuis 1996, **Surfimage** est un moteur d’indexation et de recherche d’images par le contenu qui cristallise la majorité des recherches menées à IMEDIA. Il est écrit en C++ avec interfaçage en Tcl/Tk. Il a été compilé sous Solaris, IRIX, Alpha, Linux, et NT.

Une version interactive (mais bridée) est disponible à l’adresse : <http://www-rocq.inria.fr/cgi-bin/imedia/surfimage.cgi>. Ce logiciel est actuellement en cours de rachat par Elucid Technologies, start-up de l’INRIA.

6 Résultats nouveaux

6.1 Descripteurs d’images

Mots clés : espace couleur, modèle de texture, arrangement spatial, attribut générique.

Participants : Nathalie Poirier, Alexandre Winter, Sébastien Gilles, Chahab Nastar.

Glossaire :

attribut Pour décrire le contenu d’une image, on peut utiliser divers attributs, tels que la couleur, la forme ou la texture.

représentation Pour chacun des attributs, on choisit une représentation. Par exemple, pour représenter la couleur, on peut choisir l’histogramme.

signature La signature d’une image est la traduction numérique d’une représentation. C’est généralement un vecteur de dimension très élevée.

Ce travail a été réalisé pendant le stage de DEA de Nathalie Poirier [13].

Depuis la création du projet IMEDIA, de nombreuses signatures d’images ont été développées pour traduire des attributs d’images très divers tels que la couleur, la forme et la texture. Il devenait nécessaire de faire le point sur ces descripteurs. Par conséquent, nous nous sommes intéressés successivement aux attributs de couleur, de texture, à la description spatiale d’attributs et aux descripteurs génériques.

6.1.1 Décrire la couleur des images

De nombreuses méthodes ont déjà été employées pour traduire numériquement la couleur d’une image. Notre but était de choisir un espace couleur efficace pour la recherche d’images par le contenu en utilisant les histogrammes couleur. Nous avons ainsi montré expérimentalement, via des courbes précision/rappel, que l’espace couleur HSV était plus adéquat que l’espace couleur RGB ou même l’espace LUV pour rendre compte du contenu visuel des images.

6.1.2 Décrire la texture des images

Nous avons ensuite étendu le modèle multi-échelles auto-régressif (modèle MRSAR [13]) en formulant une autre signature extraite de ce modèle. Elle s'avère être bien plus efficace que les traditionnelles signatures de texture, telles les signatures de Fourier et les Ondelettes, car elle s'affranchit davantage de l'échelle de la texture (FIG. 1 et FIG. 2).



FIG. 1 – Quelques lacunes des descripteurs de texture actuels ; le système Surfimage retrouve les images les plus similaires à la requête (image entourée de bleu en haut à gauche) en se fondant sur les descripteurs de Fourier.



FIG. 2 – Le modèle MRSAR s'affranchit davantage de l'échelle de la texture que les descripteurs de Fourier ; le système Surfimage retrouve les images les plus similaires à la requête (image entourée de bleu en haut à gauche) en se fondant sur le descripteur MRSAR.

6.1.3 Vers des descripteurs génériques

Enfin, nous avons axé notre étude sur des signatures combinant plusieurs informations sur l'image, traduisant par exemple l'arrangement spatial des couleurs ou celui des formes (vecteurs cohérents et autocorrélogrammes [13]). Ce type de signature peut être étendu à toute sorte d'attribut. Il constitue de ce fait un pas vers les descripteurs génériques d'images. Nous les avons testés avec succès sur la couleur et la forme (orientation des contours). Dans ces cas, les performances sont bien meilleures que les descripteurs spécifiques sans pour autant nécessiter un temps de calcul important (FIG. 3 et FIG. 4).

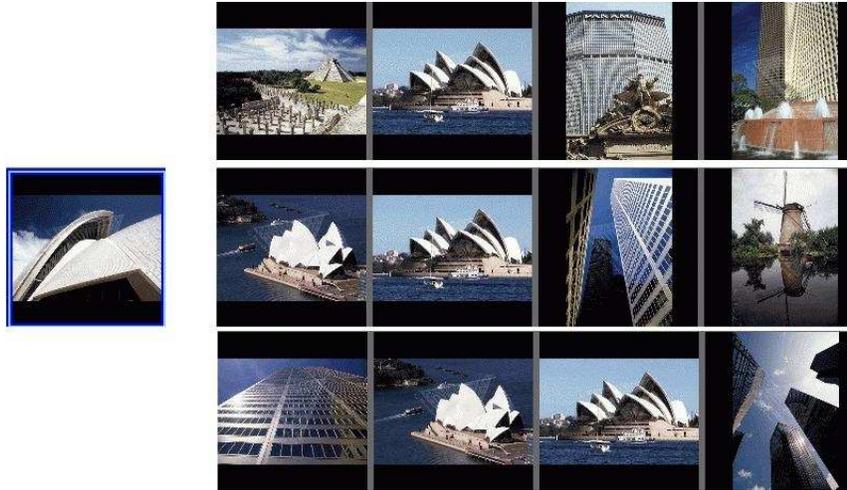


FIG. 3 – De haut en bas, l'image requête, les résultats avec l'histogramme couleur usuel, ceux obtenus avec les vecteurs cohérents de couleur et ceux obtenus avec les autocorrélogrammes de couleurs (espace HSV).

6.2 Reconnaissance d'empreintes digitales

Participants : Sylvain Bernard, Chahab Nastar, Nozha Boujema.

