



INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

*Project-Team Imedia*

*Images and Multimedia : Indexing,  
Retrieval and Navigation*

*Rocquencourt*

THEME COG

*Activity*  
*R*  
*Report*

2005



## Table of contents

<b>1. Team</b>	<b>1</b>
<b>2. Overall Objectives</b>	<b>1</b>
2.1. Overall Objectives	1
<b>3. Scientific Foundations</b>	<b>2</b>
3.1. Introduction	2
3.2. Modeling, construction and structuring of the feature space	2
3.3. Pattern recognition and statistical learning	3
3.3.1. Statistical learning and object detection	3
3.3.2. Clustering methods	4
3.4. Interactive search and personalization	4
3.5. Cross-media indexing	5
<b>4. Application Domains</b>	<b>5</b>
4.1. Application Domains	5
<b>5. Software</b>	<b>6</b>
5.1. IKONA/MAESTRO Software	6
<b>6. New Results</b>	<b>7</b>
6.1. Construction and organization of the visual feature space	7
6.1.1. Feature spaces structuring	7
6.1.1.1. Behaviour of SR-trees with points of interest distributions	7
6.1.1.2. Multiple queries	8
6.1.2. Visual saliency	9
6.1.3. Silhouette shape descriptor for image retrieval	10
6.1.4. Adaptive satellite images segmentation by level set multiregion competition	12
6.1.5. 3D indexing : 2D/3D shape descriptors	12
6.1.6. Geometric consistency of local descriptors	14
6.1.7. Discriminant local features selection	15
6.2. Interactive cross-modal retrieval	15
6.2.1. Region labelling using a point-based coherence criterion	15
6.2.2. Cross-modal image retrieval with active relevance feedback	16
6.2.3. Semantic cartography of a database from a user's query	19
6.3. Clustering and learning	19
6.3.1. Active semi-supervised clustering	19
6.3.1.1. Ambiguousness criterion	21
6.3.1.2. Non-redundancy criterion	22
6.3.1.3. Experimental evaluation	22
6.3.2. Kernels for similarity learning	23
6.3.2.1. The intermediate matching kernel	23
6.3.2.2. The LCCP for the selection of kernel parameters	24
6.3.2.3. Kernel PCA for similarity invariant shape recognition and its application to leaf species	25
6.4. Software	27
6.4.1. IKONA/MAESTRO software	27
6.4.2. aceMedia software integration	27
<b>7. Other Grants and Activities</b>	<b>27</b>
7.1. National Initiatives	27
7.1.1. Industrial contract with INA [2004-2007]	27

---

7.1.2.	BIOTIM Project (exploiting Text-IMage ressources in BIODiversity) within the national initiative "Masses of data"[2003-2006]	27
7.1.3.	QuerySat Project within the national initiative "ACI Masses de Données")[2004-2007]	27
7.1.4.	InfoMagic[2005-2008]	27
7.2.	European Initiatives	27
7.2.1.	Integrated European Project "AceMedia"[2004-2007]	27
7.2.2.	European Network of Excellence "MUSCLE"[2004-2007]	28
7.2.3.	European Network of Excellence "DELOS2 [2004-2007]"	28
7.3.	International Initiatives	28
7.3.1.	ViMining Project INRIA-NII (INRIA Associated Team Program)[2004-2005]	28
7.3.2.	STIC Project INRIA-Tunisian universities "INISAT"[2003-2006]	28
<b>8.</b>	<b>Dissemination</b>	<b>28</b>
8.1.	Leadership with scientific community	28
8.1.1.	Nozha Boujema	28
8.1.2.	Valérie Gouet-Brunet	29
8.1.3.	Nicolas Hervé	29
8.1.4.	Jean-Philippe Tarel	29
8.1.5.	Anne Verroust-Blondet	29
8.2.	Teaching	29
8.2.1.	Sabri Boughorbel	29
8.2.2.	Valérie Gouet-Brunet	30
8.2.3.	Jean-Philippe Tarel	30
8.2.4.	Anne Verroust-Blondet	30
<b>9.</b>	<b>Bibliography</b>	<b>30</b>

# 1. Team

## Team leader

Nozha Boujemaa [Research Director (DR2)]

## Administrative assistant

Laurence Bourcier [shared with Salsa and Micmac project-team]

## Staff members

Anne Verroust-Blondet [Research Scientist (CR1)]

Jean-Paul Chièze [Senior Technical Staff, half-time]

## External collaborators

Olivier Buisson [Research Scientist at Institut National de l'Audiovisuel]

Michel Crucianu [Professor at CNAM]

Valérie Gouet-Brunet [Assistant Professor at CNAM]

Jean-Philippe Tarel [Research Scientist (CR1) at LCPC]

## Invited Professor

Peter Belhumeur [Professor at Columbia University, USA (January to July 2005)]

Donald Geman [Professor at Johns Hopkins University, USA (February, March, April 2005)]

Masashi Inoue [Assistant Professor at NII (National Institute of Informatics) Japon (3rd October to 1st November 2005)]

## Technical staff

Mathieu Coutaud [Junior Technical Staff]

Nicolas Hervé [Technical Staff]

Alexis Joly [Technical Staff]

## Post-doctoral Fellow

Sabri Boughorbel [ACET Scientifique]

Marin Ferecatu [ACET Scientifique]

Ithéri Yahiaoui [ACET Scientifique]

## Ph. D. Student

Olfa Besbes [Joint tutorship with Sup'Com, national grant since 1/9/2004]

Mohamed Chaouch [INRIA grant, Télécom Paris]

Nizar Grira [INRIA grant, Orsay University Paris XI]

Hichem Houissa [INRIA grant, Orsay University Paris XI]

Julien Law-To [CIFRE grant with INA since 1/5/2004]

# 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 specialties. The main ones are image analysis, pattern recognition, statistical learning, human-machine interaction and database systems.

