INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Project-Team Imedia

Images and Multimedia : Indexing, Retrieval and Navigation

Rocquencourt

THEME COG

Activity Report

2006

Table of contents

1. Team

Head of the team
Nozha Boujemaa [Research Director (DR), HdR]

Assistante de projet
Laurence Bourcier [Secretary (SAR) Inria (shared with Salsa and Micmac project-team)]

Personnel Inria
Alexis Joly [Research Associate (CR) Inria (since October 2006)]
Anne Verroust-Blondet [Research Associate (CR) Inria, HdR]
Jean-Paul Chièze [Senior Technical Staff Inria (half-time)]

Research scientists (external)
Olivier Buisson [Research Scientist at Institut National de l'Audiovisuel]
Michel Crucianu [Professor at CNAM, HdR]
Valérie Gouet-Brunet [Assistant Professor at CNAM]
Marie-Luce Viaud [Research Scientist at Institut National de l'Audiovisuel]
Itheri Yahiaoui [ATER at CNAM till August 31 2006 and Assistant Professor at Reims University since September 1st 2006]

Invited Professor
Donald Geman [Professor at Johns Hopkins University, USA (January and June 2006), HdR]

Post-doctoral Fellow
Jaume Amores [Postdoctoral fellow since March 1st 2006]
Sabri Boughorbel [ACET Scientifique Inria (till February 28 2006)]
Marin Ferecatu [ACET Scientifique Inria]
Alexis Joly [ACET Scientifique Inria (from January 1st till September 30 2006)]
Hong Tang [Postdoctoral fellow since March 1st 2006 (shared with ENST Paris)]

Technical staff
Mehdi Bouabta [Junior Technical Staff Inria since September 1st 2006]
Mathieu Coutaud [Junior Technical Staff Inria]

Ph. D. Student
Olfa Besbes-Abdelhak [Joint tutorship with Sup'Com, national grant since September 1st 2004]
Mohamed Chaouch [INRIA grant, Télécom Paris since October 1st 2005]
Nizar Grira [INRIA grant, Paris-Sud University until April 30 2006]
Nicolas Hervé [INRIA grant, Paris-Sud University since December 1st 2005]
Hichem Houissa [INRIA grant, Paris-Sud University]
Julien Law-To [CIFRE grant with INA since May 1st 2004]

Interns
Sahbi Bahroun [Sup'Com, Tunis, since September 18 2006]
Ahmed Rebaï [Sup'Com, Tunis, from March 1st till September 30 and from November 1st 2006]

2. Overall Objectives

2.1. Overall Objectives

One of the consequences of the increasing ease of use and significant cost reduction of computer systems is the production and exchange of more and more digital and multimedia documents. These documents are fundamentally heterogeneous in structure and content as they usually contain text, images, graphics, video and sounds.

Information retrieval can no longer rely on text-based queries alone; it will have to be multi-modal and to integrate all the aspects of the multimedia content. In particular, the visual content has a major role and represents a central vector for the transmission of information. The description of that content by means of image analysis techniques is less subjective than the usual keyword-based annotations, whenever they exist. Moreover, being independent from the query language, the description of visual content is becoming paramount for the efficient exploration of a multimedia stream.

In the IMEDIA group we focus on the intelligent access by visual content. With this goal in mind, we develop methods that address key issues such as content-based indexing, interactive search and image database navigation, in the context of multimedia content.

Content-based image retrieval systems provide help for the automatic search and assist human decisions. The user remains the *maître d'oeuvre*, the only one able to take the final decision. The numerous research activities in this field during the last decade have proven that retrieval based on the visual content was feasible. Nevertheless, current practice shows that a usability gap remains between the designers of these techniques/methods and their potential users.

One of the main goals of our research group is to reduce the gap between the real usages and the functionalities resulting from our research on visual content-based information retrieval. Thus, we apply ourselves to conceive methods and techniques that can address realistic scenarios, which often lead to exciting methodological challenges.

Among the "usage" objectives, an important one is the ability, for the user, to express his specific visual interest for a *part of* a picture. It allows him to better target his intention and to formulate it more accurately. Another goal in the same spirit is to express subjective preferences and to provide the system with the ability to learn those preferences. When dealing with any of these issues, we keep in mind the importance of the response time of such interactive systems. Of course, what value the response time should have and how critical it is depends heavily on the domain (specific or generic) and on the cost of the errors.

Our research work is then at the intersection of several scientific specialities. The main ones are image analysis, pattern recognition, statistical learning, human-machine interaction and database systems. It is structured into the following main themes :

1. Image indexing : this part mainly concerns modelling the visual aspect of images, by means of image analysis techniques. It leads to the design of image signatures that can then be obtained automatically.
2. Clustering and statistical learning : generic and fundamental methods for solving problems of pattern recognition, which are central in the context of image indexing.
3. Interactive search and personalisation : to let the system take into account the preferences of the user, who usually expresses subjective or high-level semantic queries.
4. Cross-media indexing, and in particular bimodal *text + image* indexing, which addresses the challenge of combining those two media for a more efficient indexing and retrieval.

More generally, the research work and the academic and industrial collaborations of the IMEDIA team aim to answer the complex problem of the intelligent access to multimedia content.

3. Scientific Foundations

3.1. Introduction

We group the existing problems in the domain of content-based image indexing and retrieval in the following themes : image indexing, pattern recognition, personalisation and cross-media indexing. In the following we give a short introduction to each of these themes.

3.2. Modelling, construction and structuring of the feature space

Keywords: *image analysis*, *image features and signatures*, *indexing of visual content*, *matching* , *pattern recognition*, *visual appearance*, *visual similarity*.

Participants: Nozha Boujemaa, Alexis Joly, Mohamed Chaouch, Valérie Gouet-Brunet, Itheri Yahiaoui, Hichem Houissa, Anne Verroust-Blondet, Ahmed Rebai, Jean-Paul Chièze.

Content-based indexing *the process of extracting from a document (here a picture) compact and structured significant visual features that will be used and compared during the interactive search.*

The goal of the IMEDIA team is to provide the user with the ability to do content-based search into image databases in a way that is both intelligent and intuitive to the users. When formulated in concrete terms, this problem gives birth to several mathematical and algorithmic challenges.

To represent the content of an image, we are looking for a representation that is both compact (less data and more semantics), relevant (with respect to the visual content and the users) and fast to compute and compare. The choice of the feature space consists in selecting the significant *features*, the *descriptors* for those features and eventually the encoding of those descriptors as image *signatures*.

We deal both with generic databases, in which images are heterogeneous (for instance, search of Internet images), and with specific databases, dedicated to a specific application field. The specific databases are usually provided with a ground-truth and have an homogeneous content (faces, medical images, fingerprints, etc.)

Note that for specific databases one can develop dedicated and optimal features for the application considered (face recognition, etc.). On the contrary, generic databases require generic features (colour, textures, shapes, etc.).

We must not only distinguish generic and specific signatures, but also local and global ones. They correspond respectively to queries concerning parts of pictures or entire pictures. In this case, we can again distinguish approximate and precise queries. In the latter case one has to be provided with various descriptions of parts of images, as well as with means to specify them as regions of interest. In particular, we have to define both global and local similarity measures.

When the computation of signatures is over, the image database is finally encoded as a set of points in a high-dimensional space : the feature space.
A second step in the construction of the index can be valuable when dealing with very high-dimensional feature spaces. It consists in pre-structuring the set of signatures and storing it efficiently, in order to reduce access time for future queries (tradeoff between the access time and the cost of storage). In this second step, we have to address problems that have been dealt with for some time in the database community, but arise here in a new context : image databases. The diversity of the feature spaces we deal with force us to design specific methods for structuring each of these spaces.

3.3. Pattern recognition and statistical learning

Statistical learning and classification methods are of central interest for content-based image retrieval [40] [45].

We consider here both supervised and unsupervised methods. Depending on our knowledge of the contents of a database, we may or may not be provided with a set of *labelled training examples*. For the detection of *known* objects, methods based on hierarchies of classifiers have been investigated. In this context, face detection was a main topic, as it can automatically provide a high-level semantic information about video streams. For a collection of pictures whose content is unknown, e.g. in a navigation scenario, we are investigating techniques that adaptatively identify homogeneous clusters of images, which represent a challenging problem due to feature space configuration.

3.3.1. Statistical learning and object detection

Keywords: *Statistical learning*, *boosting*, *kernel methods*, *object retrieval*, *object detection*.

Participants: Donald Geman, Jaume Amores, Nicolas Hervé.

Object detection is the most straightforward solution to the challenge of content-based image indexing. Classical approaches (artificial neural networks, support vector machines, etc.) are based on induction, they construct generalization rules from training examples. The generalization error of these techniques can be controlled, given the complexity of the models considered and the size of the training set.