Mots clés : empreinte digitale, reconnaissance, mise en correspondance.

Il s'agit du travail de thèse de Sylvain Bernard (CIFRE avec la société THOMSON-CSF IDMATICS). THOMSON-CSF IDMATICS fournit des systèmes de production de titres d'identité (carte d'identité, passeport, permis de conduire). Ses clients étant des Etats (Gabon, Kenya, Pologne), cette société doit gérer de très grandes bases d'images (empreintes, visages). THOMSON-CSF IDMATICS a besoin d'un système propriétaire de reconnaissance des empreintes digitales (AFIS).

Nous voyons un AFIS comme un système d'indexation et de recherche d'images sur une base spécifique. Par conséquent, nous exploitons un grand nombre de connaissances propres au domaine de la reconnaissance à l'aide d'empreintes digitales.

Pour savoir si deux empreintes proviennent du même doigt, les experts détectent leur ensemble de minuties (bifurcations ou points de fin de lignes de crêtes) et les mettent en



FIG. 4 – *De haut en bas: image requête, résultats avec l'histogramme des orientations de contours multi-résolutions, résultats avec les vecteurs cohérents de contours et résultats avec l'autocorrélogramme des contours.*

correspondance. Si le nombre de minuties communes est suffisant, on affirme que les deux empreintes proviennent du même doigt.

Du fait de la grande taille des bases de données, une classification préalable des empreintes permet un gain de temps considérable (FIG. 5).

Nous avons concentré nos recherches sur la détection automatique des minuties et la classification des empreintes. Nous avons obtenu trois résultats majeurs :

1. une segmentation robuste et rapide de l'image conduisant à la détection efficace des minuties,
2. un contrôle qualité de l'empreinte,
3. une détection des centre(s) et delta(s) de l'empreinte donnant une information de classe.

6.2.1 La segmentation robuste de l'image

Pour des questions de robustesse, nous ne détectons pas directement les minuties, mais nous procédons en trois étapes :

- Nous détectons les lignes de crêtes et les vallées de l'empreinte. Cette segmentation utilise la connaissance a priori qu'une empreinte a localement une forte structure périodique et orientée, que nous modélisons par une ondelette de Gabor. Une image améliorée est ainsi obtenue par un filtrage multiéchelle.
- Cette image améliorée est ensuite binarisée ce qui correspond, en tout point, à décider de son appartenance à une ligne de crête ou à une vallée.

- Une squeletisation des lignes de crêtes nous permet une détection efficace et rapide des minuties.

6.2.2 Contrôle de qualité

Nous avons poussé la segmentation jusqu'à la détection des zones de bruit de l'empreinte et du fond de l'image. Nous pouvons ainsi attribuer une note de qualité à l'empreinte, ce qui nous permet de rejeter les empreintes de mauvaise qualité en entrée du système. Cet algorithme a été intégré à l'offre THOMSON-CSF IDMATICS.

6.2.3 Détection des centre(s) et delta(s) de l'empreinte

Nous avons développé un algorithme de détection des centre(s) et delta(s) de l'empreinte. Leur nombre et leur position relative apporte une information de classe qui sera, dans un proche avenir, fusionnée à la sortie d'un réseau de neurone développé l'année passée [5].

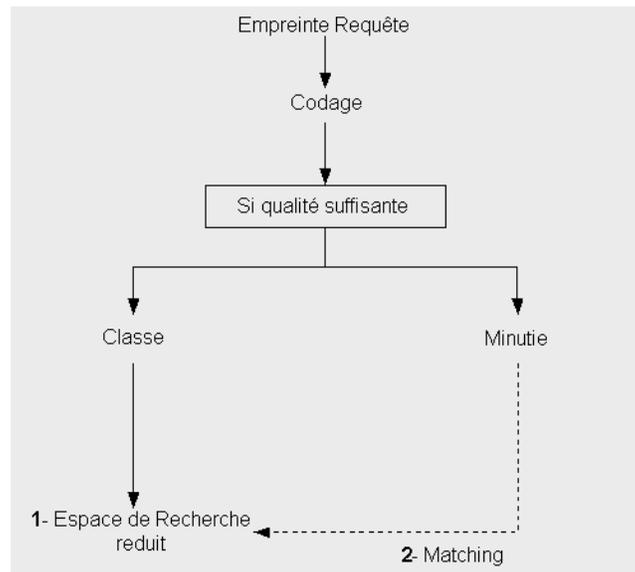


FIG. 5 – Schéma global de la recherche.

6.3 Détection hiérarchique de visages par apprentissage statistique

Mots clés : Détection, apprentissage statistique, visage, corrélation.

Participants : François Fleuret, Donald Geman.

Nous nous intéressons à la détection d'objets, et plus précisément de visages, dans des scènes en niveaux de gris, sans information de mouvement ou de profondeur (FIG. 6). L'approche que nous proposons consiste à construire durant une phase d'apprentissage une représentation très

hiérarchisée de l'objet "visage", afin de pouvoir organiser la détection de manière algorithmiquement efficace.

