

Avant-projet IMEDIA

Images et Multimédia : Indexation, Recherche et Navigation

Rocquencourt

THÈME 3B



*R*apport
d'Activité

2000

Table des matières

| | | |
|----------|---|-----------|
| 1 | Composition de l'équipe | 3 |
| 2 | Présentation et objectifs généraux | 4 |
| 3 | Fondements scientifiques | 4 |
| 3.1 | Indexation d'images et espace de description | 5 |
| 3.2 | Requêtes partielles | 5 |
| 3.3 | Recherche interactive | 6 |
| 3.4 | Indexation Multimédia | 7 |
| 4 | Domaines d'applications | 7 |
| 5 | Logiciels | 8 |
| 5.1 | Logiciel Ikona/Maestro | 8 |
| 6 | Résultats nouveaux | 8 |
| 6.1 | Descripteurs d'images : Bases génériques | 8 |
| 6.1.1 | Intégration de la couleur et de la texture dans une signature unique . . . | 8 |
| 6.1.2 | Signature couleur adaptative | 14 |
| 6.2 | Descripteurs d'images : Bases spécifiques | 14 |
| 6.2.1 | Indexation de bases d'empreintes digitales | 14 |
| 6.2.2 | Indexation d'une base de visages | 16 |
| 6.3 | Généralisation de la classification compétitive | 18 |
| 6.4 | Requêtes partielles | 19 |
| 6.4.1 | Requêtes par régions d'intérêt | 19 |
| 6.4.2 | Détection et localisation de visages | 24 |
| 6.4.3 | Segmentation des Visages | 25 |
| 6.5 | Recherche interactive | 27 |
| 6.5.1 | Modèles de questions/réponses pour une recherche interactive | 27 |
| 6.5.2 | Navigation | 28 |
| 6.6 | Indexation multimédia | 31 |
| 6.6.1 | Estimation automatique de mots-clés | 31 |
| 7 | Contrats industriels (nationaux, européens et internationaux) | 33 |
| 7.1 | Relations industrielles | 33 |
| 7.1.1 | Contrat Thomson-CSF Idmatics | 33 |
| 7.1.2 | Contrat Thomson Multimedia | 33 |
| 7.1.3 | Contrat avec LookThatUp | 33 |
| 8 | Actions régionales, nationales et internationales | 33 |
| 8.1 | Actions nationales | 33 |
| 8.1.1 | Projet du Réseau National de Recherche en Télécommunications (RNRT) RECIS (Recherche et Exploration par le Contenu Image et Son) | 33 |

| | | |
|-----------|---|-----------|
| 8.1.2 | Projet du P rogramme pour la R echerche et l' I nnovation dans l' A udiovisuel et le M ulti M édia (PRIAMM) "MédiaWorks" | 33 |
| 8.1.3 | Action de Recherche Collaborative SIRIUS | 34 |
| 8.2 | Actions européennes | 34 |
| 8.2.1 | Projet Européen "STOP" | 34 |
| 8.3 | Actions internationales | 34 |
| 8.3.1 | Projet STIC (9902) "Indexation et recherche d'images par le contenu". . | 34 |
| 8.4 | Visites, et invitations de chercheurs | 34 |
| 9 | Diffusion de résultats | 34 |
| 9.1 | Animation de la Communauté scientifique | 34 |
| 9.2 | Enseignement | 35 |
| 10 | Bibliographie | 35 |

1 Composition de l'équipe

Responsable scientifique

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

Assistante de projet

Laurence Bourcier [à temps partiel dans le projet]

Personnel INRIA

Jean-Paul Chièze [IR1, à temps partiel dans le projet]

Conseiller scientifique

Donald Geman [Professeur à l'université du Massachusetts et à l'ENS Cachan]

Chercheurs post-doctorants

Constantin Vertan [Contrat RNRT Recis, jusqu'au 17/10/2000]

Ingénieurs experts

François Fleuret [depuis 1/03/2000, à mi-temps depuis octobre 2000]

Valérie Gouet [depuis décembre 2000]

Doctorants

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

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

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

Hichem Sahbi [Bourse de la coopération Franco-Algérienne depuis le 1/10/1999]

Marin Ferecatu [depuis le 20/09/2000]

François Fleuret [jusqu'au 31/01/2000]

Stagiaires

François Tonnin [avril-septembre 2000]

Roland Moquet [septembre-décembre 2000]

Moez Kamoun [juin-juillet 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 (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 l'élaboration et l'optimisation des descripteurs génériques de l'aspect visuel des images ainsi que de ceux spécifiques à une catégorie de bases d'images dédiées à une application particulière ;
2. les requêtes partielles 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 regroupe notre activité dans les deux thèmes suivants : "le contrôle de pertinence" et "la navigation intelligente" dans les bases d'images ;
4. l'indexation plurimédia et particulièrement 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, statistique, base de données image.

Participants : Nozha Boujemaa, Constantin Vertan, Hichem Sahbi, Sylvain Bernard, Jean-Paul Chièze.

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 multimédias) par le contenu, d'une manière ergonomique et intuitive pour l'utilisateur. Ce problème posé en des termes très concrets 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, 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 ...).

Une deuxième phase de construction d'index peut s'avérer utile dans le contexte d'un espace de représentation de dimension très importante. 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".

3.2 Requêtes partielles

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

Participants : Nozha Boujemaa, Julien Fauqueur, François Fleuret, Donald Geman.

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 subdiviser les images et constituer ces zones d'intérêt d'une manière efficace et pertinente,
- comment gérer l'indexation de la base par les signatures partielles et les mesures de similarité perceptuelle locale,
- comment gérer le problème des relations spatiales dans le contexte des requêtes partielles multiples : avec ou sans contraintes de positions relatives, d'invariances morphologiques...

3.3 Recherche interactive

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

Participants : Donald Geman, Nozha Boujemaa, Roland Moquet, Bertrand Le Saux, François Fleuret.

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, François Fleuret, Julien Fauqueur, Donald Geman.

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 *métadatas* (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.

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 objectif 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 un fournisseur d'images 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 dessin. IMEDIA a été contacté par le ministère de la culture au sujet de ses archives en images. Par manque de moyens humains, ces travaux n'ont pas encore commencé.

– 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 pour, le cas échéant, intégrer les standards dans sa méthodologie.

5 Logiciels

5.1 Logiciel Ikona/Maestro

Participants : Toute l'équipe , François Fleuret [correspondant pour Ikona].

L'évolution de nos travaux de recherche ainsi que la validation des algorithmes que nous développons nécessitent une architecture logicielle différente de celle du moteur "Surfimage". Surfimage a été 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. Ikona est écrit en Java, donc portable et facilement exploitable dans le contexte du WWW. Les figures 1 et 5.1 montrent l'interface du logiciel lors d'une requête.

Ce nouveau système fait déjà l'objet d'une collaboration industrielle avec Thomson Multimédia (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.

6 Résultats nouveaux

6.1 Descripteurs d'images : Bases génériques

6.1.1 Intégration de la couleur et de la texture dans une signature unique

Mots clés : couleur, texture de couleurs, histogrammes pondérés.

Participants : Constantin Vertan, Nozha Boujemaa.

L'approche classique pour l'indexation d'images couleur consiste à calculer la distribution statistique de premier ordre (l'histogramme) de la couleur contenue dans l'image, en utilisant un espace de représentation de couleur et une quantification convenables. Pour une image de

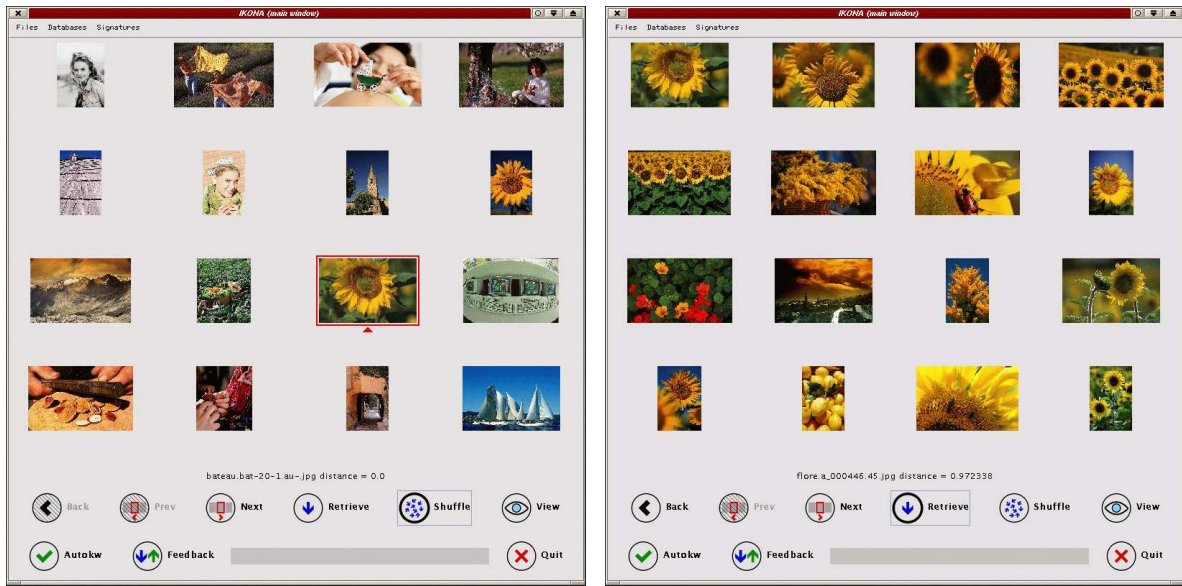


FIG. 1 – Interface du logiciel Ikona. À gauche, première page avec des images prises au hasard dans une base généraliste. L'image sur laquelle va porter la requête est entourée. À droite se trouve le résultat de la requête.