Our research on object detection addresses the design of invariant kernels and algorithmically efficient solutions as well as boosting method for similarity learning. We have developed several algorithms for face detection based on a hierarchical combination of simple two-class classifiers. Such architectures concentrate the computation on ambiguous parts of the scene and achieve error rates as good as those of far more expensive techniques.

3.3.2. Clustering methods

Keywords: *clustering*, *competitive agglomeration*, *membership*, *number of classes*, *pattern recognition*.

Participants: Nozha Boujemaa, Nizar Grira, Michel Crucianu, Itheri Yahiaoui.

Unsupervised clustering techniques automatically define categories and are for us a matter of visual knowledge discovery. We need them in order to :

- Solve the "page zero" problem by generating a visual summary of a database that takes into account all the available signatures together.
- Perform image segmentation by clustering local image descriptors.
- Structure and sort out the signature space for either global or local signatures, allowing a hierarchical search that is necessarily more efficient as it only requires to "scan" the representatives of the resulting clusters.

Given the complexity of the feature spaces we are considering, this is a very difficult task. Noise and class overlap challenge the estimation of the parameters for each cluster. The main aspects that define the clustering process and inevitably influence the quality of the result are the clustering criterion, the similarity measure and the data model.

We investigate a family of clustering methods based on the competitive agglomeration that allows us to cope with our primary requirements : estimate the unknown number of classes, handle noisy data and deal with classes (by using fuzzy memberships that delay the decision as much as possible).

3.4. Interactive search and personalisation

Keywords: *expression of preferences*, *interaction with the user*, *relevance feedback*, *semantic gap*, *statistical learning*, *subjective clustering*.

Participants: Marin Ferecatu, Donald Geman, Nozha Boujemaa, Michel Crucianu, Hichem Houissa, Jean-Paul Chièze.

We are studying here the approaches that allow for a reduction of the "semantic gap" There are several ways to deal with the semantic gap. One prior work is to optimize the fidelity of physical-content descriptors (image signatures) to visual content appearance of the images. The objective of this preliminary step is to bridge what we call the numerical gap. To minimize the numerical gap, we have to develop efficient images signatures. The weakness of visual retrieval results, due to the numerical gap, is often confusingly attributed to the semantic gap. We think that providing richer user-system interaction allows user expression on his preferences and focus on his semantic visual-content target.

Rich user expression comes in a variety of forms :

- allow the user to notify his satisfaction (or not) on the system retrieval results–method commonly called relevance feedback. In this case, the user reaction expresses more generally a subjective preference and therefore can compensate for the semantic gap between visual appearance and the user intention,
- provide precise visual query formulation that allows the user to select precisely its region of interest and pull off the image parts that are not representative of his visual target,
- provide a mechanism to search for the user mental image when no starting image example is available. Several approaches are investigated. As an example, we can mention the logical composition from visual thesaurus. Besides, learning methods related to information theory are also developed for efficient relevance feedback model in several context study including mental image retrieval.

3.5. Cross-media indexing and retrieval

Keywords: *hybrid indexing and search*, *information theory*, *textual annotation*.

Participants: Marin Ferecatu, Nicolas Hervé, Nozha Boujemaa, Michel Crucianu.

We have described, up to now, our research approaches in using the visual content alone. But when additional information is available, it may prove complementary and potentially valuable in improving the results returned to the user. We may cite here *metadata* (file name, date of creation, caption, etc.) but also the textual annotations that are sometimes available. We must note that annotations usually carry high-level information related to a prior knowledge of the context. The use of these sources of information implies that we can speak of multimedia indexing.

We can think of several approaches for combining textual and visual information in the context of indexing and retrieval. As examples, we may cite the automatic textual annotation of images based on similarities between visual signatures or the propagation of textual annotations relying on the interaction between textual ontologies and visual ontologies. We also investigate methods that allow automatic textual annotation from visual content analysis. This part of our research activities is yet another solution for the reduction of the "semantic gap".

4. Application Domains

4.1. Application Domains

- **Security applications** Examples : Identify faces or digital fingerprints (biometry). Biometry is an interesting specific application for both a theoretical and an application (recognition, supervision, ...) point of view. Two PhDs were defended on themes related to biometry. Our team also worked with a database of images of stolen objects and a database of images after a search (for fighting pedophilia). We are currently collaborating with the Ministry of the Interior.
- **Multimedia** Examples : Look for a specific shot in a movie, documentary or TV news, present a video summary. Our team has a collaboration with the TV channel TF1 in the context of a RIAM project. Text annotation is still very important in such applications, so that cross-media access is crucial.
- **Scientific applications** Examples : environmental images databases : fauna and flora; satellite images databases : ground typology; medical images databases : find images of a pathological character for educational or investigation purposes. We have an ongoing project on multimedia access to biodiversity collections.

- **Culture, art and education** Examples : encyclopedic research, query by example of paintings or drawings, query by a detail of an image. IMEDIA has been contacted by the French ministry of culture and by museums for their image archives.
  Finding a specific texture for the textile industry, illustrating an advertisement by an appropriate picture. IMEDIA is working with a picture library that provides images for advertising agencies.
- **Telecommunications** Examples : image representation and content-based queries stand as the basis of MPEG-4 and MPEG-7. IMEDIA does not contribute to their normative aspects but is interested in the latest results related to the MPEG-7 group. Note that the signatures developed by IMEDIA can be used with this norm.

5. Software

5.1. IKONA/MAESTRO Software

Keywords: *CBIR*, *User interface*, *image retrieval by content*, *relevance feedback*.

Participants: Nozha Boujemaa, Marin Ferecatu, Nicolas Hervé, Jean-Paul Chièze, Mathieu Coutaud, Alexis Joly, Mehdi Bouabta, François Fleuret.

IKONA is a framework for building Content Based Image Retrieval software prototypes. It has been designed and implemented in our team during the last four years [42]. The current version is fully generic and is highly adaptable to any CBIR scenario thanks to its level of abstraction. As a research environment, IKONA offers support to the researchers in their work by providing stable and tested tools. As an application, it can easily be deployed and used by non-specialist users.

IKONA is based on a client/server architecture. The communication between the two components is achieved through a proprietary network protocol. It is a set of commands the server understands and a set of answers it returns to the client. The communication protocol is extensible, i.e. it is easy to add new functionalities without disturbing the overall architecture. It is also modular and therefore can be replaced by any new or existing protocol dealing with multimedia information retrieval.

The main processes are on the server side. They can be separated in two main categories:

- offline processes: data analysis, features extraction and structuration
- online processes : answer the client requests

The images are characterised with **Global** signatures that are implemented in the server :

- Generic signatures : Color, Shape and Texture features investigated at the IMEDIA Group.
- Specific signatures : Faces and signatures for fingerprints.
- Annotations : Some keywords.

Besides, two **local** signatures are included : The region-based description and the point-based one. The server uses image signatures and offers several types of query paradigms, available to the user through the graphical interfaces of the clients :

- **query by global example** : The user selects an entire image as visual query.
- **partial queries** : the user is looking for regions in images that are visually similar to a the selected region.
- **relevance feedback on global and partial query**: the user interacts with the system in a feedback loop, by giving positive and negative examples to help the system identify the category of images she/he is interested in [42];
- **mental image search**: Two different methods are investigated. The first is Target Image Search with relevance feed-back model based on mutual information, the second one consist on Logical Query Composition.

We have developed two main clients that can communicate with the server. A good starting point for exploring the possibilities offered by IKONA is our web demo, available at http://www-rocq.inria.fr/cgi-bin/imedia/ikona. This CGI client is connected to a running server with several generalist and specific image databases, including more than 23,000 images. It features query by example searches, switch database functionality and relevance feedback for image category searches. The second client is a desktop application. It offers more functionalities. More screen-shots describing the visual searching capabilities of IKONA are available at http://www-rocq.inria.fr/imedia/cbir-demo.html.

The architecture of this client/server software and several visual signatures were a subject of a deposit to APP.

6. New Results

6.1. Construction and organization of the visual feature space

6.1.1. Video Copy detection

Keywords: *Content-Based Video Copy Detection*, *Video description*.

Participants: Julien Law-To, Valerie Gouet-Brunet, Olivier Buisson [INA], Nozha Boujemaa.

(a) Similar videos but not copies (Two different talks of Koffy Annan)

(b) Two copies (same video source) (*Gala du Midem.* G. Ulmer 1970 (c) INA)

Figure 1. Copy / similarity .

Motivation and challenge.