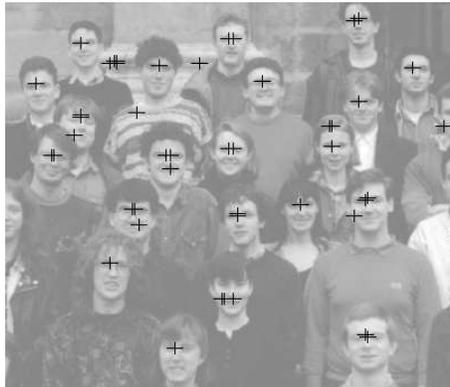


FIG. 6 – *Exemple de détection de visages.*

Nous introduisons l'idée de ρ -décomposition, qui nous permet de construire des structures géométriques qui ont une forte probabilité de présence sur les visages, et une probabilité très faible sur le "fond". Elles peuvent être construites sans avoir à estimer simultanément un grand nombre de paramètres interdépendants, et nous évitons ainsi l'écueil classique du sur-ajustement aux données que l'on rencontre dans la plupart des classificateurs non-paramétriques (par exemple les réseaux de neurones).

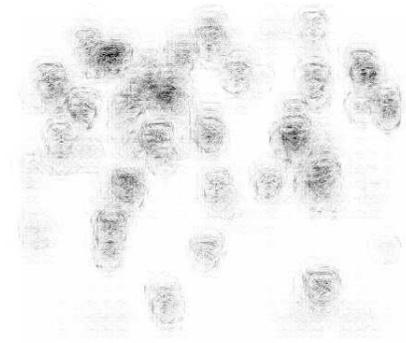


FIG. 7 – *Intensité de l'utilisation des zones de l'image pendant la détection. La hiérarchisation de la représentation permet de concentrer le calcul là où les visages sont effectivement présents.*

Nous utilisons ensuite ces structures pour construire plusieurs détecteurs dédiés à des sous-

ensembles de visages dont la pose est contrainte. Chacun de ces détecteurs a une forme hiérarchique et ne teste la présence de structures complexes que si l'hypothèse "visage" n'a pas pu être rejetée en ne considérant que des structures plus simples.

Enfin, ces détecteurs dédiés sont combinés de manière à déterminer efficacement s'il y a un visage présent à un endroit donné de la scène, et quelle est sa localisation précise. Cette hiérarchisation poussée permet d'accumuler les calculs aux endroits où il y a ambiguïté, et a contrario de les réduire au maximum partout ailleurs (FIG. 7). L'intégralité du processus de détection sur la scène de la figure 6 demande 0.6 seconde sur un PC standard.

Nous justifions théoriquement la légitimité de cette approche en prouvant d'une part que les structures ρ -décomposables ont effectivement une probabilité élevée d'être présentes dans les images de visages, et d'autre part en donnant des résultats préliminaires sur l'optimalité de la manière dont sont combinés les détecteurs dédiés.

Ces travaux sont décrits dans [1] et [6].

6.4 Interrogation et consultation bi-modales texte/image d'une base d'images

Mots clés : image, mot-clé, théorie de l'information, probabilités, classification.

Participants : Sébastien Gilles, Chahab Nastar, Nozha Boujemaa.

L'enjeu de cette étude est le développement d'outils théoriques permettant de manipuler attributs textuels et attributs visuels pour l'interrogation d'une base d'images éventuellement annotées ou munies de mots-clés. Le spectre de l'information contenue dans une telle base est donc très large (sémantique et syntaxique). A cet égard, cette étude aborde la question de la fusion et de l'intégration d'informations multi-modales et hétérogènes, question essentielle pour les futurs systèmes d'accès intelligents à des bases de données multimédia (vidéo, image, son, texte). Par ailleurs, il y a un lien fort entre cette étude et les initiatives MPEG-7 et XML, visant à standardiser la manière de décrire des documents multimédias, puisque les algorithmes que nous développons peuvent être envisagés à terme comme des méthodes définies sur des objets multimédias.

Nous avons commencé par envisager plusieurs scénarios "utilisateur" typiques, d'abord au niveau de l'interrogation de la base :

- interrogation par une image,
- interrogation par le texte,
- interrogation bi-modale texte/image.

L'interrogation faite, l'utilisateur peut rechercher des informations différentes :

- recherche d'images similaires,
- recherche de mots-clés probables.

Par exemple, nous avons intitulé “propagation de mots-clés” la technique (se rapprochant de la classification) consistant à interroger une base avec une image externe à la base et de récupérer en sortie non seulement les images similaires mais aussi les mots-clés les plus probables pour cette image. Au cœur du problème est la définition d’une métrique permettant de comparer texte et image. Étant donnée la nature doublement hétérogène des données (sémantique et discret pour le texte, syntaxique et (pseudo-)continu pour l’image), nous avons commencé par définir un nouvel espace de représentation des mots et des images :

Chaque mot W est vu à la fois comme un état possible (au sens des variables aléatoires) et comme une variable aléatoire prenant valeur dans un ensemble d’états, avec une distribution discrète de probabilités de transition vers ces états. Chaque mot est donc représenté de façon duale, soit en tant qu’état, soit en tant que variable aléatoire. La remarque importante est qu’il suffit de connaître la distance entre deux états pour définir des probabilités de transition entre ces deux états. Par exemple, on peut définir la notion *GUITARE* par l’ensemble d’états {instrument, Dylan, corde, musique, manche, mediator} avec une probabilité associée à chacun des états. Pour la distance inter-mots, une solution consiste à utiliser un thesaurus.

Nous appliquons le même procédé aux images : sachant que nous disposons d’une notion de distance entre des images (la distance basée sur le contenu), on peut représenter une image comme une variable aléatoire prenant valeur dans un ensemble d’états, avec des probabilités de transition dépendant uniquement de la distance entre les états.