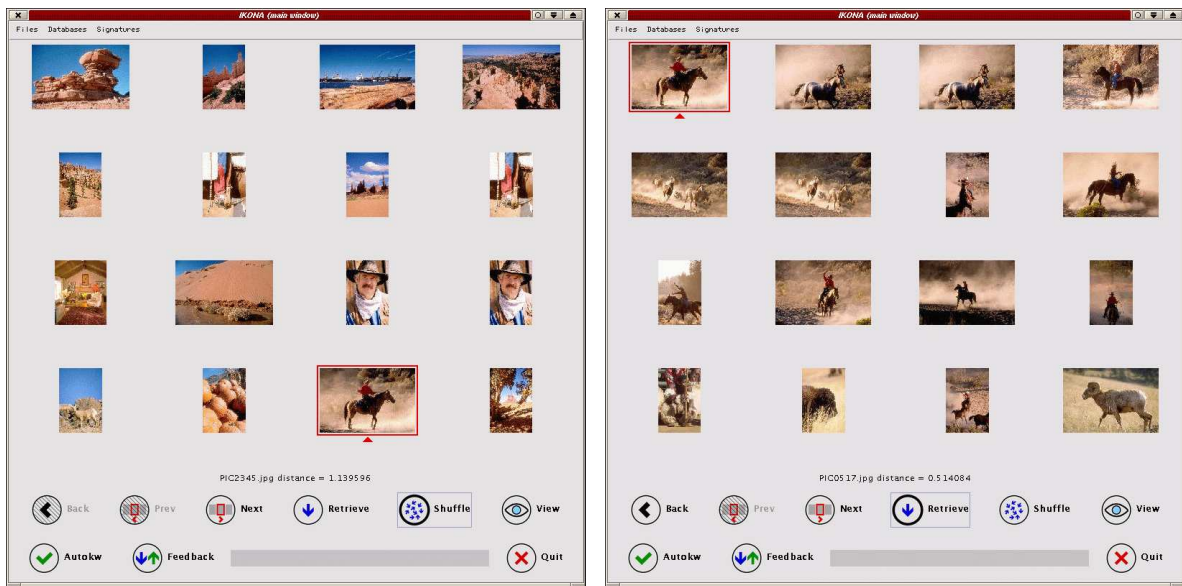


FIG. 2 – Interface du logiciel Ikona. La requête est faite dans une deuxième base généraliste.

taille $M \times N$ pixels, dont le pixel $f(i, j)$ a la couleur \mathbf{c} , l'histogramme peut être exprimé à l'aide du symbole de Kroneker δ comme suit :

$$h(\mathbf{c}) = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \delta(f(i, j) - \mathbf{c}), \quad \forall \mathbf{c} \in \mathcal{C} \quad (1)$$

Cette approche a pourtant prouvé ses limites, à travers son caractère global et la non-conservation de l'arrangement spatial des couleurs dans une image, qui constitue pourtant une propriété visuelle importante. En général, cette lacune est contournée par la combinaison de l'histogramme de couleur avec d'autres signatures, décrivant la structure et la texture. La description des images selon les statistiques d'ordre supérieur, englobant à la fois l'information de couleur et l'arrangement spatial des couleurs, est plus puissante du point de vue de la précision avec laquelle on retrouve certains types d'images, mais nécessite un temps de calcul important. La plupart des techniques de description de la texture sont appliquées à la composante de luminosité seulement, faute d'outils appropriés au caractère multicomposante de la couleur. Même si on a proposé des techniques pour la modélisation adaptative de la couleur par des nombres complexes [15, 10], la caractérisation séparée de la texture nous semble peu efficace.

L'approche innovante formulée en [12, 13, 16] est l'intégration des descripteurs de couleur et de texture dans une seule structure, à la fois simple et flexible : l'histogramme pondéré.

$$\tilde{h}(\mathbf{c}) = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} w(i, j) \delta(f(i, j) - \mathbf{c}), \quad \forall \mathbf{c} \in \mathcal{C} \quad (2)$$

Ainsi, on peut combiner l'histogramme des couleurs (la description classique du contenu de couleur d'une image) avec un attribut local $w(i, j)$, qui caractérise le voisinage de chaque pixel. Cette approche se révèle différente des approches classiques, étant donné que la signature résultante est plus compacte (permettant ainsi la réduction du temps de recherche dans la base d'images pour une requête donnée) et permet plus de flexibilité, via les multiples instantiations possibles de $w(i, j)$.

On peut caractériser le voisinage de chaque pixel par des attributs primaires comme :

- l'importance de la transition au pixel courant ;
- le nombre de pixels N_{ij} du voisinage de taille D ayant la même couleur c que le pixel courant (i, j) ,

$$N_{ij}(\mathbf{c}) = \sum_{m=-D/2}^{D/2} \sum_{n=-D/2}^{D/2} \delta(f(i+m, j+n) - \mathbf{c}) \text{ avec } c = f(i, j); \quad (3)$$

- le nombre maximal de pixels du voisinage de taille D ayant une même couleur N_{\max} ;
- la somme des distances de couleur entre le pixel courant et les autres pixels de son voisinage de taille D ,

$$s_{ij} = \sum_{m=-D/2}^{D/2} \sum_{n=-D/2}^{D/2} d(f(i+m, j+n), f(i, j)). \quad (4)$$

Ces informations de base peuvent se décliner selon plusieurs pondérations, comme l'usage du Laplacien (voir équation 5) [12] (dont le réglage de $k \in \{0, 1\}$ et de $\alpha \in \mathbb{R}$ permet de renforcer la prise en compte des pixels situés sur les contours ou à l'intérieur des régions uniformes), de la probabilité, de l'entropie normalisée. Le détail de ces pondérations avec d'autres descripteurs du voisinage se trouve dans [12, 13, 16, 14, 11]

$$w(i, j) = (k + |\Delta(i, j)|)^\alpha \quad (5)$$

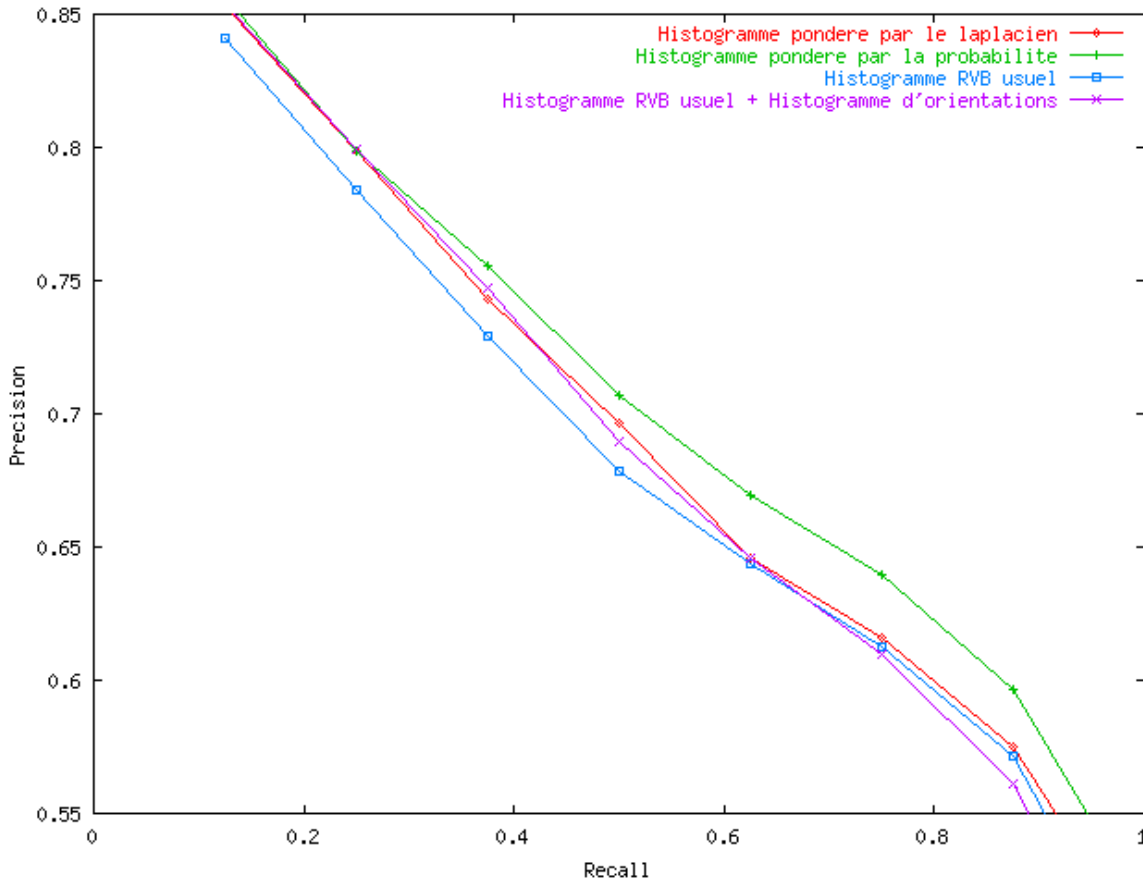


FIG. 3 – Courbes précision-rappel pour des requêtes dans la base généraliste de test selon l'histogramme pondéré par le laplacien, l'histogramme pondéré par la probabilité locale de la couleur, l'histogramme usuel. La pondération des histogrammes accroît la performance de manière significative; la performance de la recherche selon l'histogramme pondéré est semblable ou meilleure à la performance de la combinaison des histogrammes de couleur et d'orientation, mais pour un coût de recherche inférieur.