Due to the increasing broadcasting of digital video content (TV Channels, Web-TV, Video Blogs, etc), finding copies in a large video database has become a critical new issue and Content Based Copy Detection (CBCD) presents an alternative to the watermarking approach to identify video sequences. Video archive professionals need to trace the uses of their video databases and CBCD is a powerful tool for this task. At the same time, it provides a useful tool for Video Web servers to quickly and automatically control the copyrights of the videos uploaded by the users. A crucial difficulty is the fundamental difference between a copy and the notion of similar image encountered in Content-Based Video Retrieval (CBVR): a copy is not an identical or a near replicated video sequence but rather a transformed video sequence. These photometric or geometric transformations (gamma and contrast transformations, overlay, shift, etc) can greatly modify the signal, and therefore a copy can in fact be visually less similar than other kinds of videos that might be considered similar (see figure 1). CBVR applications aim to find similar videos in the same visual category, like for example soccer games, but most of these detections would clearly represent false alarms in a CBCD application.

Proposed Concept.

We propose a solution called *ViCopT* for Video Copy Tracking [28], [27], [30]. It involves estimating and characterising trajectories of complementary points of interest throughout the video sequence. Building trajectories of points in videos is a recent topic for video content indexing. We plan on taking advantage of such trajectories to characterise the spatio-temporal content of videos. First, the redundancy of the local description along the trajectory can be efficiently summarised with a reduced loss of information. Second, the trajectory properties will allow the local description to be enriched by adding a spatial and temporal behavior of this point. Analysing the trajectories obtained makes it possible to assign a label of behavior to a local descriptor. The aim is to provide a *rich*, *compact* and *generic* video content description, which will be used with a robust voting function dedicated to copy detection. The behavior of a point along a trajectory can be seen as a *temporal context* of this point. This work is done in collaboration with INA, within the scope of the CIFRE thesis of Julien Law-To. Close to this work, the concept proposed by J. Sivic and A. Zisserman in [52] also involves points of interest in video sequences. The concept is different from our work because it is based on finding objects in the video and so temporal behavior is not considered.

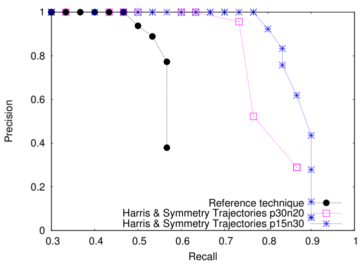
(a) *10e concours eurovision de la chanson.* 1965 (c)ORTF.

(b) Left: *Les duos de l'impossible* 2005. Right: *Système deux.* C. Fayard 1975 (c)INA

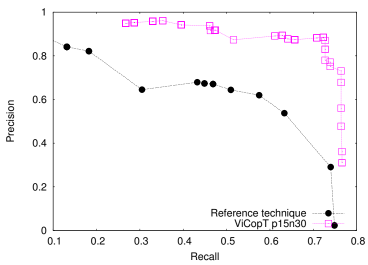
Figure 2. Detection on Real cases: video queries on the left and videos detected on the right.

Experimental results.

All the experiments were carried out on 700 hours of videos chosen randomly from the video archive database stored at INA. These videos were TV sequences from several kinds of programs (sports events, news shows, talk shows). The reference technique used for comparison is the one detailed in [50].



(a) Detecting Video segments



(b) Detecting Video frames

Figure 3. Evaluation of ViCopT.

ViCopT (cf. Figures 2 and 3) shows a real improvement in the performances compared with a state-of-the art technique. The precision of each detection is also improved by our system. Other tests have been done on real life cases by monitoring TV and by using videos found on the web.

6.1.2. Adaptive satellite images segmentation by level set multiregion competition

Keywords: *adaptive multispectral image segmentation*, *discrimination power*, *level set theory*, *multiregion competition*, *textured / non-textured regions*.

Participants: Olfa Besbes, Nozha Boujemaa, Ziad Belhadj [SUP'COM - Tunisia].

The application of various imaging satellite sensors results in huge quantities of data. Therefore, effective querying and browsing are needed to easily and selectively exploit these remote sensing image archives. Region-Based Image Retrieval (RBIR) is a powerful tool since it allows to search images containing similar objects of a reference image. It requires the satellite image to be parsed into its constituent patterns. Satellite images consist of multiple layers of stochastic processes such as texture, texton, interest point, line, curve, region, and object processes. This is even more true today with high resolution images such as IKONOS, SPOT-5 and QuickBird data. This observation motivates our work in perceptual grouping, which subsumes image segmentation as region process, and naturally integrates perceptual organization and object recognition. In order to solve the perceptual grouping task, we use the hierarchical Bayesian inference framework [51] so that the different layers such as geometric structure, texton and region processes cooperate to robustly explain the input image. Our first work on image segmentation is a variational method that combines adaptively texture and spectral cues according to their discrimination power [35]. As all PDE methods, it cannot handle dimension changes which need reversible jumps as proposed in Markov chain Monte Carlo methods. Our goal is then to develop an effective MCMC-based perceptual grouping algorithm to search for globally optimal solution in a complex solution space of all possible visual pattern compositions. It involves both stochastic diffusions by partial differential equation and reversible jumps with Metropolis-Hastings method. In particular, we are interested in Swendsen Wang-Cut algorithm [38] since it can simulate reversible split-merge moves and thus find globally optimal solution for a general posterior probability. In future work, we intend to extend this work in two directions: First, improving the perceptual grouping by using a multi-scale framework and prior learned dictionaries. Second, describing effectively the extracted stochastic patterns for partial queries in satellite image databases.

6.1.3. 3D indexing : 2D/3D shape descriptors

Keywords: *2D/3D descriptor*, *3D descriptor*, *3D indexing*, *3D model retrieval*, *depth-buffer image*, *silhouette*.

Participants: Mohamed Chaouch, Anne Verroust-Blondet.

We have pursued the work on 3D shape retrieval based on 2D views we begun last year. To take into account the diversity of information contained in the 2D views of a 3D model, relevance index models have been introduced. The relevance values associated to the silhouettes or the depth-buffer images are then used when computing the dissimilarity between two 3D shapes. Two 3D-shape retrieval methods have been proposed: the enhanced silhouette approach (ESA) and the enhanced depth-buffer approach (EDBA). Their efficiency have been evaluated on the Princeton Shape Benchmark (cf. the precision-recall curves on Figure 4). This work is described in [20].

We participated to the SHREC - 3D shape retrieval contest organised by the European Network of Excellence AIM@SHAPE in 2006. Our depth-buffer approach EDBA obtained good retrieval results in the SHREC classification (cf. [54]).

The ESA and EDBA methods can be tested on the demonstrator built for the "Description, Matching and Retrieval by Content of 3D Objects" project of the European Network of Excellence DELOS II (http://www-rocq.inria.fr/cgi-bin/imedia/3d.cgi [39]). This demonstrator shows also 3D-shape retrieval approaches developed by our DELOS partners (University of Firenze, University of Modena and ENIC).

6.1.4. Spatial visual query configuration

Keywords: *Weighted fuzzy orientations*, *histogram of angles*, *spatial composition*.

Participants: Hichem Houissa, Nozha Boujemaa, Hichem Frigui.

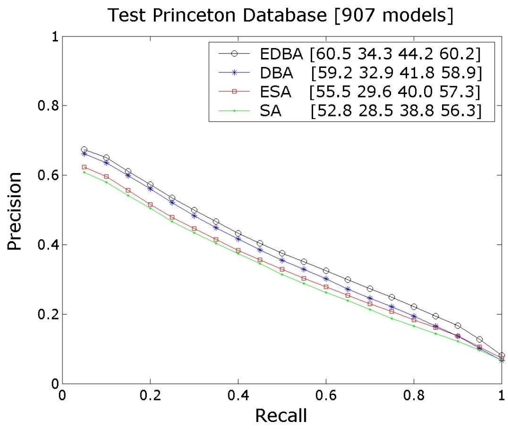
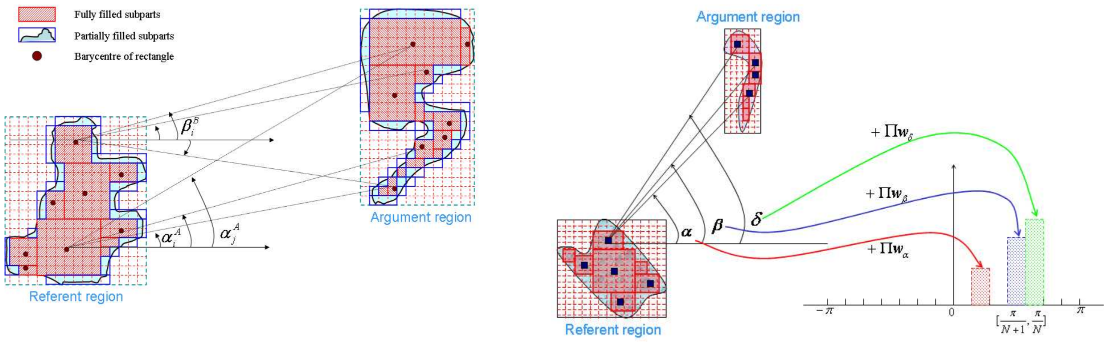


Figure 4. Precision-recall curves for the global average using the ESA and EDBA approaches and the two initial approaches without relevance index (SA and DBA). The mean NN, FT, ST and DCG values are given in the legends.