L’intérêt de cette représentation est qu’il devient possible de comparer une image avec un mot ou un concept, ou un ensemble d’images avec un ensemble de mots-clés. Pour ceci, on utilise l’information mutuelle, une métrique statistique issue de la théorie de l’information qui mesure la prédictabilité d’une variable aléatoire, étant donnée une autre variable aléatoire. Comme la nature ou l’existence d’une fonction liant les deux variables n’entre pas en ligne de compte, cette métrique est très bien adaptée au cas du texte et de l’image [2]. L’information mutuelle entre un mot et une image peut être vue comme un moyen de quantifier le degré de cohérence entre les aspects sémantiques et syntaxiques. Ainsi, pour une variable aléatoire textuelle W et une variable aléatoire image I , on peut calculer :

$$MI(W, I) = H(W) + H(I) - H(W, I)$$

Cette métrique nécessite de calculer les probabilités jointes de co-occurrence d’un mot et d’une image $P(w, i)$ (les minuscules référant aux états plutôt qu’aux variables). Au premier ordre, on peut simplement compter le nombre d’occurrences de la paire (w, i) dans la base. Le résultat est alors fortement dépendant de l’indexation présente dans la base. Pour y remédier, on peut calculer cette probabilité jointe en prenant en compte les états proches de l’image et proches du mot, à l’aide d’une technique d’estimation de densité, par exemple les fenêtres de Parzen.

Si l’on peut quantifier le degré de cohérence entre le texte et l’image, on peut donc s’intéresser au *spectre visuel d’un mot-clé*, c’est-à-dire à l’ensemble des images dont les mots-clés sont proches, muni d’une distribution de probabilités. On sait d’emblée qu’il n’y a pas une correspondance bijective entre un mot et une image, mais l’on sait aussi (autre extrême) qu’il n’y a pas indépendance entre les deux. La métrique développée permet de quantifier la cohérence

de cette relation.

Entre autres, le système peut tenter de raisonner sur sa propre connaissance. Par exemple, si, dans une taxonomie de langage (ou les mots sont organisés en catégories hiérarchiques), un mot W_1 est un père de W_2 , le spectre visuel $Sp(W_1)$ doit être d'entropie supérieure au spectre visuel $Sp(W_2)$. Sous réserve d'une base représentative, la réciproque peut être obtenue. Le système pourrait dans ce cas organiser sa propre taxonomie et ordonner la base d'images en groupes sémantiques, ou en groupes visuels. Le système peut aussi, lorsqu'on l'interroge à partir d'une image externe sans mot-clé (FIG. 8), suggérer des mots-clés probables pour cette image (propagation de mots-clés). L'idée est ici de déterminer l'ensemble d'états $W = \{w_1, \dots, w_n\}$ qui maximise l'information mutuelle avec les images proches de l'image requête. Les travaux futurs concerneront l'adaptation de cette approche aux requêtes partielles.

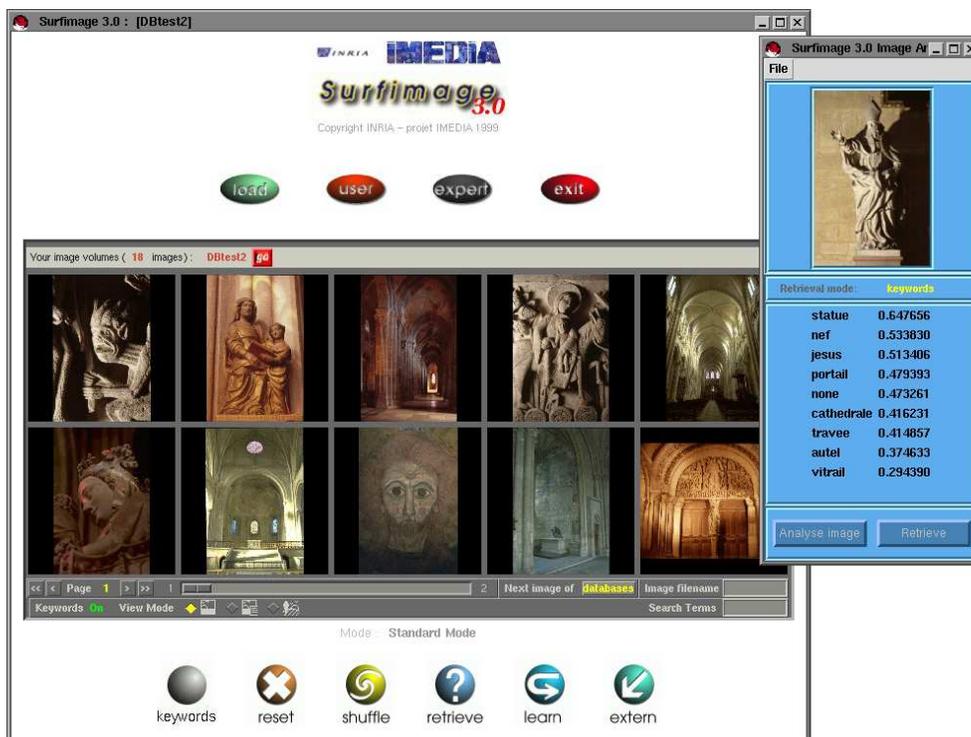


FIG. 8 – Une requête visuelle à partir d'une image externe à la base sans mot-clé. Dans un premier temps, Surfimage retrouve les images les plus similaires. Dans un deuxième temps, Surfimage détermine les mots-clés les plus probables pour cette image, en utilisant l'algorithme de maximisation de l'information mutuelle. Les images de la base sont munies d'une série de mot-clés, non visibles à l'écran. En l'occurrence, les mots-clés les plus probables ont été retournés avec les scores suivants: ("statue", 0.64), ("autel", 0.53), ("vitrail", 0.51), ("cathédrale", 0.479). On remarque que le mot-clé de loin le plus probable est "statue", qui décrit bien l'image requête.