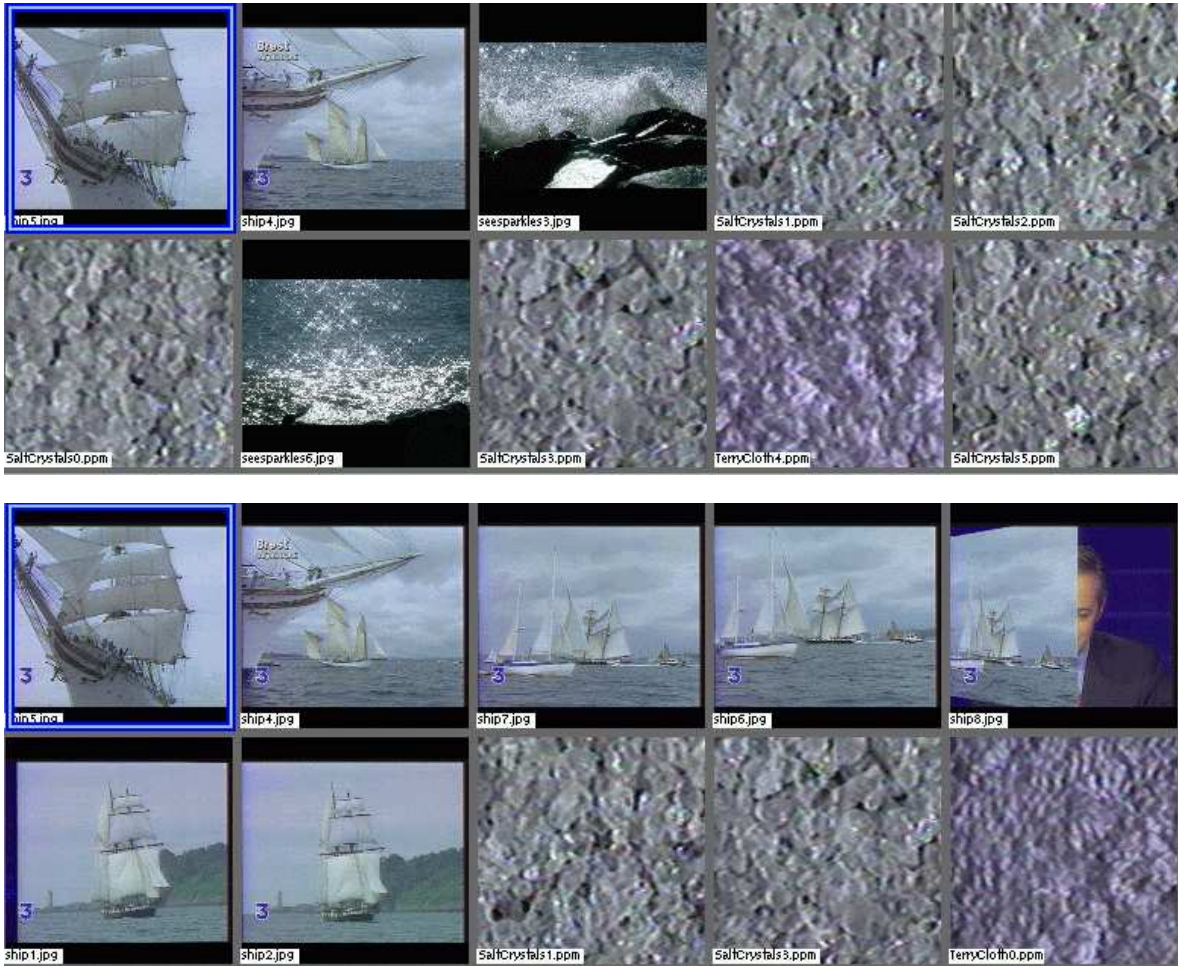


FIG. 4 – Résultats de l'indexation; pour une même image requête (en haut, à gauche, contournée en bleu) en utilisant l'espace de couleur RVB et la métrique L_1 . Bloc d'images de haut: indexation par l'histogramme usuel. Bloc d'images en bas: indexation par l'histogramme pondéré par l'entropie. Les cadres-clés ont été gracieusement fournis par l'INA - Institut National de l'Audiovisuel.

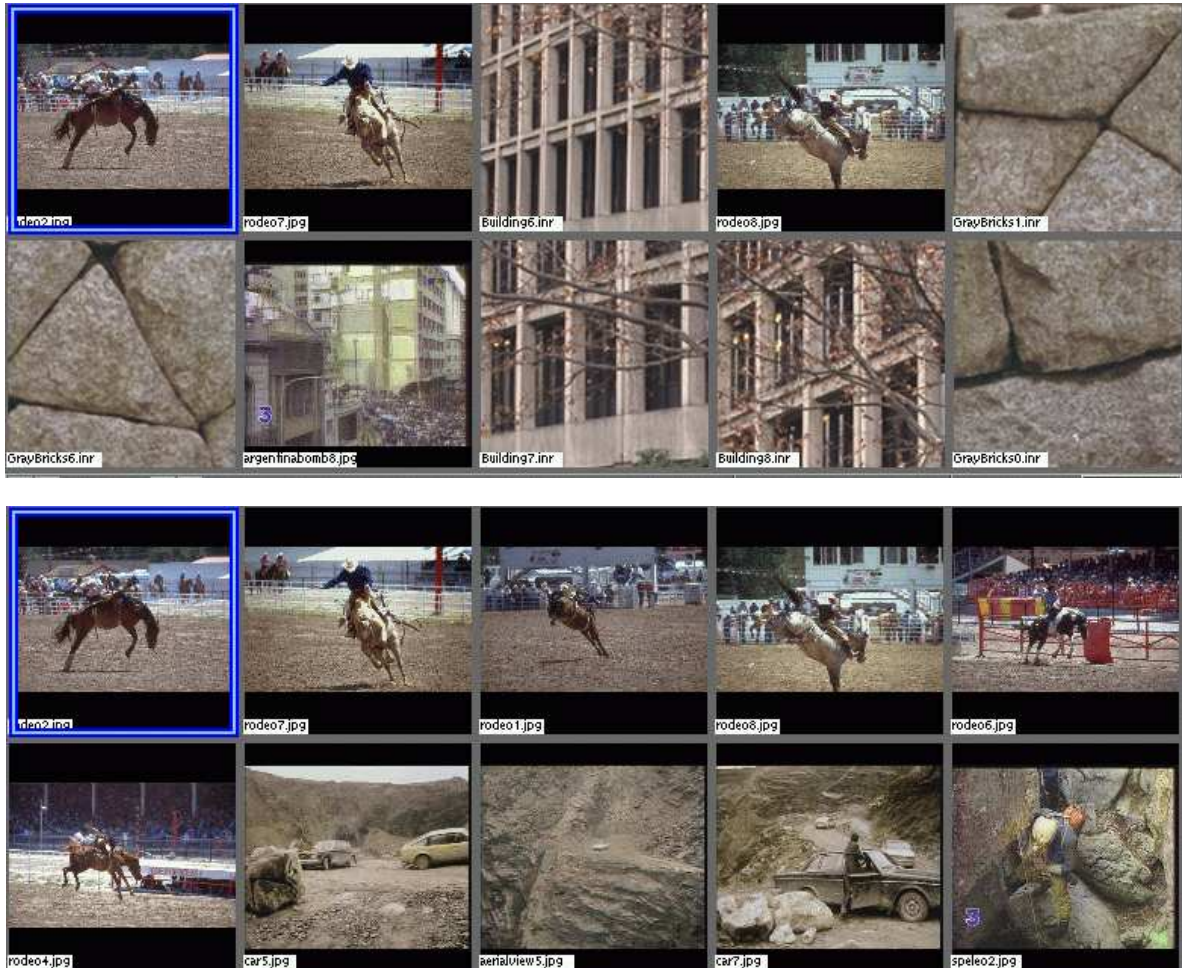


FIG. 5 – Résultats de l'indexation; pour une même image requête (en haut, à gauche, entourée en bleu) en utilisant l'espace de couleur RVB et la métrique L_1 . Bloc d'images de haut : indexation par l'histogramme usuel. Bloc d'images en bas : indexation par l'histogramme pondéré par la probabilité. Les cadres-clés ont été gracieusement fournis par l'INA - Institut National de l'Audiovisuel.

6.1.2 Signature couleur adaptative

Mots clés : couleur, représentation adaptative, Kolmogorov.

Participants : Francois Tonnin, Francois Fleuret, Nozha Boujemaa.

La plupart des signatures couleur (ou même texture) produisent un échantillonnage systématique et régulier de l'espace des descripteurs. L'avantage est que la comparaison des signatures est aisée dans ces conditions. L'inconvénient se traduit parfois par le manque de représentativité du contenu visuel. L'objectif de ce travail est d'explorer un nouveau type de signature couleur procédant par échantillonnage adaptatif de l'espace des descripteurs. Une première solution est donnée par une signature basée sur une approximation de la norme L^∞ entre les fonctions de répartition des couleurs.

L'argument théorique sous-jacent est le même que celui utilisé pour justifier le test de Kolmogorov-Smirnov : la loi de la norme infinie entre deux fonctions de répartition empiriques d'une même loi L ne dépend pas de L .

Durant l'indexation, l'algorithme génère pour chaque image un arbre de représentation de la distribution des couleurs dans l'espace choisi (RGB ou HSV). Chaque noeud interne de cet arbre correspond à la couleur moyenne des pixels représentés dans le sous-arbre correspondant.

L'évaluation de la distance se fait en parcourant cet arbre et en calculant online les fonctions de répartition, et l'écart maximum qui les sépare. Ce travail constitue une étude préliminaire. La méthode dans son état actuel n'est pas exploitable vu le temps de calcul important.

6.2 Descripteurs d'images : Bases spécifiques

6.2.1 Indexation de bases d'empreintes digitales

Mots clés : empreintes digitales, signatures spécifiques, matching de points, classification.

Participants : Sylvain Bernard, Nozha Boujemaa.

Il s'agit d'un travail développé dans le cadre de la 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 états (Gabon, Kenya, Pologne ...), cette société doit gérer de très grandes bases d'images (empreintes, visages).

Nous exploitons un grand nombre de connaissances propres au domaine de la reconnaissance manuelle de personnes par empreintes digitales. Pour savoir si deux empreintes proviennent du même doigt, les experts détectent leur ensemble de minuties (points de fin ou bifurcations de lignes de crêtes) [4]. Après leur mise en correspondance et si le nombre de minuties communs est suffisant, il est alors affirmé que les deux empreintes proviennent du même doigt. Dans le cas de la recherche dans de grandes bases d'empreintes, une classification préalable de celles-ci est alors nécessaire afin de permettre un gain de temps considérable [4].

Cette année, nous avons axé nos recherches sur la détection automatique des minuties et le matching de minuties.

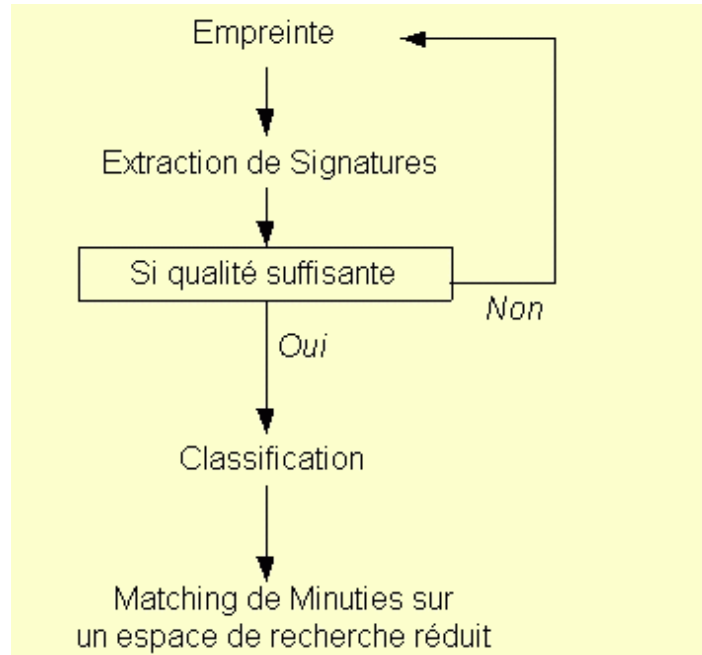


FIG. 6 – Schéma global du système pour la recherche d'une empreinte dans la base.

La détection automatique des minuties Pour des questions de robustesse, nous ne détectons pas directement les minuties, mais nous procédons en 3 étapes [4] :

- 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 structure périodique et orientée, que nous modélisons par une ondelette de Gabor. Une image améliorée est ainsi obtenue par un filtrage multi-échelle.
- 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 squelettisation des lignes de crêtes nous permet une détection efficace et rapide de ses fins de lignes et bifurcations.

La segmentation de l'image permet non seulement de détecter les lignes de crêtes et les vallées mais aussi de délimiter les zones bruitées de l'empreinte et le fond de l'image. Le rapport $\frac{\text{surface des zones bruitées}}{\text{surface empreinte}}$ donne une note sur la qualité globale de l'empreinte, ce qui nous permet de rejeter les empreintes de mauvaise qualité en entrée du système. Après une importante phase d'optimisation de code, le temps nécessaire à la détection des minuties a été divisé par 5.