We developed a novel approach to model the spatial relations between the compose visual patches. We define the Weighted Angle Spatial Histogram (WASH) that jointly uses the (ir)regularity and angles between regions of interest to determine their relative spatial position [26]. Depending on the topological aspect of the considered couple of regions, we propose a histogram that meets the following requirements: **(1)** Only visually relevant parts of each region are considered and adaptively weighted during angle computation, **(2)** coarse segmentation resulting in rough and inaccurate contours is balanced by means of appropriate weighting functions, **(3)** the proposed histogram is invariant to pairwise permutation of objects.



(a) Adaptive size rectangles are embedded into regions of interest (b) Weighted angular orientations are computed

Figure 5. WASH construction procedure

The principle of WASH relies basically on an adaptive grid over each region of interest, and a weighted angle computation between respective centers of gravity (Fig 5).

In practice, the use of WASH is to enhance more robustness in the retrieved results when the user composes his mental query image. Typically, the user is proposed a Visual Thesaurus summarising all regions of the database in categories and represented by prototypes. He selects the visual patches he is interested in and arranges them on a white board in a way that corresponds to his mental image.

(a) Query

(b) Landscape visual composition

(c) Query

(d) Seascape visual composition

(e) Query

(f) Harbor and mountain visual composition

Figure 6. Retrieval results using both visual thesaurus and spatial relations

Some resulting queries are shown in Fig 6 where images are ranked with respect to increasing histogram intersection distance. Notice that the confidence threshold is decreased for purpose for these results in order to retrieve most relevant images. Obviously, the user may reject some images afterwards.

This work has been done inside a collaboration with University of Louisville, KE, USA (Departement of Computer Engineering and Computer Science), cofunded by INRIA and NSF.

6.1.5. Local descriptors and visual saliency

Keywords: *image retrieval*, *interest points*, *local topology*, *local descriptors*, *visual saliency*.

Participants: Ahmed Rebai, Alexis Joly, Nozha Boujemaa.

In computer vision, interest points are used to reduce data processing and to focus on stable and informative contents. To achieve this goal, we have to extract the minimum of points which describe the best a given image. The problem with standard interest points' detectors consists in the extraction of several corners and

T-junctions without a visual saliency assigned to the detected points. This leads to a considerable increase of the output data. One way to circumvent this problem consists in selecting points which have better visual interpretability. Within this framework, we developed a new interest points' detector. The specificity of this detector is that it enables the detection of points from different topological natures. It can detect centers of circles, centers of logarithmic spirals, or more generally centers of portions of logarithmic spirals as well as intersections of such curves (cf. Figure 7).

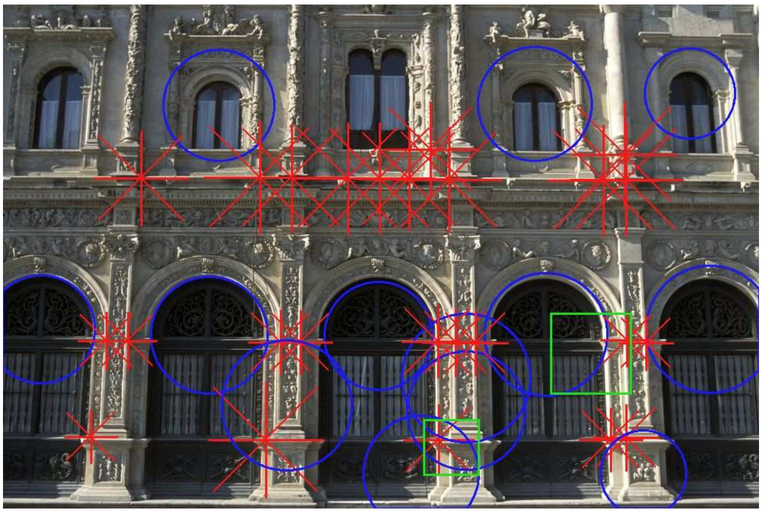


Figure 7. This is a detection example. Points detected at $\theta = 0°$ are represented by circles, those detected at $\theta = 45°$ by squares and finally those detected at $\theta = 90°$ by cross.

These points are called "Constant Tangential Angle Elected Interest Points". The transform is based on votes made by vectors tuned by an angle θ form the gradients' directions. By varying θ between $[-\frac{\pi}{2}, \frac{\pi}{2}]$, we construct a more rich analysis space. We expect that the $3D$-maxima detected in this space are more stable to different image transformations. In addition, the developed algorithm was integrated into the IMEDIA CBIR software to make use of it in various applications. Experimental results show that the repeatability measure is comparable to that of Harris's detector. The present work is described in [31], [37]. In future work, we will study the relationship between the three dimensional theta space and the scale space theory. Then we will test the transform in some computer applications such as indexing and object recognition.

6.1.6. Silhouette shape descriptor for image retrieval

Keywords: *content-based image retrieval*, *partial queries*, *shape descriptor*.

Participants: Itheri Yahiaoui, Nozha Boujemaa, Nicolas Hervé.

Last year, we designed a new shape descriptor that we called Directional Fragment histogram (DFH). This new shape descriptor was designed within BIOTIM project. Our main objective was to propose a new shape descriptor with real time performance for image retrieval in botanical databases compared with existing shape descriptors.

Different versions of the DFH descriptor were implemented and tested. In order to evaluate the performances of this new shape descriptor, we compared it with some shape descriptors of literature, in particular those used in Mpeg7 as CSS and EOH [32]

For three botanical databases, DFH and CSS-MPEG7 descriptors give comparable retrieval performances. However, CSS computation time is approximately 250 times that of the DFH.

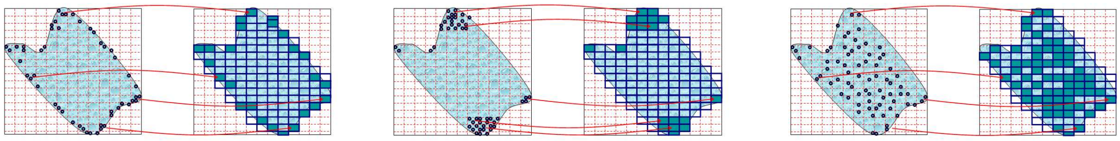
In order to obtain a generic descriptor, we are studying the possibility of adapting our DFH to binary images without use of the originals for gradient computation.

6.2. Interactive retrieval

6.2.1. Relational categorization of adaptive local descriptors toward interactive composition from Visual Thesaurus

Keywords: *Earth Mover's distance*, *Harris points of interest*, *relational clustering*.

Participants: Hichem Houissa, Nozha Boujemaa.



(a) Points of interest are located on the contours due to rough region (b) Points of interest are confined in small parts of the region because of coarse segmentation (c) Points of interest cover the whole region denoting of a textured structure

Figure 8. Several cases of spatial points dispersion: (a) on contours (b) in tiny parts of the region and (c) scattered region-wide

Fauqueur and Boujemaa [43] introduced a new Query By Visual Thesaurus paradigm that enriches the QBVE by the creation of the "page zero". This paradigm stipulates that if the user has forgotten his query image or simply has a vague idea about it, he can compose his query by selecting several visual patches on a Visual Thesaurus (VT). VT is then a new query alternative that overcomes the absence of starting example image and offers the possibility to combine multiple visual patches in order to retrieve the target mental image. The Visual Thesaurus is a summary of all the regions in the database; each category contains similar regions according to low-level visual features. The construction of the VT should be as reliable as possible to reflect the human visual categorization. We have introduced a novel semantic region labelling criterion based on point of interest dispersion. This point-based coherence criterion (PCC) leads to label regions through topological and spatial dispersion of points of interest [24], [25]. PCC is evaluated on single coarse regions using Harris colour points detector that catches the local photometric variability on small sites (few pixels). As far as points of interest are concerned, a homogeneous region is more likely to contain less points than a textured one because points catch the local photometric variability around a very small site (Fig 8).

Once regions are separated into homogeneous and textured categories, we measure the similarity between pairs of regions knowing that homogeneous ones are described by means of fix-size colour distribution, whereas textured patches are assessed through multi-dimensional descriptors based on the Hilbert's invariants combination. Hence, we use suitable metrics that cope with unbalanced sets of points. The Earth Mover's Distance (EMD) and Hausdorff distance are investigated for reliable similarity measures [23]. Furthermore, the relational clustering is used in order to generate the visual categories according to the similarity measures computed previously. The resulting clusters are represented by their prototypes -the region having the highest fuzz membership degree to the underlying category- and the whole set of prototypes (visual thesaurus) is proposed to the user as a page zero to initiate his mental image search by visual composition [22].