6.5 Requêtes partielles

Mots clés : segmentation, région, zone d'intérêt, objets.

Participants : Nozha Boujemaa, Itheri Yahiaoui, Alexandre Winter, Chahab Nastar.

Résumé : *Nos recherches récentes se sont intéressées à la caractérisation des propriétés de l'image spatialement localisées et à leurs mesures de similarité. En effet, l'utilisateur est le plus souvent intéressé par 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. Dans l'objectif de formuler des requêtes partielles (multiples ou non) sur une image, nous avons entrepris des travaux pour répondre à sa première phase, à savoir la subdivision de l'image. Différentes approches de subdivision d'images sont possibles pour permettre l'interrogation d'une base à partir d'une zone d'intérêt: approches par contourage manuel de la zone, approche par subdivision systématique de l'image, subdivision automatique par des méthodes de segmentation. Nous allons présenter nos travaux relevant des deux dernières approches, étant donné que nous considérons que le contourage manuel de chaque objet d'intérêt dans une image peut constituer rapidement une tâche lourde pour l'indexation d'une base surtout si elle contient un nombre très important d'images (quelques centaines de milliers d'images).*

6.5.1 Segmentation systématique

La première approche développée [7] consiste à procéder par subdivision systématique en maillage multirésolution de l'image. Elle représente un moyen simple et efficace pour sélectionner des parties de l'image et faire une requête. Ce mode de subdivision de l'image permet à l'utilisateur de préciser localement des sous-images en évitant l'étape difficile de la segmentation d'autant plus que toutes les images dans une base ne se présentent pas sous forme de collection d'objets. Dans le cas de requêtes multiples sur la même image, nous construisons leur rectangle englobant afin de figer leurs relations spatiales. Une mesure de similarité appropriée a été mise au point pour gérer l'invariance par translation de ce masque regroupant les sélections multiples de l'utilisateur. Les signatures de **Surfimage** sont calculées et combinées sur chacune des sous-images de la grille multirésolution. Ainsi, la simplicité de ce mode d'interaction flexible cache une complexité dans l'espace des signatures résolue par une structure du type kd-tree. Nous présentons des résultats (FIG. 9) obtenus sur une base de 3670 images hétérogènes contenant des tableaux de grands maîtres, textures, visages, et scènes naturelles.

6.5.2 Segmentation automatique par DFDM

L'idée est la suivante [11, 12]: pour décider qu'une région est cohérente, on estime que la distribution d'un certain attribut d'image est stable sur cette région. En d'autres termes, on ne s'appuie que sur la distribution (complexe, multimodale etc.) de l'attribut, évitant ainsi de *modéliser* avec une connaissance a priori toujours discutable (e.g. distribution gaussienne).

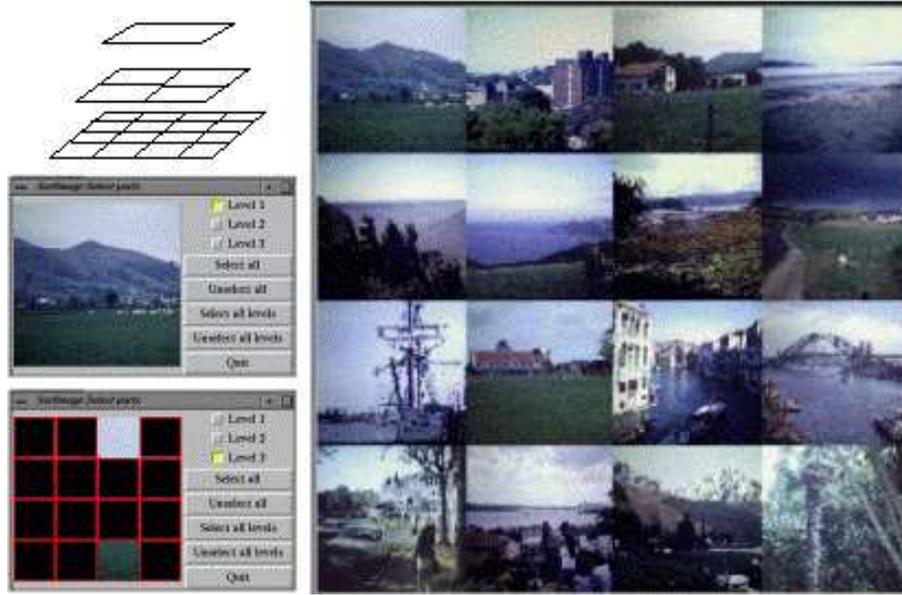


FIG. 9 – Requête partielle par subdivision systématique.