Le matching de minuties Il doit tenir compte du fait qu'entre deux jeux de minuties issus du même doigt, nous trouvons en général :

- une transformation globale (translation, rotation),

- des déformations locales,
- la présence de minuties dans le premier jeu et leur absence dans le second (et inversement).

Nous avons mis au point un algorithme de matching fondé sur la transformée de Hough. Notre matching tient compte de la relation spatiale des minuties entre elles. Cet algorithme permet un recalage plus robuste des deux jeux de minuties par détermination de la déformation globale qui les lie. L'algorithme de Matching donne des résultats très encourageants avec de faibles taux d'erreurs (FAR et FRR). Nous espérons pouvoir l'améliorer à la fois en précision, en tenant compte notamment des déformations locales, et en temps.

Les algorithmes de détection des minuties et de matching seront très prochainement intégrés aux produits THOMSON-CSF IDMATICS.

6.2.2 Indexation d'une base de visages

Mots clés : Eigenfaces, programmation dynamique, Apprentissage statistique.

Participants : Hichem Sahbi, Nozha Boujemaa.

Résumé : *Dans le cadre de l'indexation des bases d'images et plus particulièrement des bases de données contenant des visages, nous modélisons la distribution des visages dans des espaces à n dimensions en utilisant la transformée PCA connue sous le nom de "eigenfaces". Cette signature à la fois très efficace pour capturer les variations inter- et intra- personne(s) présente une forte sensibilité aux variations d'intensité, à la pose 2D ou 3D, et nécessite un très bon recalage de l'information. D'une manière générale, une phase d'extraction préalable des visages dans une image est nécessaire (cf. 6.4). Nous procédons ensuite au recalage (features matching) des visages en utilisant la programmation dynamique. La signature d'ordre n associée à chaque visage représente une multi-distance d'un visage recalé sur n prototypes bien choisis.*

Une première phase consiste à recalcr les visages qui ont été extraits par une méthode de segmentation "coarse to fine" présentée en 6.4.

Dans cette étape nous supposons que les visages extraits sont normalisés en intensité et en échelle, le but est de réduire les effets de distorsion des visages afin de les recalcr entre eux. L'approche de recalage adoptée est basée sur une extraction systématique des imageries pour chaque visage de la BD et un recalage de ces imageries sur des prototypes (on peut prendre comme prototypes les n visages les plus frontaux de la base, tels que deux images quelconques ne se rapportent pas au même individu). Ce recalage est effectué en utilisant une adaptation de l'algorithme de programmation dynamique ^[Ney84]. Cette approche est applicable pour un recalage d'informations incomplètes d'où la possibilité de définir des signatures pour des visages partiellement occultés (lunettes, etc.).

[Ney84] H. NEY, «The Use of a One-Stage Dynamic Programming Algorithm for Connected Word Recognition», *IEEE Trans. ASSP* 32, 2, 1984, p. 1042-1062.

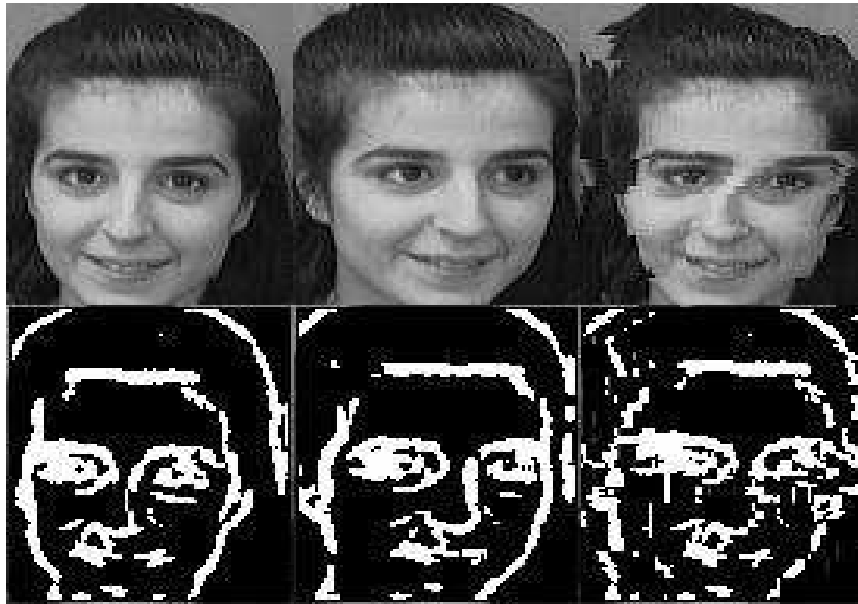
La phase de reconnaissance des visages passe généralement par un schéma d'indexation. Dans notre cas, pour indexer les visages détectés, on applique les étapes suivantes :

1. Construire les n prototypes visages P_j , $i = 1..n$ les plus frontaux.
2. Pour chaque visage V_i ayant un index T de taille n :
 - (a) Effectuer un recalage entre V_i et chaque prototype P_j ;
 - (b) Positionner $T_j = distance(V_i, P_j)$.



(A)

(B)



(C)

FIG. 7 – Recalage en utilisant (A) Les images contours versus (B) Les images de niveau de gris. (C) Le recalage de deux images de contours fournit des résultats meilleurs par rapport aux images de niveau de gris.

Ainsi pour chaque visage, un index d'ordre n est construit, la distance choisie entre les deux visages recalés V_i et P_j représente la somme des distances euclidiennes entre les projections

des imaugettes recalées dans un espace propre construit par apprentissage sur la base d'Olivetti, cette distance s'exprime par :

$$Trans(V_i, P_j) = \sum_k^m \sum_l^m \alpha_{kl} d_{kl}(\Phi(V_{ik}), \Phi(P_{jl})) \quad (6)$$

m est le nombre d'images extraites dans chaque visage, Φ est une application qui fournit pour une image sa projection dans l'espace propre, et α_{kl} est appelé "facteur de matching" (ou coefficient de Kronecker) et indique si deux images V_{ik} et P_{jl} d'un visage V_i et d'un prototype P_j sont mises en correspondance ou non. La métrique d_{kl} est liée à α_{kl} comme décrit ci dessous :

$$\alpha_{kl} = \begin{cases} 0 & \text{si } d_{kl} \text{ est une métrique d'insertion ou de suppression.} \\ 1 & \text{si } d_{kl} \text{ est une métrique de substitution.} \end{cases}$$

6.3 Généralisation de la classification compétitive

Mots clés : classification non-supervisée, régularisation entropique, optimisation.

Participant : Nozha Boujema.

Une des questions les plus délicates dans un problème de classification est celui du nombre optimal de classes. La plupart des algorithmes de partitionnement supposent connu cette information a priori. Dans le cadre de l'indexation et de recherche d'images par le contenu, les méthodes de classification sont utiles pour la résolution de problèmes de natures différentes. Nous citons par exemple :

- la segmentation d'images, par classification dans l'espace des attributs visuels, pour générer des masques pertinents pour les **requêtes partielles**. Dans ce cas, le nombre de classes correspond au nombre de couleurs et/ou textures pertinentes dans l'image,
- la catégorisation d'une base d'images permettant d'offrir une **navigation** intelligente. Dans ce cas, le nombre de classes correspond au nombre de catégories d'objets (voitures, coupes, etc.) ou de thèmes (tableaux de peinture, scènes de villes, paysages, images de sport, etc.) représentés dans la base.

Dans le contexte des grandes bases d'images, il n'est pas viable de pouvoir disposer de cette information pour chaque image ou pour chaque base. Des méthodes non-supervisées permettant la détermination automatique du nombre de classes optimal sont nécessaires. Dans cet objectif, nous avons développé de nouvelles méthodes de classification non-exclusives [1] procédant par compétition agglomérative et les avons situées dans un cadre générique.

Nous avons proposé une nouvelle famille de fonctions objectives. Le principe repose sur la combinaison de critères à effets antagonistes traduisant la fidélité aux données (J1) d'une part et la réduction de complexité de l'espace des partitions (J2) d'autre part, et qui s'exprime plus généralement de la manière suivante :

$$J = J_1 + \alpha J_2 \quad (7)$$

Nous avons proposé et considéré les deux cas particuliers suivants :

$$J = \sum_{i=1}^C \sum_{j=1}^N u_{i,j}^2 d_{i,j}^2 - \frac{\alpha}{N} \sum_{i=1}^C \sum_{j=1}^N u_{i,j} \log(u_{i,j}) \quad \text{avec} \quad \sum_{i=1}^C u_{i,j} = 1 \quad (8)$$

$$J = \sum_{i=1}^C \sum_{j=1}^N u_{i,j}^2 d_{i,j}^2 - \alpha \sum_{i=1}^C \sum_{j=1}^N p_i \log(p_i) \quad \text{avec} \quad \sum_{i=1}^C u_{i,j} = 1 \quad \text{et} \quad p_i = \frac{1}{N} \sum_{j=1}^N u_{i,j} = \frac{N_j}{N} \quad (9)$$

Les termes de réduction de complexité possibles (ici régularisation entropique) ne sont autres que des critères de validité, déjà proposés dans la littérature [Bez81], dont l'utilisation devient plus intéressante dans ce nouveau modèle. L'explication détaillée se trouve dans [1]. Nous avons montré ([6], [7]) que les solutions optimales s'expriment de la manière suivante :

$$u_{i,j} = u_{i,j}^{fcm} + u_{i,j}^{biais} \quad \text{avec} \quad u_{i,j}^{fcm} = \frac{\frac{1}{d_{i,j}^2}}{\sum_{s=1}^C \frac{1}{d_{s,j}^2}} \quad (10)$$

et, pour la fonction (1) ,

$$u_{i,j}^{biais} = \frac{1}{d_{i,j}^2} \frac{\alpha}{2N} \left(\log(u_{i,j}) - \frac{\sum_{s=1}^C \frac{\log(u_{s,j})}{d_{s,j}^2}}{\sum_{s=1}^C \frac{1}{d_{s,j}^2}} \right) \quad (11)$$

pour la fonction (2) ,

$$u_{i,j}^{biais} = \frac{1}{d_{i,j}^2} \frac{\alpha}{2N} \left(\log(p_i) - \frac{\sum_{s=1}^C \frac{\log(p_s)}{d_{s,j}^2}}{\sum_{s=1}^C \frac{1}{d_{s,j}^2}} \right) \quad \text{avec} \quad p_i = \frac{N_j}{N}, \quad p_s = \frac{N_s}{N} \quad (12)$$

On note que, dans les deux cas, u^{biais} s'exprime en fonction de la différence des paramètres de la classe courante et d'une moyenne pondérée des paramètres des autres classes. Le signe de u^{biais} affaiblit ou renforce la population de la classe selon le cas. Ce mécanisme est à l'origine de la compétition entre les classes pour attirer les données de l'espace. Les classes de faible cardinalité disparaissent. Le nombre optimal est obtenu, à la convergence, après équilibre entre ces deux termes antagonistes.