6.2.2. Interactive retrieval from remote sensing repositories using active relevance feedback

Keywords: *active learning*, *kernel function*, *remote sensing repositories*, *sample selection*.

Participants: Marin Ferecatu, Nozha Boujemaa.

As the resolution of remote sensing imagery increases, the full complexity of the scenes becomes increasingly difficult to approach. Data access by time of acquisition, geographical position or type of sensor is often less important than the content of the scene, e.g. structures, objects, etc. Interesting applications involve complicated spatial and structural relationships among image objects. In this context, we argue that relevance feedback (RF), method traditionally used for searching complex image classes in generic image databases, can be successfully employed with databases of satellite images.

We evaluate, in the context of satellite image retrieval, two general improvements of SVM-based relevance feedback method. First, we focus on the criterion employed by the system for selecting the images presented to the user at every feedback round. Our relevance feedback scheme use an active learning selection criterion that minimizes redundancy between the candidate images shown to the user. Second, for image classes having very different scales, we find that a high sensitivity of the SVM to the scale of the data brings about a low retrieval performance. Insensitivity to scale is desirable in this context and we show how to obtain it by the use of specific kernel functions [44].

Experimental evaluation of both ranking and classification performance on a groundtruth database of satellite images confirms the effectiveness of our approach [13]. In Fig. 9 we show a visual comparison of our relevance feedback method (right) with the standards query by example technique.

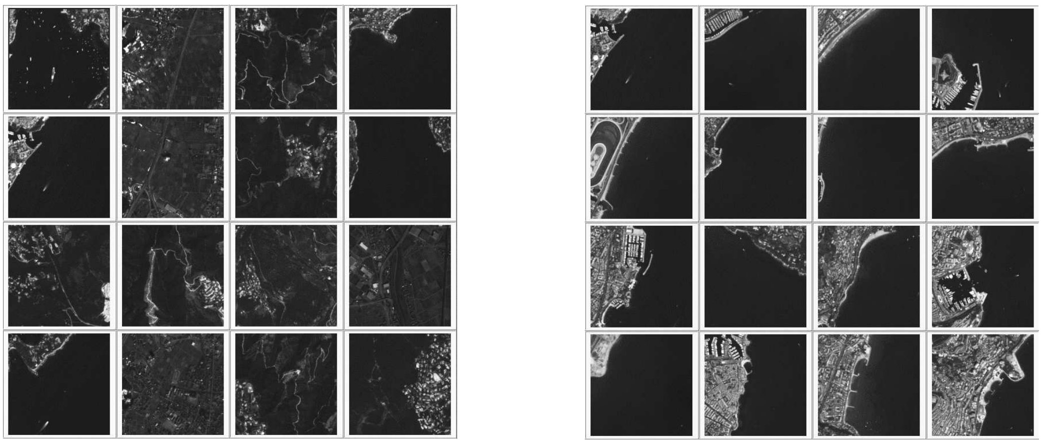


Figure 9. Search with relevance feedback (right): looking for urban coast areas (results returned after marking 5 images as "relevant" and 8 images as "irrelevant"). Compared to the query by example case (left), the results proposed by relevance feedback are much closer to the intended target class.

6.2.3. Interactive Image Retrieval by Mental Matching

Keywords: *Bayesian learning*, *category search*, *mental image*, *relevance feedback*.

Participants: Marin Ferecatu, Donald Geman, Nozha Boujemaa.

Standard query scenarios in image retrieval, e.g. query by example or interactive browsing by relevance feedback, are affected by the famous page zero problem: to start a query they all need some first examples that are relevant to the target. Finding relevant examples by random sampling is not feasible even for medium sized databases (a few thousands of images). Solutions to this problem include query by sketch, or query by visual composition of image patches; but these solutions are not mature enough to be applied for large databases. To answer this problem, we propose an interactive search model which is based on information theory and statistical inference. We allow the search to start from any random set of images, and then we use relevance feedback to iteratively guide the system to the target class. To help the system minimize the number of iterations needed for finding the target class, we first cluster the image database into small classes that have a high degree of internal semantic coherence. The display algorithm involves a Bayesian relevance feedback model and an optimality criterion based on conditional entropy. Performance is measured by the expected number of iterations necessary to match the class of the query. To further improve the practical results with large databases, we design the metric and the response model such as to be consistent with human behavior. Preliminary experiments with several generalist ground truth image databases showed very promising results [21].

6.2.4. Semantic cartography of a database from a user's query

Keywords: *Euler diagrams*, *data visualisation*, *graph visualisation*.

Participants: Anne Verroust-Blondet, Marie-Luce Viaud [INA].

We have pursued the work performed with Marie-Luce Viaud from INA (Institut National de l'Audiovisuel). In order to help the user to structure and navigate inside a multimedia database, we want to built new visualisation tools. In the context of digital libraries, we carried on our work on Euler Diagram drawing (cf. [53]) and we studied the characteristics of graph visualisation methods. This has be done in part inside INFOM@GIC project (Subtask 1.12 : "Recherche, extraction, analyse et fusion d'information", cf. [55], [46]).

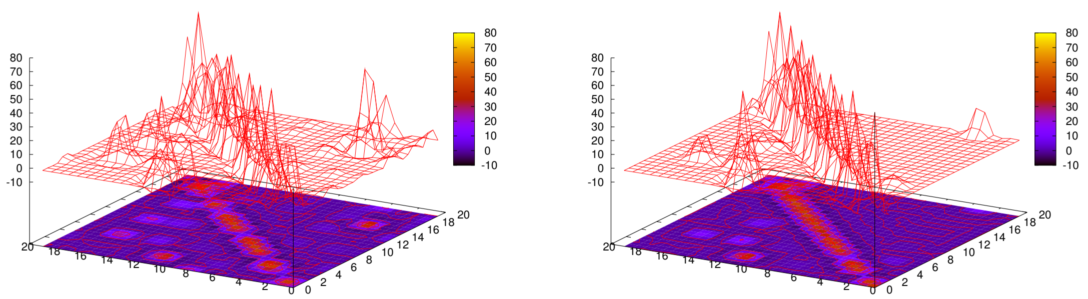
6.3. Clustering and learning

6.3.1. A local search augmentation of the fuzzy c-means algorithm

Keywords: *Clustering*, *Fuzzy c-means*, *local search*.

Participants: Nizar Grira, Michel Crucianu, Nozha Boujemaa, Michael Houle.

In an attempt to improve the quality of the clustering partitions, we were first concerned by adding a simple semantic information in the form of pairwise constraints between data-items to guide the clustering process results towards the ones expected by the user [16].



With local search without local search)

Figure 10. Results obtained on a database of 20 objects drawn randomly from the COIL-100 database: left, using the fuzzy c-means (FCM) algorithm and, right, using FCM augmented by a local search refinement step.

In a second attempt to improve the convergence of the proposed partitional algorithms [48], [47], we compared the convergence performance of two well studied and widely used algorithms namely the k-means (from which our algorithms derive) and k-medoids [49]. In the the k-means case, we are given a set S of n data items in a d dimensional *vector* space and asked to *determine* a set of k points called centers or prototypes so as to minimise the squared-error distortion (sum of squared distances from each data point to its closest center). While in k-medoids, we are given a set S of n patterns in a *metric* space and desire to *locate* k medians minimising the sum of distances from each pattern to the nearest medoid. After conducting series of experiments, we came up with the fact that in high dimensional feature spaces, k-medoids behave much better than k-means due to its exhaustive search for new centers. However, exhaustiveness means a time complexity that is quadratic in the size of input data, while one of our challenges is to make the partitioning linear in the size of the input and thus suitable for real world applications. Our proposal consists in augmenting k-means iterations (or the iterations of FCM) with a local search procedure inspired by the k-medoid approach to find better positions for the centers.

The refinement step uses a hierarchy of uniform samples to provide a set of center candidates that can improve an FCM partition by swapping them in and out of the k current output prototypes. The final method [8] is then a hybrid use of FCM together with a local search approach. It was tested on a set of image databases and results showed empirically that the proposed algorithm improved the partitions of the previous mentioned algorithms.

6.3.2. On boosting multiple patch-based rankings for object recognition and retrieval.

Keywords: *Object recognition*, *learning*, *object retrieval*.

Participants: Jaume Amores, Petia Radeva, Nicu Sebe, Qi Tian, Jie Yu.

We face with the problem of accurate object recognition and retrieval in the presence of unsegmented and cluttered images. Objects are characterised using constellations of local parts, and relevant local parts are learned in a weakly supervised manner. This is achieved by extracting multiple rankings of local patches: those local patches that rank positive images in top positions convey relevant information about the object, and efficient selection of constellations of these patches can be obtained by Boosting. Conducted experiments show that this approach outperforms other state-of-the-art bag-of-features approaches. Related to this work, we are investigating how to efficiently incorporate spatial and contextual information in the object description and how to estimate suitable distance functions for accurate ranking.