Plus précisément, on calcule, de manière multi-résolution et multi-attributs, un DFDM (Differential Feature Distribution Map). Le DFDM est une image qui mesure la stabilité de l'attribut considéré en chaque pixel (FIG. 10):

$$D(i, j) = (d_1 + d_2 + d_3 + d_4)/4 \tag{1}$$

où:

$$d_1 = d(X_1, X_5) \tag{2}$$

$$d_2 = d(X_2, X_6)/\sqrt{2} \tag{3}$$

$$d_3 = d(X_3, X_7) \tag{4}$$

$$d_4 = d(X_4, X_8)/\sqrt{2} \tag{5}$$

L'organigramme du calcul du DFDM est montré sur la figure 11.

Des exemples de segmentation sont visibles sur la figure 12 ainsi qu'à l'adresse: <http://www-rocq.inria.fr/~awinter/Segmentation.html>.

6.5.3 Segmentation par classification compétitive

La détection de zones saillantes, par segmentation automatique, reste très attractive pour la recherche de similarité par le contenu. Ce mode de sélection partielle pourrait être considéré comme très ambitieux, compte tenu des difficultés connues de la segmentation, à moins de préciser les objectifs à atteindre. En effet, dans un contexte de recherche d'image par le contenu, le problème de la segmentation ne se pose pas de la même manière que dans un contexte de

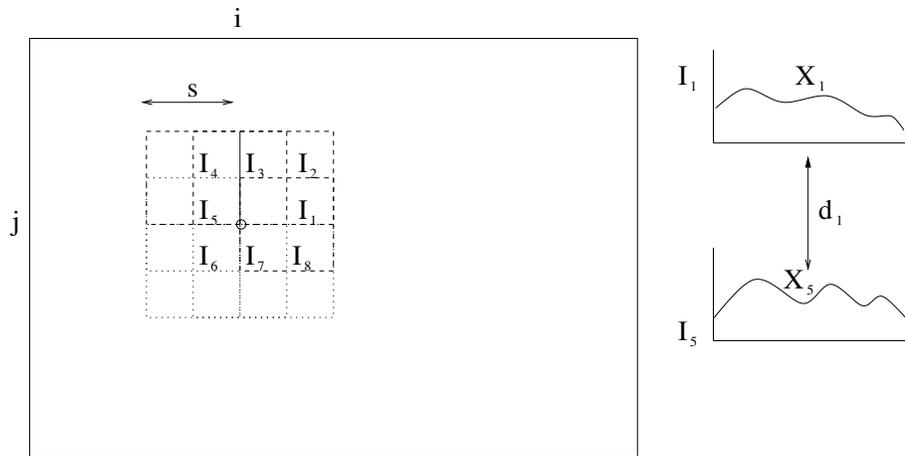


FIG. 10 – Le DFDM unifie le calcul dans l'espace image et dans l'espace des caractéristiques.

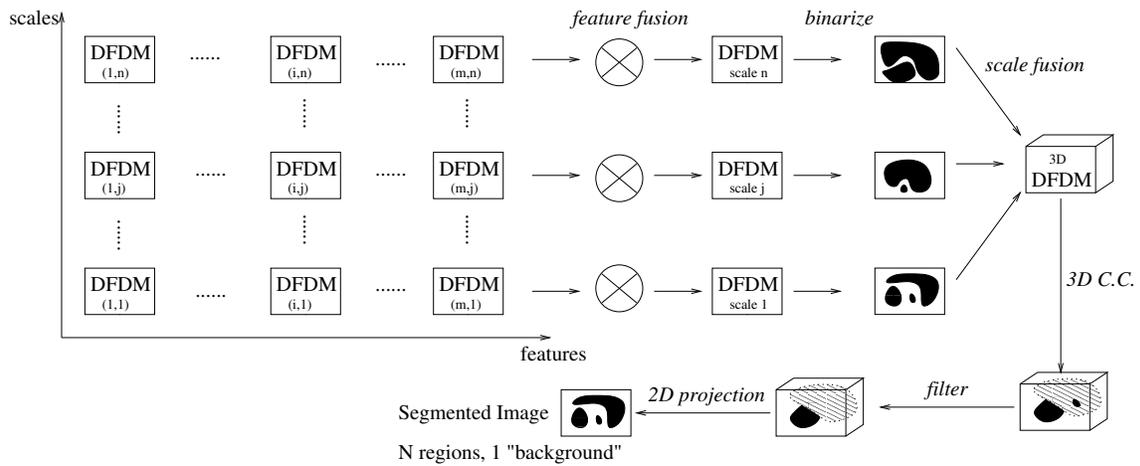


FIG. 11 – Organigramme de la segmentation par DFDM.



FIG. 12 – Exemple de segmentation par DFDM.

reconnaissance d'objets. Notre méthode de segmentation repose sur un algorithme de classification non-exclusive par compétition agglomérative dans un espace de descripteurs de couleur. La propriété importante de cette méthode est sa capacité à déterminer automatiquement le nombre de classes contrairement à la plupart des méthodes existantes. La fonction objectif est une combinaison de deux termes à effets antagonistes qui crée le mécanisme de compétition entre les classes. Les classes de faible cardinalité sont amenées à disparaître. A la convergence, un nombre de classes optimal résulte de l'équilibre atteint entre le terme de fidélité aux données et le terme de réduction de complexité. Les zones dominantes ainsi formées constitueront les masques utilisés pour effectuer des requêtes partielles. Ce travail est soumis à publication [14]. La figure 13 présente un résultat de segmentation obtenu en prenant comme descripteur



FIG. 13 – Segmentation de la région usine présentant une variabilité importante de couleurs.

la distribution des couleurs locales autour d'un pixel. D'autres résultats sont présentés dans <http://www-rocq.inria.fr/~boujemaa/Partielle2.html>.