Ce travail fait l'objet d'un chapitre en cours de rédaction sur les méthodes de classification avec Rajesh Davé (New Jersey Institute of Technology)

6.4 Requêtes partielles

6.4.1 Requêtes par régions d'intérêt

Mots clés : Graphe d'Adjacence de Régions (R.A.G.), composantes connexes,

[Bez81] J. C. BEZDEK, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, 1981.

segmentation, classification, requête partielle.

Participants : Julien Fauqueur, Nozha Boujemaa.

Ce thème touche plusieurs contextes applicatifs. Il est abordé dans la thèse de Julien Fauqueur qui s'inscrit dans le cadre de la collaboration Priamm "MediaWorks" (cf. 7.1)

Lors d'une recherche d'images dans une base, il est courant que l'utilisateur ne s'intéresse qu'à une région particulière de l'image de référence. Il souhaite retrouver toutes les images comportant une région similaire, ou éventuellement un agencement spatial de régions similaires. Pour le système d'indexation, cela se traduit par une évaluation locale des signatures sur les zones d'intérêt.

Effectuer une requête globale à partir de toute l'image suppose d'intégrer les signatures visuelles du fond de l'image dans la requête. Celui-ci représente du bruit par rapport à la région d'intérêt de l'utilisateur. Ainsi le résultat retrouvé, après le calcul des similarités globales, ne conviendra pas à l'intention de l'utilisateur, étant donné qu'il prendra en compte l'apparence visuelle totale de l'image.

Une fois la zone d'intérêt identifiée, un masque est créé afin d'évaluer les signatures uniquement sur la partie de l'image sélectionnée.

La méthode la plus évidente consiste à détourner manuellement les régions jugées pertinentes. Cette idée est écartée lorsqu'il s'agit d'effectuer ce traitement sur les milliers d'images d'une base.

La méthode la plus simple consiste à effectuer une subdivision systématique de l'image : par exemple en une grille de 16 rectangles [5]. Mais, étant indépendante du contenu de l'image, cette subdivision ne permet pas d'offrir une précision de recherche satisfaisante : elle constitue seulement une première approche de sélection de partie d'images.

Nous avons retenu la méthode qui consiste à effectuer une *segmentation automatique non-supervisée* des images. Les masques sont alors définis par les régions de l'image dont les pixels voisins sont visuellement homogènes. La segmentation est opérée par classification (agglomération compétitive) des pixels de l'image dans l'espace des attributs.

Une plateforme de test a été développée afin de mettre en oeuvre divers algorithmes de segmentation par classification et différentes méthodes de fusion de régions couleurs adjacentes. Ces algorithmes sont en cours d'évaluation.

Pour représenter toutes ces informations, on génère le Graphe d'Adjacence de Régions de l'image, dit R.A.G. pour *Region Adjacency Graph*.

Chaque *noeud* du graphe représente une *région* et stocke les informations relatives aux attributs visuels et géométriques.

Chaque *arête* représente la *relation entre deux régions adjacentes* et stocke les attributs associés : adjacence, disposition relative, contraste, ...

Le R.A.G. permet donc de représenter les informations relatives au contenu de chaque image de la base dans une structure logicielle manipulable par le système d'indexation.

La mesure de similarité entre deux images est alors définie comme le *score de mise en correspondance entre les deux graphes* à partir des attributs des noeuds et arêtes. Le RAG-requête peut représenter une *requête partielle simple* ou bien une *requête partielle multiple*.

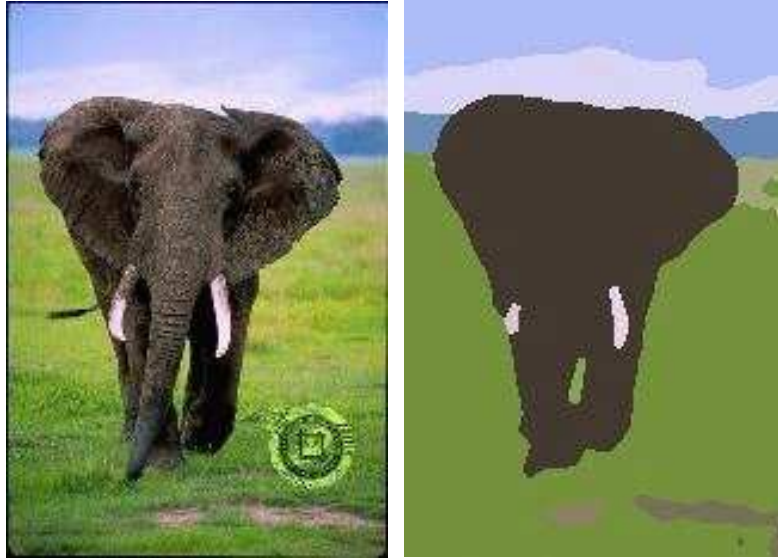


FIG. 8 – *Image originale / Image segmentée.*

Nos travaux récents ont porté sur le développement du mode de requête partielle simple. Pour une image donnée, nous avons l'ensemble des opérations suivantes :

1. Segmentation de l'image.
2. Détermination des composantes connexes et contours de l'image segmentée.
3. Création du RAG à partir des composantes connexes qui constituent les régions de l'image, donc les noeuds du graphe.
4. Restriction du calcul des signatures d'images sur chaque composante connexe.
5. Restriction du RAG aux régions d'intérêt : l'utilisateur doit pouvoir formuler sa requête à partir d'un sous-graphe du RAG. Il est donc possible d'extraire un sous-graphe à partir de la structure du RAG.
6. Mise en place de la procédure de comparaison locale.

Note : Sur la base de critères de couleur et/ou topologiques, les opérations de fusion des noeuds du RAG permettent une composition différente de la segmentation.

La capture d'écran ci-dessous (cf. Fig.9, premier bloc d'images) montre le résultat d'une requête partielle simple sur la région visage de la première image (imagerie dans le cadre bleu) (notons ici que la région visage n'est pas identifiée en tant que telle mais représente une région générique). Les images suivantes présentent les régions (identifiées par leur contour magenta) dont les signatures locales ont été jugées proches de celle de la région requête.

Les régions retournées correspondent également à des visages et, en particulier, les deux premières représentent le même visage avec des arrière-plans différents de la requête. Il est

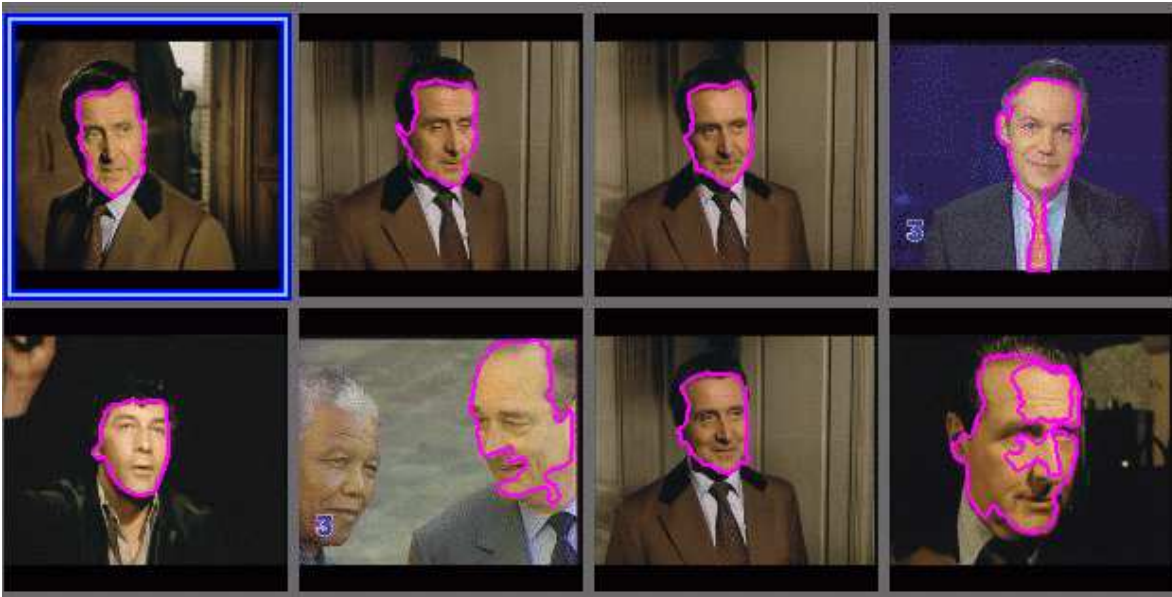


FIG. 9 – Résultat d'une requête partielle simple / résultat d'une requête globale dans les mêmes conditions.

intéressant de noter que l'image du présentateur du journal télévisé et celle de Monsieur Chirac comportent un fond visuellement différent du fond de la première image. Ceci met en valeur la pertinence de l'évaluation des signatures uniquement sur la région d'intérêt, à savoir le visage.

Plusieurs développements sont encore nécessaires, en particulier celui de l'adaptation des signatures globales au contexte des requêtes par régions d'intérêt avant d'investiguer les requêtes partielles multiples.

Le mode de requête partielle que nous venons de décrire est basé sur une subdivision de l'image est appelé *requête partielle par l'exemple*, car les régions proviennent d'images réelles. Un autre mode interactif important est celui des *requêtes partielles par ébauche graphique*.

Le scénario de requête partielle présenté dans les sections précédentes suppose que l'utilisateur dispose d'une image illustrant l'objet (ou plus généralement la partie) qu'il recherche.

Les *requêtes partielles par ébauche graphique* offrent la possibilité à l'utilisateur de construire interactivement sa requête en dessinant un agencement de régions correspondant à l'idée qu'il se fait de l'image requête.

L'interface graphique présentée dans cette section (figure 10) a été développée en Java par **Moez Kamoun** lors de son stage au sein de l'équipe. Il s'agit d'un éditeur graphique qui permet à l'utilisateur de dessiner plusieurs régions de couleurs ou de textures prédéfinies. Les régions sont tracées librement, remplies par la couleur/texte sélectionnée. Leurs contours sont fermés automatiquement. La démarche globale, mise à part l'expression de la requête utilisateur, reste sensiblement la même que celle des requêtes par régions exemples. Nous essayerons de résoudre les spécificités de ce mode dans nos travaux futurs.



FIG. 10 – Interface de requête graphique.

6.4.2 Détection et localisation de visages

Mots clés : détection de visages, apprentissage statistique, représentation hiérarchique.

Participants : François Fleuret, Donald Geman.

Nous avons développé un algorithme rapide de détection de visages. Étant donnée l'image en noir et blanc d'une scène dans laquelle sont présents des visages vus de face, ce logiciel est ainsi capable de produire la liste de tous les visages présents, et d'estimer pour chacun d'entre eux la position des yeux et de la bouche (cf. figure 11).