For incorporating contextual information, we introduce a framework based on Generalised Correlograms, a very sparse and discriminant representation that leads to efficient recognition when combined with Joint Boosting and Inverted Files.

For obtaining accurate distance functions, in [19], [33], [34] we propose learning the distance estimation by boosting in a product space. Results show that the derived distance function is more accurate than traditional metrics in a classification context.

6.4. Cross-modal indexing and retrieval

Keywords: *automatic image annotation*, *cross-modal retrieval*, *cross-modal indexing*.

Participants: Nicolas Hervé, Marin Ferecatu, Nozha Boujemaa, Alexis Joly.

The combination of visual content and textual metadata is a promising approach to bridge the semantic gap.

We continue our effort regarding the early fusion of visual descriptors and textual image annotation toward reliable concept search. An extensive study on the impact on the relevance feedback results has been conducted [11]

In addition, we studied deeply the automatic annotation methods for cross-modal approach for indexing. This work has been carried out in the Infom@gic project [41]. Furthermore, we start to investigate new approaches for automatic image annotation. Theses methods have been tested within the ImagEval benchmark 7.1.5 where IMEDIA won all the 3 image retrieval tasks. Some results on the ImagEval dataset are shown on Figure 11 and 12.

(a) Sunglasses (b) US Flag

Figure 11. First results for two queries of the 4th ImagEval task : Objects and objects class retrieval

(a) Art representation (b) Colour, Outdoor, Day, Natural scene

Figure 12. First results for two queries of the 5th ImagEval task : Semantic attributes extraction

6.5. Software

6.5.1. IKONA/MAESTRO software

Keywords: *CBIR*, *image retrieval by content*, *relevance feedback*, *user interface*.

Participants: Mathieu Coutaud, Mehdi Bouabta, Marin Ferecatu, Nicolas Hervé, Alexis Joly, Jean-Paul Chièze.

This year, new descriptors (namely local descriptors) and query paradigms have been integrated with unified interfaces defined last year. For example, the relevance feedback is now supported by the server. A mass annotation mechanism has also been developed and several parts of the graphical client have been rewritten to allow more interaction. Efforts are now concentrated to use a real relational database management system and still to migrate other << research code >> into the server.

6.5.2. aceMedia software integration

Keywords: *CBIR*, *aceMedia*, *image retrieval by content*.

Participants: Mathieu Coutaud, Nicolas Hervé.

In the aceMedia european integrated project, IMEDIA is in charge of the development of the "Intelligent Search and Retrieval" application module. This module brings together the software of four research teams that work on different multimedia information retrieval paradigms. This year the integration in the PCS User Interface have been finalised, and the support of video has started.

7. Other Grants and Activities

7.1. National Initiatives

7.1.1. Industrial contract with INA [2004-2007]

A co-supervision of a Phd within CIFRE context. The main topic is about optimal fine visual signatures for monitoring of INA video collections.

7.1.2. BIOTIM Project (exploiting Text-IMage ressources in BIOdiversity) within the national initiative "Masses of data"[2003-2006]

The partners of this project are the IMEDIA and ATOLL teams of INRIA Rocquencourt, the CEDRIC laboratory of the CNAM Paris, the LIFO laboratory of the University of Orléans, the Institute of Research for Development (IRD) and the National Institute for Research in Agriculture (INRA). BIOTIM is coordinated by IMEDIA. The project is financially supported by the French National Science Fund (FNS).

7.1.3. QuerySat Project within the national initiative "ACI Masses de Données")[2004-2007]

This project concerns the conception and development of content description methods for aerial and satellite images indexing and retrieval by content. This work is done jointly with ARIANA project Team (Sophia Antipolis), ENST-Paris (CNRS) and URISA research team from Sup'Com (School of Engineering - Tunis). One of the objectives is to make connection with symbolic and semantics features queries in the context of satellite image repositories.

7.1.4. InfoMagic[2005-2008]

This project is a part of IMVN (image, video et vie numérique) competitiveness pole in the region Ile de France. It aims to develop a framework for advanced multimedia search engine. The main partner is Thalès.

7.1.5. Benchmark Initiative : ImagEval

The ImagEVAL project has a national financial support but benefit of European research lab. participation (Techno-Vision, Ministère de la recherche http://www.imageval.org/presentation.html. It is animated by CEA-LIST and supervised by DGA and Nicephore Cité. It relates to the evaluation of technologies of content-based image retrieval (CBIR) and automatic image annotation in large-scale image databases. It is the first benchmark comparing image search engines using visual content and in a realistic context (images coming from professional content providers and annotated by real archivists). Blank tests were organized in April 2006 and the official 2006 campaign was performed in October. Both reached a real success with more than 10 teams including French academic teams (ENSEA, CEA, Ecole des Mines de Paris,...), European academic teams (Vienna University of Technology, University of Geneva, CERTH in Grece), and also industrial teams (Canon, LTU technologies, AdVestigo).

ImagEVAL includes 5 distinct tasks but IMEDIA participated only to the 3 tasks corresponding to the activities of the team:

- Task 1 : Transformed images retrieval

- Task 4 : Objects and objects class retrieval

- Task 5 : Semantic attributes extraction

The two other tasks deal with textual information detection in images (task 2) and hybrid text-image retrieval in documents including natural language.

IMEDIA did win all the tasks in which the team participated in for both the blank tests and the official tests (first over 6 teams for task 1, first over 3 teams for task 4, first over 6 teams for task 5). Task 1 was processed thanks to a new content-based copy detection technique based on new local descriptors coupled with global geometric constraints and an efficient indexing structure. Task 4 was processed with IMEDIA standards descriptors used in a local and multi-resolution context and processed by several new multiple instances learning strategies. Task 5 was processed with a semantical hierarchy of classifiers based on IMEDIA standards global descriptors and support vector machines. (participants: Nicolas Hervé, Alexis Joly, Nozha Boujemaa and Marin Ferecatu).

7.2. European Initiatives

7.2.1. Integrated European Project "AceMedia"[2004-2007]

"Integrating knowledge, semantics and content for user-centred intelligent media services" in the 6th Framework Program. The consortium of this project is composed of 15 industrial and academic European partners (Alinari, Belgavox, DCU, France Telecom, Fraunhofer, INRIA, ITI, Motorola, Philips, QMUL, Telefonica, Thomson, UAM, UKarlsruhe).

7.2.2. European Network of Excellence "MUSCLE"[2004-2007]

"Multimedia Understanding through Semantics, Computation and Learning" in the 6th Framework Programme. This network of excellence is composed of 42 European academic institutions. Nozha Boujemaa chairs the Workpackage "Single Media Processing" and is deputy scientific coordinator of the network.

7.2.3. European Network of Excellence "DELOS2 [2004-2007]"

"Network of excellence on Digital Libraries" in the 6th Framework Programme. This network of excellence is composed of 44 European academic institutions for the period 2004-2007.

7.2.4. Strep “TRENDS” [2006-2008]

”Trends Research Enabler for Design Specifications” in the 6th Framework Programme. This project is composed of 8 industrial and academic European partners (Robotiker, CRF, Stile Bertone, PertIMM, LPCI-ENSAM, INRIA, University of Leeds, Cardiff University). Its goal is to built an interactive software for the elaboration of design trend boards dedicated to product designers in B to C markets such as for the automotive and original equipment manufacturers. Nozha Boujemaa is workpackage leader of the workpackage "Image Content Description Technology".

7.2.5. Integrated project “VITALAS” [2007-2009]

“Video & image indexing and retrieval in the large scale” in the call6 of 6th Framework Programme. Nozha Boujemaa is the scientific co-ordinator of the project.

7.2.6. Coordination Action “CHORUS” [2007-2009]

Coordination Action in the field of Audio-Visual Search Engines accepted in the call6 of the 6th Framework Programme. Nozha Boujemaa is the scientific co-ordinator of the project.

7.3. International Initiatives

7.3.1. ViMining Project INRIA-NII (INRIA Associated Team Program)[2004-2005]

ViMining is an associated research team composed of IMEDIA group and the team of Pr. Shin’ichi Satoh from the National Institute of Informatics (NII), Japan. The major topics of common interest are : detection and description of semantic video events; organisation of the feature space; cross-media indexing and retrieval. For more information, see http://www-rocq.inria.fr/imedia/vimining/index.html and http://www-direction.inria.fr/international/EQUIPES_ASSOCIEES/index.eng.htm

7.3.2. STIC Project INRIA-Tunisian universities "INISAT"[2003-2006]

This project involves the URISA research team from the school of engineering Sup’Com in Tunis. This project aims at developing unsupervised classification methods in order to segment satellite images and organize visual database indexes.
Information about past and on-going projects are also detailed at http://www-rocq.inria.fr/imedia/projects.html.