6.6 Contrôle de pertinence

Participants : Christophe Meilhac, Alexandre Jacolin, Nicolas Israel, Chahab Nastar.

Mots clés : profilage, apprentissage, bouclage de pertinence, feedback.

Le contrôle de pertinence est un sujet que nous avons abordé dès 1997 [9]. Il s'agit d'effectuer un apprentissage statistique online des interactions utilisateurs pour mieux répondre à ses besoins (affinage de la requête, catégorisation, navigation). Dans [9], nous avons proposé une estimation des densités des images positives et négatives fournies par l'utilisateur dans une approche analogue au maximum de vraisemblance. Plus récemment [8], nous avons replacé cette problématique dans un cadre Bayésien permettant de résoudre le problème de la catégorisation subjective des images de la base à partir de quelques exemples fournis par l'utilisateur. La méthode évalue des distributions non paramétriques et utilise la règle de décision de Bayes pour estimer si une image est pertinente, non-pertinente, ou ambiguë pour l'utilisateur. Plusieurs stratégies utilisateur sont proposées : rechercher les images les plus probables, lever l'ambiguïté... [8] (FIG. 14). On a de plus montré quantitativement que la méthode est dix fois meilleure que le hasard pour retrouver une image-cible.

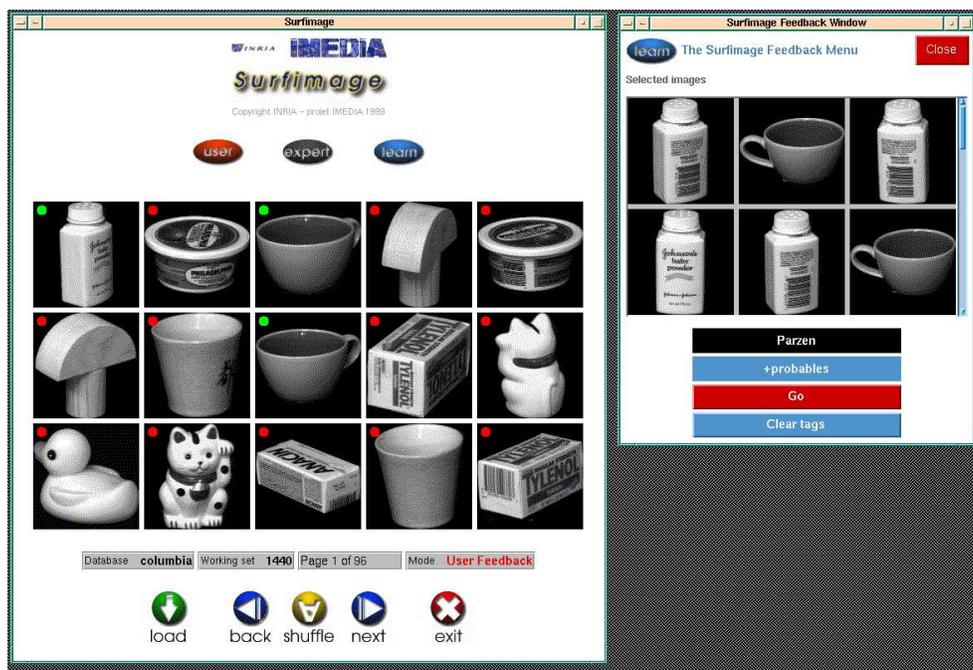


FIG. 14 – Les exemples positifs de l'utilisateur permettent, lors de la navigation, d'associer une pastille verte (pertinence) ou rouge (non-pertinence) à toutes les images de la base.

7 Contrats industriels (nationaux, européens et internationaux)

7.1 Relations industrielles

Participants : Alexandre Winter, Sylvain Bernard, Nozha Boujemaa, Chahab Nastar.

7.1.1 Contrat Alcatel

Le second contrat entre IMEDIA et ALCATEL-CRC (Marcoussis) s'est terminé en août 1999. Son titre était "Navigation intelligente dans les bases d'images" et il faisait suite à un précédent contrat sur "bibliothèque de signatures d'images pour l'indexation par le contenu".

7.1.2 Contrat Thomson-CSF Idmatics

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.

7.1.3 Contrat Elucid Technologies

Ce contrat de transfert du logiciel **Surfimage** est en cours de finalisation (novembre 1999) avec Elucid Technologies, jeune start-up de l'INRIA.

7.2 Actions nationales

Participants : Sébastien Gilles, Chahab Nastar, Constantin Vertan, Nozha Boujemaa.

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

C'est un projet de la catégorie exploratoire - qui implique l'avant-projet IMEDIA, France Telecom, Philips, 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.

7.2.2 Action de Recherche Collaborative SIRIUS

Cette ARC implique le projet IMEDIA ainsi que le projet Langue et Dialogue à Nancy (Malika Smail, Brigitte Simonnot). Il s'agit d'étudier l'indexation bi-modale par les mots-clés et par le contenu de l'image d'une manière originale, en particulier à travers le stage post-doctoral de Sébastien Gilles.

7.3 Actions internationales

7.3.1 Projet STIC (9902) "Indexation et recherche d'images par le contenu".

Participants : Nozha Boujemaa, Chahab Nastar.