Notre approche s'appuie sur deux aspects fondamentaux qui sont rarement combinés : l'apprentissage statistique et les représentations hiérarchiques.

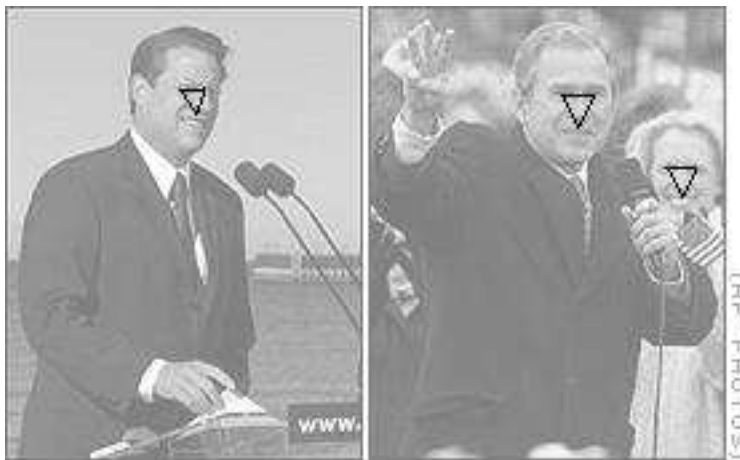


FIG. 11 – Exemple de détections de visages.

Lors d'une période d'apprentissage, nous construisons une hiérarchie de détecteurs, dédiés à des ensembles de visages plus ou moins génériques. Un détecteur ne sera utilisé à un endroit de l'image que si tous les détecteurs plus grossiers ont répondu positivement. Cette approche permet d'une part de limiter au maximum le coût algorithmique (la plupart du temps les détecteurs les plus grossiers sont capables de rejeter un emplacement de l'image, par exemple s'il s'agit d'une zone uniforme - cf. figure 12), et de surcroît d'atteindre un taux de fausses alarmes très faible (pour que l'algorithme produise une alarme, il faut que tous les détecteurs d'une séquence fassent de même).

Chacun des détecteurs a la forme d'un comptage de test booléens locaux. L'apprentissage se réduit donc à déterminer quels sont les tests locaux pertinents. Cette forme de détection par "comptage" est biologiquement vraisemblable, très robuste, et permet donc d'éviter les problèmes classiques de sur-apprentissage.

Enfin, des post-traitements éliminent les fausses alarmes isolées et estiment les poses finement en moyennant les estimations réalisées par les détecteurs individuellement. [2, 3, 8]



FIG. 12 – *Intensité d’usage des pixels lors du traitement de la scène de la figure 11. Les zones sombres indiquent où un grand nombre de détecteurs a été utilisé.*

6.4.3 Segmentation des Visages

Mots clés : Segmentation couleur, Apprentissage statistique, Modélisation floue, Classifieurs Bayésiens, Modèles structurels..

Participants : Hichem Sahbi, Nozha Boujemaa.

Nous avons développé une méthode de détection fine des régions visages [9], en partant d’une segmentation couleur grossière de l’image en entrée (Fig.13.(A,C)). Parmi les segments obtenus, on distingue ceux qui ont une distribution couleur caractérisant la peau, ces segments sont sélectionnés en recherchant ceux qui minimisent une erreur donnée par :

$$E(R_i) = \frac{1}{\|R_i\|} \sum_{(x,y) \in R_i} (\Phi(c_{(x,y)}) - c_{(x,y)})^2 \quad (13)$$

Où $c(x, y)$ représente la couleur d’un pixel, et Φ est la sortie d’un réseau de neurones entraîné en utilisant une grande population de couleurs de peau provenant de la base de données Voila et du Web. Dans ce cas l’architecture du réseau permet d’effectuer pour chaque couleur de la peau une compression/décompression (Non Linear PCA) afin de capturer non-linéairement la distribution statistique de la couleur apprise.

Les segments peau détectés serviront pour un réapprentissage de la couleur de la peau dépendant du contexte de l’image. Cette étape est effectuée en définissant la variable aléatoire $Y(c(x,y))$ (qui vaut 1 si la couleur du pixel (x,y) est une couleur de peau, 0 sinon). Dans ce cas, on modélise un classifieur Bayésien pour filtrer les pixels ayant une couleur de peau (Fig.13.(E)), cette modélisation suppose que la distribution de la couleur pour chaque région peau est gaussienne. Afin d’éliminer les pixels outliers susceptibles de fausser l’estimation des paramètres (moyenne et matrice de covariance), nous utilisons un processus de débruitage qui

suppose que la distribution de chaque région peau est divisée en deux clusters, un cluster peau non bruité et un cluster bruité. Par la suite, on utilise l'approche floue [Dav91] pour affecter chaque pixel au cluster qui lui est associé, ainsi les paramètres de la distribution gaussienne de chaque région peau sont donnés par :

$$\mu_i = \frac{\sum_{k=1}^{n_i} h_k U_{skin,k} x_k}{\sum_{k=1}^{n_i} h_k U_{skin,k}}, \quad (14)$$

$$\text{Sigma}_i = \frac{\sum_{k=1}^{n_i} h_k U_{skin,k} (x_k - \mu_i)(x_k - \mu_i)^T}{\sum_{k=1}^{n_i} h_k U_{skin,k}} \quad (15)$$

Où $U_{skin,k}$ caractérise l'appartenance d'une couleur x_k au cluster peau. Dans l'expression ci dessus, ces coefficients sont des facteurs de pondération permettant de réduire l'effet des pixels bruits lors de l'estimation des paramètres de la distribution Gaussienne.

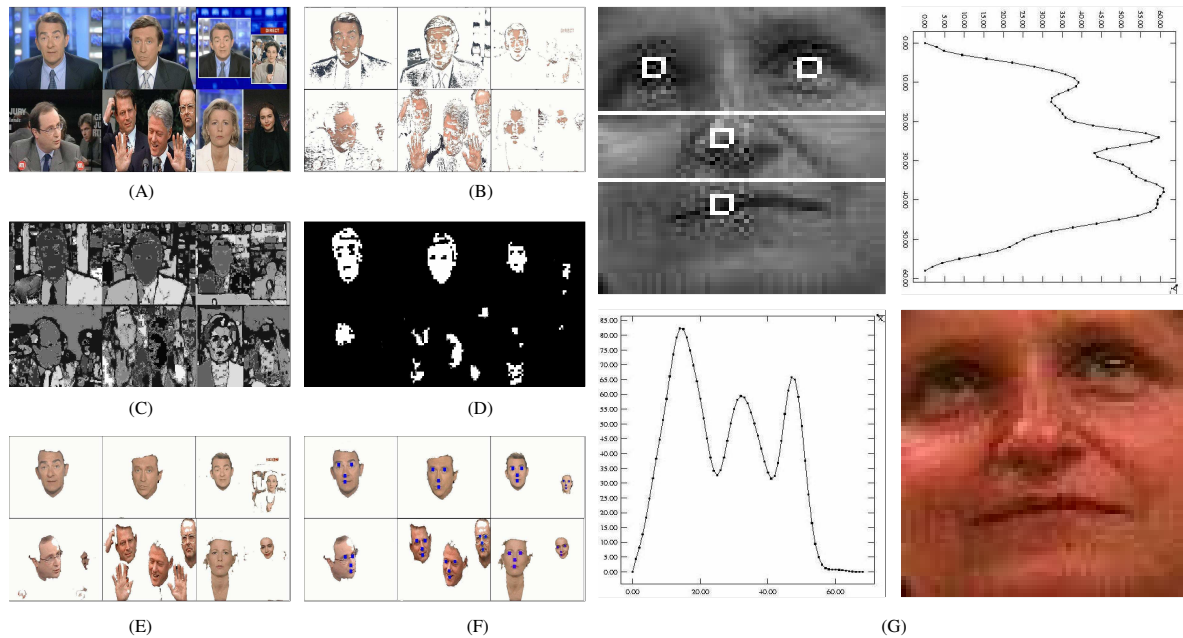


FIG. 13 – (A) L'image en entrée. (C) Segmentation en utilisant l'approche DFDM. (D) Sélection des régions peau. (B,E) Détection de la peau en utilisant respectivement un ANN et l'approche "coarse to fine". (F,G) Détection des visages frontaux en utilisant un modèle des caractéristiques visages.

Une analyse en composantes connexes est ensuite opérée pour les nouvelles régions peau détectées. Pour décider si une région peau est un visage ou non, nous effectuons une projection des intensités d'une région sur les axes vertical et horizontal (Fig.13.(F,G)). Pour chaque axe

[Dav91] R. N. DAVE, « Characterization and detection of noise in clustering », *Pattern recognition* 12, 11, 1991, p. 657–664.

on recherche les modes de variation principale (en utilisant l'analyse de Fourier, ou un filtrage progressif). Ces modes caractérisent les coordonnées horizontales et verticales des yeux, nez et la bouche données respectivement par (x_1, y_1) , (x_3, y_1) , (x_2, y_2) et (x_2, y_3) et l'hypothèse "visage" sur une région peau est vérifiée en utilisant un modèle d'apprentissage. Dans notre cas un mélange de quatre gaussiennes est utilisé où chaque cluster capture la distribution statistique des coordonnées (x, y) du feature qui lui est associé. Une décision est effectuée en utilisant une approche par maximum de vraisemblance.

6.5 Recherche interactive

6.5.1 Modèles de questions/réponses pour une recherche interactive

Mots clés : Contrôle de pertinence, recherche interactive, inférence bayésienne, réduction d'entropie..

Participants : Roland Moquet, Donald Geman.

L'utilisateur recherche une certaine image, et répond à une série de questions posées par un programme qui montre successivement un échantillon d'images de la base, en tentant d'afficher le plus rapidement possible l'image recherchée. Par exemple, à chaque étape, l'utilisateur indique laquelle parmi deux images présentées est la plus proche de ce qu'il recherche. Bien sûr, la réponse à ce type de question est subjective, et l'interaction utilisateur/programme présente une composante stochastique inhérente.

Nous avons étudié les modèles statistiques gouvernant les "questions" et les "réponses" à travers un cadre bayésien. Chaque nouvelle question (les deux images affichées) est choisie pour minimiser l'espérance de l'entropie conditionnelle de la distribution de probabilité sur toutes les images de la base, sachant les réponses données par l'utilisateur. Les réponses sont déterminées par une séquence de métriques indépendantes (sur l'espace vectoriel constitué par les attributs de l'objet), dont la distribution peut dépendre à la fois de la question et de l'image cible. La performance de l'algorithme est mesurée par le nombre moyen de questions nécessaires pour retrouver l'image cible.