8. Dissemination

8.1. Leadership with scientific community

8.1.1. Nozha Boujemaa

- Member of the Steering Board of NEM ETP (Networked and Electronic Media European Technology Platform)representing INRIA ;
- Scientific coordinator of Muscle NoE (Network of Excellence FP6);
- Scientific coordinator of Vitalas IP FP6;
- Scientific coordinator of Chorus CA FP6;
- Co-chairman, with J.Wang, of the "8th ACM SIGMM International Workshop on Multimedia Information Retrieval" in conjunction with ACM Multimedia 2006 (MIR’06)
- Co-chairman, with A. Delbimbo, R. Cucciara and E. Pauwels, of the NoE’s Muscle-Delos Summer School’06 on "Multimedia digital libraries Machine learning and cross-modal technologies for access and retrieval"
- Member of the "National Evaluation Commission" of INRIA.

- PhD Jury member : 3 PhDs (rapporteur: INSA Lyon, rapporteur: INSA Rouen, Univ. Tours) and rapporteur of HDR (Philippe Joly - IRIT Toulouse)
- Member of the JPSearch ad-hoc group (ISO/IEC JTC 1/SC 29/WG 1 - Coding of still pictures)

8.1.2. Michel Crucianu

- Scientific expert for RNTL (French national network for software technologies).
- Coordinator of the BIOTIM project of the French national initiative “Masses of data”.
- Organizer of the e-team “Active Semi-Supervised Learning” in the MUSCLE FP6 NoE.
- Journal reviewer: IEE Proc. on Vision, Image and Signal Processing, IEEE Trans. on Systems, Man and Cybernetics, Multimedia Tools and Applications, Pattern Recognition.
- PhD jury member:
  – Chairman: University of Nantes.
  – Member: University of Paris-Orsay, CNAM Paris.

8.1.3. Nicolas Hervé

- Active member of the JPSearch ad-hoc group (ISO/IEC JTC 1/SC 29/WG 1 - Coding of still pictures) that aims to produce standards to facilitate management, search, and retrieval of content in the form of still pictures.

8.1.4. Anne Verroust-Blondet

- AFIG Vice-President (Association Française d’informatique Graphique) ;
- Member of the steering committee of the CNRS GDR IG (Informatique Graphique) ;
- Member of the editorial board of the “Revue Electronique Francophone d’Informatique Graphique” ;
- Journal Reviewer : IEEE Transactions on Visualization and Computer Graphics ;

8.2. Teaching

8.2.1. Mohamed Chaouch

- 18 hours of TD on signal processing, Master 1 “Electronique et communication”, University of Marne La Vallée.

8.2.2. Michel Crucianu

- Several courses at CNAM Paris (full duty), among which “Pattern recognition and neural networks”, “Machine learning” and “Advanced databases 2: image databases” at the Master level.
- In charge of the course “Multimedia Databases” of the Master in computer science of the University Paris Dauphine.

8.2.3. Valérie Gouet-Brunet

- 192 hours in the Computer Science Department of CNAM;
- National responsible for the course "Computer Vision" of the Master research STIC - Computer Science of CNAM (6 ECTS - 60 hours);
- In charge of the course "Image indexing and retrieval" in the master SAR of the University Paris 6 (7,5 hours).

8.2.4. Nicolas Hervé

- 16h TP on Java Database Connectivity (JDBC), CNAM 3rd year Paris.

8.2.5. Hichem Houissa

- 48 hours TD of Probability and Statistics, Licence 1, Faculté Jean Monnet, University Paris 11.

9. Bibliography

Major publications by the team in recent years

[1] N. BOUJEMAA, J. FAUQUEUR, M. FERECATU, F. . FLEURET, V. GOUET-BRUNET, B. LE SAUX, H. SAHBI. *Ikona: Interactive specific and generic image retrieval*, in "International workshop on Multimedia Content-Based Indexing and Retrieval (MMCBIR'2001)", 2001.

[2] Y. FANG, D. GEMAN, N. BOUJEMAA. *An Interactive System for Mental Face Retrieval*, in "7th ACM SIGMM International Workshop on Multimedia Information Retrieval, Singapore in conjunction with ACM Multimedia 2005", November 10–11 2005.

[3] J. FAUQUEUR, N. BOUJEMAA. *Mental image search by boolean composition of region categor ies*, in "Multimedia Tools and Applications", September 2006, p. 95-117.

[4] M. FERECATU, N. BOUJEMAA, M. CRUCIANU. *Semantic interactive image retrieval combining visual and conceptual content description*, in "ACM Multimedia Systems Journal", to appear, 2006.

[5] M. FERECATU, N. BOUJEMAA. *Interactive Remote Sensing Image Retrieval Using Active Relevance Feedback*, in "IEEE Transactions on Geoscience and Remote Sensing", to appear, 2006.

[6] N. GRIRA, M. CRUCIANU, N. BOUJEMAA. *Active Semi-Supervised Fuzzy Clustering*, in "Pattern Recognition Journal", to appear, 2006.

[7] A. JOLY, O. BUISSON, C. FRÉLICOT. *Content-based Copy Retrieval using Distortion-based Probabilistic Similarity Search*, in "IEEE Transactions on Multimedia", to appear, 2006.

Year Publications

Doctoral dissertations and Habilitation theses

[8] N. GRIRA. *Catégorisation semi-supervisée dans les grandes bases d'images (Semi-Supervised Clustering in Large Image Databases)*, Ph. D. Thesis, Paris-Sud University, June 2006.

Articles in refereed journals and book chapters

[9] J. AMORES, N. SEBE, P. RADEVA. *Context-Based Object-Class Recognition and Retrieval by Generalized Correlograms*, in "IEEE TPAMI", to appear, 2006.

[10] J. FAUQUEUR, N. BOUJEMAA. *Mental image search by boolean composition of region categories*, in "Multimedia Tools and Applications", September 2006, p. 95-117.

[11] M. FERECATU, N. BOUJEMAA, M. CRUCIANU. *Semantic interactive image retrieval combining visual and conceptual content description*, in "ACM Multimedia Systems Journal", to appear, 2006.

[12] M. FERECATU, N. BOUJEMAA. *Indexation multimédia: Description, indexation et recherche des documents multimédias par leur contenu*, P. GROS (editor). , to appear, chap. Description des images fixes, Hermès Science, Traité IC2 (Information, communication et commande), 2006.

[13] M. FERECATU, N. BOUJEMAA. *Interactive Remote Sensing Image Retrieval Using Active Relevance Feedback*, in "IEEE Transactions on Geoscience and Remote Sensing", to appear, 2006.

[14] V. GOUET-BRUNET. *Encyclopédie de l'Informatique et des systèmes d'information*, J. AKOKA, I. COMYN-WATTIAU (editors). , chap. Recherche par contenu visuel dans les grandes collections d'images, Collectif Vuibert, November 2006.

[15] N. GRIRA, M. CRUCIANU, N. BOUJEMAA. *Active Semi-Supervised Fuzzy Clustering*, in "Pattern Recognition Journal", to appear, 2006.

[16] N. GRIRA, M. CRUCIANU, N. BOUJEMAA. *Fuzzy Clustering With Pairwise Constraints for Knowledge-Driven Image Categorisation*, in "IEE Proceedings - Vision, Image, and Signal Processing", vol. 153, n^{o} 3, 2006, p. 299–304.

[17] A. JOLY, O. BUISSON, C. FRÉLICOT. *Content-based Copy Retrieval using Distortion-based Probabilistic Similarity Search*, in "IEEE Transactions on Multimedia", to appear, 2006.

[18] A. VERROUST-BLONDET, M.-L. VIAUD. *Encyclopédie de l'Informatique et des systèmes d'information*, J. AKOKA, I. COMYN-WATTIAU (editors). , chap. Introduction à la synthèse d'images, Collectif Vuibert, November 2006.

[19] J. YU, J. AMORES, N. SEBE, Q. TIAN. *Advances in Imaging and Electron Physics.*, vol. 144, chap. 4: Ranking Metrics and Evaluation Measures., Academic Press, 2006.

Publications in Conferences and Workshops

[20] M. CHAOUCH, A. VERROUST-BLONDET. *Enhanced 2D/3D Approaches Based on Relevance Index for 3D-Shape Retrieval*, in "International Conference on Shape Modeling and Applications 2006 (SMI 2006), Matsushima, Japan", June 2006, p. 241-246.

[21] D. GEMAN. *Interactive Image Retrieval by Mental Matching*, in "Keynote Talk to ACM SIGMM International Workshop on Multimedia Information Retrieval MIR'06", October 2006.