C'est un projet de collaboration qui implique trois sites de l'INRIA: Rocquencourt (IMEDIA), Rennes (Vista) et Grenoble (Movi) avec l'université de Tunis II. Il finance sur un an, trois stages avec plusieurs séjours de courte durée dans les deux sens - notification juillet 99. Nozha Boujemaa est responsable, du côté français, de ce projet.

8 Diffusion de résultats

8.1 Animation de la Communauté scientifique

Chahab Nastar :

- Membre du comité de pilotage et du comité de programme de First European Workshop on Content-Based Multimedia Indexing (CBMI'99), Toulouse, octobre 1999.
- Membre du comité de programme du 12ème Congrès Francophone AFRIF-AFIA de Reconnaissance des Formes et Intelligence Artificielle (RFIA'2000), Paris, février 2000.
- Membre du comité de programme du Workshop "Multimedia Intelligent Storage and Retrieval Management", ACM Multimedia'99, Orlando, octobre 1999.
- Expertise de l'action "Cognitique" du Ministère de la Recherche.

Nozha Boujemaa :

- Secrétaire général adjoint de l'AFRIF (Association Française de Reconnaissance et Interprétation des Formes) qui remplace l'AFCET dans son rôle de regroupement de la communauté Image et Reconnaissance des Formes (membre du bureau et du conseil d'administration)- janvier 99.
- Membre du " noyau dur " du groupe thématique du GDR PRC ISIS sur la couleur - mars 99.
- Co-organisatrice d'une journée sur la segmentation et l'indexation par la couleur dans le cadre du GDR PRC ISIS (groupe thématique " couleur ") - novembre 99.

8.2 Enseignement

Chahab Nastar :

- Séminaire professionnel sur la reconnaissance du visage, Development Institute International, journée biométrie, avril 99, Paris.

Nozha Boujemaa :

- Cours sur l' indexation par la couleur, Ecole d'été sur la couleur, septembre 1999, Saint-Etienne.
- Cours au DEA IARFA (Paris VI) sur l'indexation et la recherche d'images par le contenu, janvier 1999.

- Séminaire professionnel sur " Comment la reconnaissance par l'image peut garantir le contrôle d'accès à vos réseaux et systèmes d'information? " (Biométrie), Development Institute International - thème " la sécurité des réseaux et systèmes d'information ", Juin 99 - Paris.

Donald Geman : Professeur de probabilités à l'Ecole Polytechnique.

François Fleuret : Chargé de TD et TP d'algorithmique et de TD de probabilités en DEUG MASS à l'université Paris IX - Dauphine (96h).

9 Bibliographie

Thèses et habilitations à diriger des recherches

- [1] F. FLEURET, *Détection hiérarchique de visages par apprentissage statistique*, thèse de doctorat, Université Paris VI, 1999.
- [2] S. GILLES, *Robust Description and Matching of Images*, thèse de doctorat, Oxford University, December 1998.

Articles et chapitres de livre

- [3] F. FLEURET, E. BRUNET, «DEA: An Architecture for Planning and Classification», *Neural Computation*, To appear.
- [4] J. TAREL, N. BOUJEMAA, «A coarse to Fine 3D Registration Method Based on Robust Fuzzy Clustering», *Computer Vision and Image Understanding* 73, 1, 1999.

Communications à des congrès, colloques, etc.

- [5] S. BERNARD, C. NASTAR, N. BOUJEMAA, D. VITALE, C. BRICOT, «Fingerprint image retrieval in very large databases», *in: Proceedings of the Second Workshop on Automatic Identification Advanced Technologies*, Morristown, New Jersey, 1999.
- [6] F. FLEURET, D. GEMAN, «Graded Learning for Object Detection.», *in: Proceedings IEEE CVPR Workshop on Statistical and Computational Theories of Vision, Fort Collins*, 1999.
- [7] J. MALKI, N. BOUJEMAA, C. NASTAR, A. WINTER, «Region Queries without Segmentation for Image Retrieval by Content», *in: Int. Conf. on Visual Inf. Systems*, Amsterdam, June 1999.
- [8] C. MEILHAC, C. NASTAR, «Relevance Feedback and Category Search in Image Databases», *in: IEEE International Conference on Multimedia Computing and Systems*, Florence, Italy, June 1999.
- [9] C. NASTAR, M. MITSCHKE, C. MEILHAC, «Efficient Query Refinement for Image Retrieval», *in: Computer Vision and Pattern Recognition (CVPR '98)*, Santa Barbara, June 1998.
- [10] C. NASTAR, «Indexation et recherche d'images: enjeux, méthodes et perspectives», *in: (article invité), Congrès IDT*, Paris, June 1999.

- [11] A. WINTER, C. NASTAR, « Color Image Segmentation for Region Queries in Image Databases », *in: ACM Multimedia'99, Multimedia Intelligent Storage and Management Workshop*, Orlando, Florida, 1999.
- [12] A. WINTER, C. NASTAR, « Differential Feature Distribution Maps for Image Segmentation and Region Queries in Image Databases », *in: CVPR'99, IEEE Workshop on Content-Based Access of Image and Video Libraries*, Ft Collins, Colorado, 1999.

Divers

- [13] N. POIRIER, *Descripteurs Génériques d'images pour l'indexation par le contenu*, Mémoire, Université de Nice-Sophia Antipolis, 1999.
- [14] I. YAHIAOUI, *Segmentation Compétitive pour les requêtes partielles en recherche d'images par le contenu*, Mémoire, Université de Pierre et Marie Curie, 1999.