Nous étudions le compromis fondamental entre la synchronisation du modèle utilisé (cohérence avec le comportement humain), et l'entropie du modèle (précision du modèle, et donc taux de transfert d'information). L'algorithme résultant est expérimenté par des recherches dans une base de polygones caractérisés par leur couleur (niveau de gris), leur taille, leur nombre de côtés, et leur aplatissement, et ses performances sont comparées avec les bornes d'efficacité théoriques et des modèles précédents. En conclusion, nous discutons les extensions envisageables à un engin de recherche totalement opérationnel.

Les figures ci-jointes (14 et 15) illustrent les résultats obtenus. La figure 14 montre la corrélation entre l'entropie de la réponse et l'efficacité de la recherche. Plusieurs modèles ont été testés grâce à des recherches simulées par le système. La figure 15 met en valeur l'importance de la synchronisation entre le modèle utilisé par l'algorithme et la loi suivie par les réponses réelles. La différence entre les deux lois est estimée sur un échantillon de questions : il s'agit de la synchronisation entre le modèle et les réponses réelles. La figure 14 montre la corrélation entre cette synchronisation et l'efficacité de la recherche. Dans chaque cas, l'efficacité est calculée

par le nombre moyen de questions nécessaire pour trouver l'image cible dans une base de 200 images.

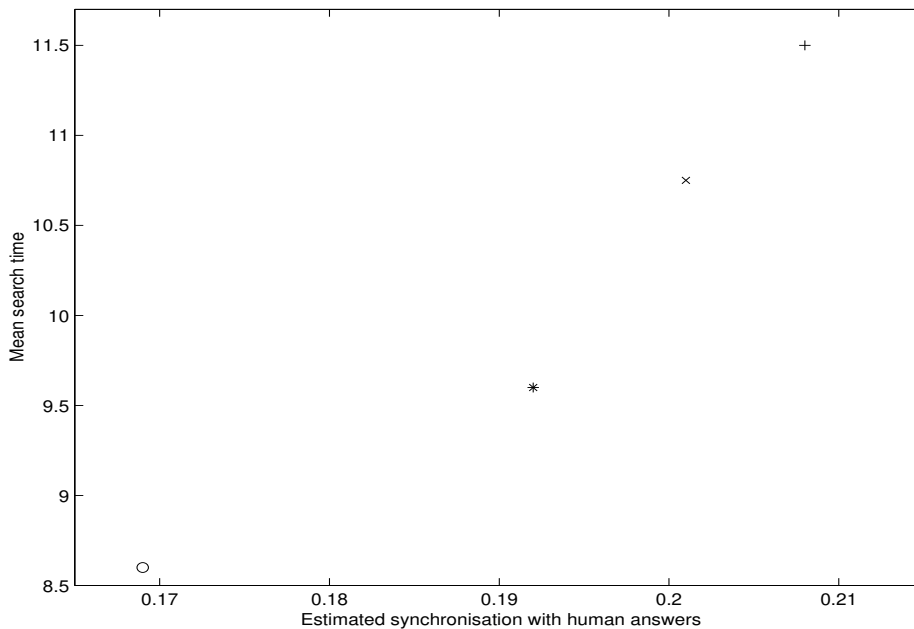


FIG. 14 – Relation entre la synchronisation et la performance pour des recherches réelles.

6.5.2 Navigation

Mots clés : classification, combinaison des signatures.

Participants : Bertrand Le Saux, Nozha Boujema.

Notre but est de classer les bases d'images afin d'offrir à l'utilisateur non pas des images prises au hasard pour commencer, mais les images les plus représentatives de la base.

- Une évaluation des signatures existantes a été entreprise afin de déterminer les plus discriminantes d'entre elles pour former notre espace de représentation. En outre, les signatures sont choisies pour apporter des informations complémentaires.
- Une image est représentée par un index (vecteur) dans un espace de signatures. Il s'agit d'abord de réduire la dimensionalité de l'espace des signatures : pour cela on utilise soit une analyse en composantes principales, soit la technique fastmap ^[FL95].
- Ensuite on doit combiner les signatures entre elles. Les moyens de combiner les signatures choisies ont été testés. Une normalisation basée sur les distances a été retenue (voir comparaisons des sets de signatures combinées : 16).

[FL95] C. FALOUTSOS, K. LIN, «FastMap: A Fast Algorithm for Indexing, Data-Mining and Visualization of Traditional and Multimedia Datasets», in: *Proc. of 8th ACM SIGMOD*, 1995.

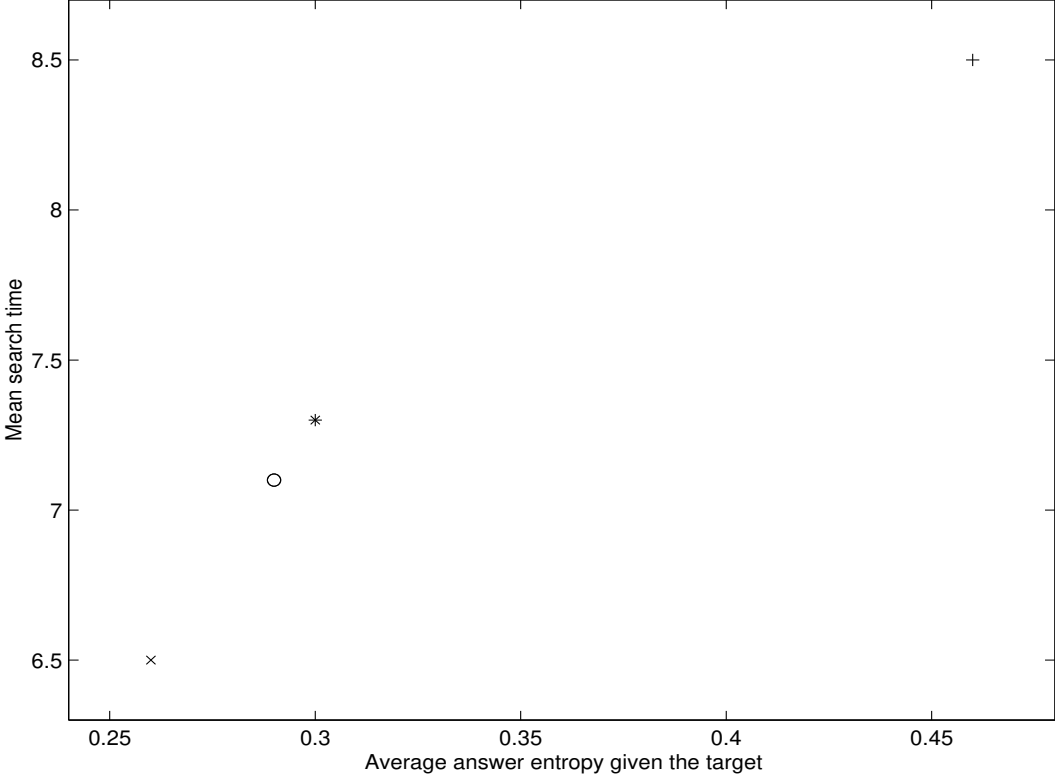


FIG. 15 – Relation entre le déterminisme du modèle utilisé et la performance pour des recherches simulées.

- Différents algorithmes de classification ont été testés sur les signatures choisies afin de classifier une base d'images : FCM [Bez81] avec distance euclidienne ou distance de Mahalanobis comme référence, algorithme de Gath et Geva [Bez81], EM [DLR77], et surtout l'Agglomération Compétitive [FK97], avec distance euclidienne ou de Mahalanobis. Les tests sont effectués sur des bases avec vérité-terrain, comme par exemple la base Columbia (objets photographiés sous différents angles de prise de vue) qui comprend 20 classes avec 72 représentants par classe. Les bases d'images comportent de nombreuses difficultés pour le clustering : les classes naturelles sont de formes variées, dans le meilleur des cas hyper-ellipsoïdales, mais souvent plus complexes; elles présentent un phénomène de recouvrement important; enfin elles sont de densités et d'étalements variés (voir figure : 17).

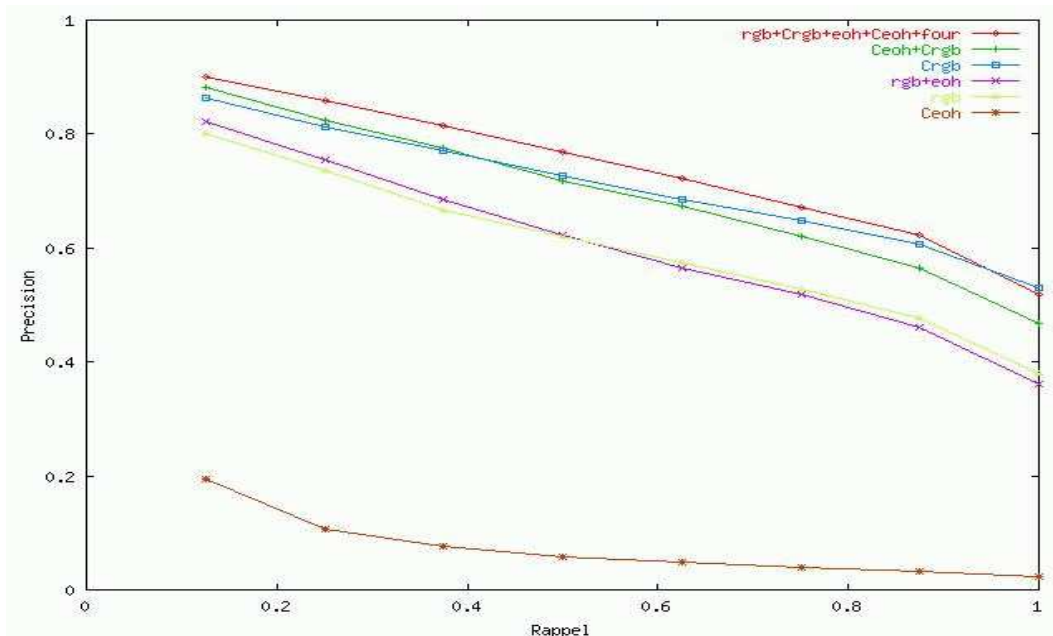


FIG. 16 – Courbes précision rappel pour des requêtes dans la base de tests généraliste, obtenues avec des combinaisons de signatures (histogrammes et corrélogrammes) de couleur et d'orientation de contours. Cinq signatures sont retenues afin de constituer un ensemble qui permet d'obtenir la courbe optimale.

-
- [Bez81] J. C. BEZDEK, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, 1981.
- [DLR77] A. DEMPSTER, N. LAIRD, D. RUBIN, « Maximum likelihood from incomplete data via the EM algorithm », *39 (Series B)*, p. 1–38, 1977.
- [FK97] H. FRIGUI, R. KRISHNAPURAM, « Clustering by competitive agglomeration », *Pattern Recognition* *30*, 1997, p. 1109–1119.