[22] H. HOUISSA, N. BOUJEMAA, H. FRIGUI. *Adaptive Visual Regions Categorization with Sets of Points of Interest*, in "Pacific-Rim Conference on Multimedia, Hangzhou, China", November 2006.

[23] H. HOUISSA, N. BOUJEMAA. *On the use of metrics for multi-dimensional descriptors clustering*, in "IEEE International Conference on Image Processing (ICIP'06), Atlanta, GA, USA", 2006.

[24] H. HOUISSA, N. BOUJEMAA. *Region labelling using a Point-Based Coherence Criterion*, in "IS&T/SPIE conference on Multimedia Content Analysis, Management and Retrieval, San Jose, CA, USA", 2006.

[25] H. HOUISSA, N. BOUJEMAA. *Regions Coherence Criterion based on Points Of Interest Configuration*, in "International Workshop on Image Analysis for Multimedia Interactive Services, Incheon, Korea", 2006.

[26] H. HOUISSA, N. BOUJEMAA. *Spatial configuration modeling for efficient Visual Thesaurus query composition*, in "International Symposium on Image/Video Communications, Hammamet, Tunisia", 2006.

[27] J. LAW-TO, O. BUISSON, V. GOUET-BRUNET, N. BOUJEMAA. *Robust Voting Algorithm Based on Labels of Behavior For Video Copy Detection*, in "ACM Multimedia, MM'06, Santa Barbara, USA", October 2006.

[28] J. LAW-TO, V. GOUET-BRUNET, O. BUISSON, N. BOUJEMAA. *Labeling Complementary Local Descriptors Behavior for Video Copy Detection*, in "International Workshop on Multimedia Content Representation, Classification and Security'06, Istanbul, Turkey", September 2006.

[29] J. LAW-TO, V. GOUET-BRUNET, O. BUISSON, N. BOUJEMAA. *Labellisation du comportement de descripteurs locaux pour la détection de copies vidéo*, in "Compression et Représentation des Signaux Audiovisuels (Coresa'06), Caen", November 2006.

[30] J. LAW-TO, V. GOUET-BRUNET, O. BUISSON, N. BOUJEMAA. *Local Behaviours Labelling for Content Based Video Copy Detection*, in "18th Int. Conf. on Pattern Recognition, ICPR'06, Hong Kong, China", August 2006.

[31] A. REBAI, A. JOLY, N. BOUJEMAA. *Constant Tangential Angle Elected Interest Points*, in "ACM SIGMM International Workshop on Multimedia Information Retrieval (MIR'06)", 2006.

[32] I. YAHIAOUI, N. HERVÉ, N. BOUJEMAA. *Shape-Based Image Retrieval in Botanical Collections*, in "Pacific-Rim Conference on Multimedia, Hangzhou, China", November 2006.

[33] J. YU, J. AMORES, N. SEBE, Q. TIAN. *A New Study on Distance Metrics as Similarity Measurement*, in "IEEE International Conference on Multimedia and Expo (ICME 2006), Toronto, Canada", July 2006.

[34] J. YU, Q. TIAN, J. AMORES, N. SEBE. *Toward Robust Distance Metric Analysis for Similarity Estimation*, in "IEEE International Conference on Computer Vision and Pattern Recognition (CVPR 2006)", vol. 1, 2006, p. 316-322.

Internal Reports

[35] O. BESBES, Z. BELHADJ, N. BOUJEMAA. *Adaptive Satellite Images Segmentation by Level Set Multiregion Competition*, Technical report, n^o RR-5855, INRIA Rocquencourt, March 2006, http://hal.inria.fr/inria-00070171.

[36] J. LAW-TO, V. GOUET-BRUNET, O. BUISSON, N. BOUJEMAA. *Labelling the Behaviour of Local Descriptors for Selective Video Content Retrieval*, Technical report, n^o RR-5821, INRIA Rocquencourt, January 2006, http://hal.inria.fr/inria-00070204.

Miscellaneous

[37] A. REBAI. *Étude et intégration d'un nouveau détecteur de points d'intérêt pour la recherche d'images par le contenu visuel*, June 2006, Mémoire de fin d'études, École Supérieure des Communications de Tunis (SUP'COM).

References in notes

[38] A. BARBU, S. ZHU. *Generalizing Swendsen-Wang to Sampling Arbitrary Posterior Probabilities.*, in "IEEE Trans. Pattern Anal. Mach. Intell.", vol. 27, n^o 8, 2005, p. 1239–1253.

[39] S. BERRETTI, M. CHAOUCH, R. CUCCHIARA, A. DEL BIMBO, M. DAOUDI, C. GRANA, P. PALA, J.-P. VANDEBORRE, A. VERROUST-BLONDET, M. WORRING. *Description, Matching and Retrieval by Content of 3D Objects*, August 2006, DELOS Research ACtivities 2006.

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

[41] N. BOUJEMAA, O. BUISSON, M. FERECATU, N. HERVÉ, A. JOLY, J.-M. RENDERS, A. SAULNIER, M.-L. VIAUD. *Etat de l'art des travaux actuels et Objectifs scientifiques*, July 2006, INFOM@GIC, document D1.4-4, tâche T1.4 "Recherche multimedia".

[42] N. BOUJEMAA, J. FAUQUEUR, M. FERECATU, F. FLEURET, V. GOUET-BRUNET, B. LE SAUX, H. SAHBI. *Ikona: Interactive specific and generic image retrieval*, in "International workshop on Multimedia Content-Based Indexing and Retrieval (MMCBIR'2001)", 2001.

[43] J. FAUQUEUR, N. BOUJEMAA. *New Image Retrieval Paradigm: Logical Composition of Region Categories*, in "IEEE International Conference on Image Processing (ICIP'03)", 2003, p. 601-604.

[44] M. FERECATU, N. BOUJEMAA, M. CRUCIANU. *Active Relevance Feedback for Interactive Satellite Images retrieval*, in "Proc. of ESA-EUSC'05: Image Information Mining - Theory and Application to Earth Observation", October 2005.

[45] F. FLEURET. *Détection hiérarchique de visages par apprentissage statistique*, Ph. D. Thesis, Université Paris-VI, Paris, 2000.

[46] H. GOËAU, A. SAULNIER, M.-L. VIAUD, L. DENOYER, A. VERROUST-BLONDET, S. LEROY. *D1.12-4 Etat de l'art sur les stratégies de consultation*, June 2006, INFOM@GIC document D1.12-4, sous-tâche ST1.12 "Stratégie de consultation".

[47] N. GRIRA, M. CRUCIANU, N. BOUJEMAA. *Active semi-supervised fuzzy clustering for image database categorization*, in "7th ACM SIGMM international workshop on Multimedia Information Retrieval (MIR '05), New York, NY, USA", ACM Press, 2005, p. 9–16.

[48] N. GRIRA, M. CRUCIANU, N. BOUJEMAA. *Semi-supervised fuzzy clustering with pairwise-constrained competitive agglomeration*, in "The IEEE International Conference on Fuzzy Systems, Fuzz'IEEE 2005", May 2005, p. 80–86.

[49] A. K. JAIN, R. C. DUBES. *Algorithms for clustering data*, Prentice-Hall, Inc., 1988.

[50] A. JOLY, C. FRELICOT, O. BUISSON. *Feature statistical retrieval applied to content-based copy identification*, in "IEEE International Conference on Image Processing (ICIP'04)", vol. 1, October 2004, p. 681 - 684.

[51] T. LEE, D. MUMFORD. *Hierarchical Bayesian inference in the visual cortex.*, in "Journal of Optical Society of America, A. .", vol. 20, n^o 7, 2003, p. 1434–1448.

[52] J. SIVIC, A. ZISSERMAN. *Video Google: A Text Retrieval Approach to Object Matching in Videos*, in "International Conference on Computer Vision", vol. 2, October 2003, p. 1470–1477.

[53] J. THIÈVRE, M.-L. VIAUD, A. VERROUST-BLONDET. *Using Euler Diagrams in Traditional Library Environment*, in "Electronic Notes in Computer Science", vol. 134, June 2005, p. 189–202.

[54] R. C. VELTKAMP, R. RUIJSENAARS, M. SPAGNUOLO, R. VAN. ZWOL, F. TER. HAAR. *SHREC2006: 3D Shape Retrieval Contest*, Technical report, n^{o} UU-CS-2006-030, Institute of Information and Computing Sciences, Utrecht University, 2006.

[55] M.-L. VIAUD, A. SAULNIER, J. THIÈVRE, A. VERROUST-BLONDET, P. LAVAL, S. LEROY. *Etat de l'art des différents composants - Consultation*, June 2006, INFOM@GIC, document D1.12-3, sous-tâche ST1.12 "Stratégie de consultation".