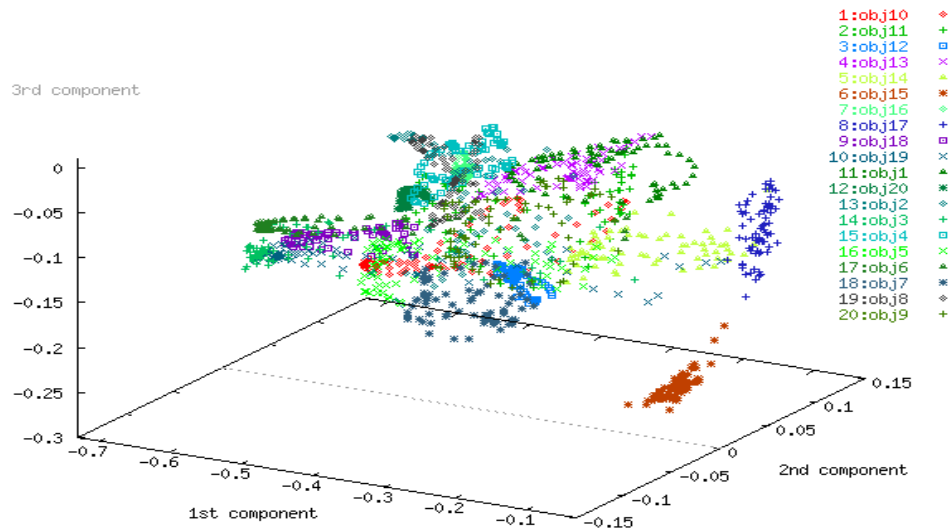


FIG. 17 – Classes originales de la base Columbia, selon la signature en niveau de gris, après projection sur les 3 composantes principales.

6.6 Indexation multimédia

6.6.1 Estimation automatique de mots-clés

Mots clés : image et texte, statistiques, similarités visuelles.

Participants : François Fleuret, Nozha Boujemaa.

Ce travail s'inscrit dans le cadre du projet PRIAMM "MédiaWorks" (cf. 7.1). L'objectif est d'étudier différents scénarios d'indexation et de recherche bi-modale image et texte dans les archives audiovisuelles des journaux télévisés (20H) de TF1. Dans cette partie nous considérons des signatures caractérisant l'apparence visuelle calculée sur toute l'image. **Certaines** images de la base possèdent, **individuellement**, des annotations textuelles. Un scénario de coopération entre l'indexation visuelle et l'indexation textuelle consiste à propager (suggérer) des mots clés à partir des images annotées vers les images non-annotées. Il s'agit alors de générer (proposer) automatiquement une liste de mots-clés pertinents pour une image en intégrant ses similarités visuelles avec le reste de la base.

Différentes approches sont possibles pour réaliser cette propagation.

L'approche que nous proposons consiste à déterminer dans un premier temps, à l'aide de méthodes classiques, les images visuellement similaires à l'image requête. Ensuite, un système de vote associe à chaque mot-clé un score dépendant de sa fréquence d'apparition dans les images similaires, pondérée par les distances à l'image requête.

Cette procédure génère donc finalement une liste de mots-clés ordonnés par pertinence (cf.

figure 18) qui sont finalement affichés dans le champ "keywords". Les mots ainsi trouvés sont proposés pour l'image qui est encadrée par la sélection de l'utilisateur. L'abréviation "gp" désigne "gros plan", annotation utilisée par les documentalistes.

L'utilisation sous-jacente des techniques classiques d'estimations de la similarité visuelle permet à l'utilisateur de paramétrer finement le modèle qui lui paraît adapté au type d'images qu'il considère (visages, paysages, peintures, etc.)



FIG. 18 – Interface de requêtes par similarités visuelles et d'estimations automatiques de mots-clés pertinents.

7 Contrats industriels (nationaux, européens et internationaux)

7.1 Relations industrielles

Participants : Nozha Boujemaa, Sylvain Bernard, Jean-Paul Chièze, François Fleuret, Hichem Sahbi.

7.1.1 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.

7.1.2 Contrat Thomson Multimedia

Une collaboration a été récemment mise en place selon deux voies. La première s'inscrit dans le cadre d'une thèse portant sur la détection et la reconnaissance de visages. La deuxième concerne l'indexation dans les bases génériques d'une manière générale et leur usage dans les applications domestiques.

7.1.3 Contrat avec LookThatUp

Il concerne le transfert du logiciel Surfimage.

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, Constantin Vertan.

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

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 Action de Recherche Collaborative SIRIUS

Participants : Nozha Boujemaa, Marin Ferecatu.

Cette ARC implique le projet IMEDIA ainsi que le projet Langue et Dialogue à Nancy (Malika Smail). Il s'agit d'étudier l'indexation bi-modale par les mots-clés et par le contenu de l'image. Cette collaboration a été interrompue avec le départ de Sébastien Gilles à LookThatUp. Elle a repris récemment avec le recrutement de Marin Ferecatu.

8.2 Actions européennes

8.2.1 Projet Européen "STOP"

Participants : Nozha Boujemaa, Valérie Gouet.

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.

8.3 Actions internationales

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

Participant : Nozha Boujemaa.

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.

8.4 Visites, et invitations de chercheurs

Raghu Krishnapuram (Colorado School of Mines, États-Unis), Stan Sclaroff (Boston University, États-Unis), Lambert Shomacker (NICI, Pays-Bas) ont visité pour une courte durée et animé des séminaires à IMEDIA.

9 Diffusion de résultats

9.1 Animation de la Communauté scientifique

Nozha Boujemaa :

- Membre du Conseil d'administration de l'AFRIF (Association Française de Reconnaissance et Interprétation des Formes) qui remplace l'AF CET dans son rôle de regroupement de la communauté Image et Reconnaissance des Formes .

- Membres de cinq jurys de thèse en 2000 (LIP6 - univ. Paris6, LISI - univ. Lyon1, LIGI2P - Ecole des Mines d'Alès, LE2i - univ. de Bourgogne, INSERM - Univ. de Tours)
- Membre de comité de programme des conférences : CGIP'2000, CGIV'2001, TAIMA'2001.
- Membre du " noyau dur " du groupe thématique du GDR PRC ISIS sur la couleur.
- Chargée de la rédaction d'un chapitre sur l'indexation couleur dans un ouvrage collectif du GDR.

Séminaires invités de Nozha Boujemaa :

- "Classification compétitive non supervisée, application à l'indexation et à la recherche d'images", 16 novembre 2000, Lip6 - univ. Paris 6.
- "Indexation et recherche d'image par le contenu", 10 mars 2000, LEP - Philips.
- "Les défis de l'indexation d'images" - 29 mars 2000, Club Iliatech (Club des partenaires industriels de l'INRIA Rocquencourt).

9.2 Enseignement

Nozha Boujemaa : Cours sur les méthodes d'indexation partielle à l'Ecole Nationale des Ponts et Chaussées, Marne-La-Vallée, cours à l'école de printemps ETASM'2000 (Politechnica de Roumanie - Bucharest), cours à l'Ecole Supérieure de Communication (Sup'Com -Tunis).

Donald Geman : Professeur de probabilités à l'Ecole Polytechnique, chargé de cours au DEA de l'ENS Cachan.

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

Cours de C++ undergraduate à l'université de Chicago (CS116 30h), cours de vision en graduate (15h).

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 de Méthodologie fonctionnelle et mathématiques appliquées au langage Caml, de TP de Programmation fonctionnelle en langage Caml en DEUG MIAS, et de TP de programmation C en DU Informatique à l'université Paris XI - Orsay.

10 Bibliographie

Thèses et habilitations à diriger des recherches

- [1] N. BOUJEMAA, *"Sur la classification non-exclusive en analyse d'images"*, habilitation à diriger des recherches, Université de Versailles-Saint-Quentin, 2000.
- [2] F. FLEURET, *Détection hiérarchique de visages par apprentissage statistique*, thèse de doctorat, Université Paris-VI, Paris, 2000.

Articles et chapitres de livre

- [3] F. FLEURET, D. GEMAN, «Coarse-to-fine visual selection», *International Journal of Computer Vision*, 2000, to be published.

Communications à des congrès, colloques, etc.

- [4] S. BERNARD, N. BOUJEMAA, D. VITALE, C. BRICOT, «AFIS as Domain Specific CBIR System», *in: Proc. of IEEE Workshop on Signal Processing Applications*, Brisbane, Australia, 2000.
- [5] N. BOUJEMAA, C. NASTAR, J. MALKI, «Requêtes partielles sans segmentation pour la recherche d'images par le contenu», *in: Proceedings of RFIA2000*, 2000.
- [6] N. BOUJEMAA, «Generalized Competitive Clustering for Image Segmentation», *in: Proc. of NAFIPS'2000*, Atlanta, Georgia, 13-15 Jul. 2000.
- [7] N. BOUJEMAA, «On Competitive Unsupervised Clustering», *in: Proc. of ICPR '2000*, Barcelona, Spain, 3-8 Sept. 2000.
- [8] F. FLEURET, D. GEMAN, «Apprentissage hiérarchique pour la détection de visages», *in: Proceedings of RFIA2000*, 2, p. 349–357, 2000.
- [9] H. SAHBI, N. BOUJEMAA, «From Coarse To Fine Skin and Face Detection.», *in: Proc. of 8th ACM International Conference on Multimedia*, 2000.
- [10] C. VERTAN, N. BOUJEMAA, «Color Texture Classification by Normalized Color Space Representation», *in: Proc. of ICPR '2000*, 3, p. 584–587, Barcelona, Spain, 3-8 Sept. 2000.
- [11] C. VERTAN, N. BOUJEMAA, «Embedding Fuzzy Logic in Content Based Image Retrieval», *in: Proc. of NAFIPS'2000*, p. 85–90, Atlanta, Georgia, 13-15 Jul. 2000.
- [12] C. VERTAN, N. BOUJEMAA, «Spatially constrained color distributions for image indexing», *in: Proc. of CGIP'2000*, p. 261–265, Saint Etienne, France, 1-4 Oct. 2000.
- [13] C. VERTAN, N. BOUJEMAA, «Upgrading Color Distributions for image retrieval: Can we do better?», *in: Proc. of Visual 2000*, Lyon, France, 2-4 Nov. 2000.
- [14] C. VERTAN, N. BOUJEMAA, «Using Fuzzy Histograms and Distances for Color Image Retrieval», *in: Proc. of CIR'2000*, Brighton, United Kingdom, 4-5 May 2000.
- [15] C. VERTAN, M. CIUC, N. BOUJEMAA, «On the introduction of a chrominance spectrum and its applications», *in: Proc. of CGIP'2000*, p. 214–218, Saint Etienne, France, 1-4 Oct. 2000.
- [16] C. VERTAN, M. CIUC, V. BUZULOIU, N. BOUJEMAA, «Color distributions and outlier detection: joint use and mutual influences», *in: Proc. of Communications'2000*, Bucharest, Romania, 7-9 Dec. 2000.