Our work is structured into the following main themes :

1. Image indexing : this part mainly concerns modeling 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 personalization : 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, personalization and cross-media indexing. In the following we give a short introduction to each of these themes.*

### 3.2. Modeling, 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, Valérie Gouet-Brunet, Ithéri Yahiaoui, Hichem Houissa, Sabri Bougorbel, Jean-Philippe Tarel, Nizar Grira, Anne Verroust-Blondet, Hichem Sahbi, 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 (color, 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.

Also, since the arrival of Anne Verroust-Blondet, we have been investigating the problem of 3D model description, in order to complete our approach of the description of the visual appearance in 2D and 3D.

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. A collaboration on this topic is under way with Michel Scholl (INRIA/CNAM).

### 3.3. Pattern recognition and statistical learning

Statistical learning and classification methods are of central interest for content-based image retrieval [39] [48].

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 *labeled 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, algorithmic optimization, kernel methods.*

**Participants:** Donald Geman, Hichem Sahbi, Sabri Boughorbel, Jean-Philippe Tarel.

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. 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. The computational efficiency we are looking for has the effect of a regularization constraint : it favors structurally simple classifiers, which have good generalization properties.

Beside this work focusing on the trade-off between error rate and computational cost, we are working on the design of invariant kernels for vision. We have worked on the scale invariance of kernel methods based on the triangular kernel, and we have unified kernel methods and the matching of points of interest by designing matching kernels.

These high invariance of matching schemes to the view-based representation underlying support vector machines or other kernel methods.

### 3.3.2. Clustering methods

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

**Participants:** Nozha Boujemaa, Nizar Grira, Michel Crucianu, Hichem Sahbi, 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 personalization

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

**Participants:** Marin Ferecatu, Yuchun Fang, Donald Geman, Nozha Boujemaa, Michel Crucianu, Hichem Houissa, Jean-Paul Chièze, Jean-Philippe Tarel.

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

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

**Participants:** Marin Ferecatu, Valerie Gouet-Brunet, Nozha Boujemaa, Michel Crucianu, Hichem Sahbi.

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 : encyclopaedic 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:** Nicolas Hervé, Mathieu Coutaud, Jean-Paul Chièze, Marin Ferecatu.

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 [40]. 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 characterized 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 [40];

- **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 developped 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 fonctionnalities. More screenshots 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. Feature spaces structuring

**Keywords:** *high-dimensional feature spaces, image index management, local descriptors, multi-dimensional indexing.*

**Participants:** Nouha Bouteldja, Valérie Gouet-Brunet, Michel Scholl.

Several categories of image descriptors are studied in the IMEDIA group. Some of them, the global descriptors in particular, allow the interrogation of high-dimensional databases (about 500,000 images) in real-time with standard hardware. Other descriptors, such as the local descriptors involving points of interest, currently only allow the interrogation of small databases (about 3,000 images). Our objective is to fit to scale the descriptors developed at IMEDIA. For the moment, we focus on the local approaches, that do not allow real-time responses for the databases encountered in various applications.

##### 6.1.1.1. Behaviour of SR-trees with points of interest distributions

This year, our first contribution was to experimentally assess whether the so-called *curse of dimensionality* phenomenon [66] is reached with various categories of distributions related to dimensions encountered with local descriptors (usually between 8 and 30). We compare the performance of a single sphere query when the collection is indexed by a tree structure (an SR-tree in our experiments) to that of a sequential scan. The tested distributions were :

1. synthetic clustered distributions ( $D_C$ )
2. synthetic uniform distributions ( $D_U$ )
3. real distributions ( $D_R$ , obtained by an extraction and a local description of points of interest from a set of generalist images).

Some of the experiments realized are illustrated in figure 1; it shows the ratio of the sequential scan CPU time over the CPU time obtained with the SR-tree traversal, versus the number of neighbors for  $D_R$  and  $D_C$ .

Clearly, for the real dataset as well as for the clustered distribution, the curse of dimensionality is not reached. The speed up for dimension  $d = 29$  is significant even for large values of the number of neighbors. In contrast, as experimented earlier in the literature, there is no gain with the uniform data set (not shown here). These trends were confirmed by the study of the ratio of nodes acceded during the tree traversal. This ratio remains small for the  $D_R$  and  $D_C$  distributions. All the experiments performed show that when the dimension of the space is moderate, i.e. between 8 and 30, the tree structure indeed performs well : it exhibits a significant gain wrt to sequential scan even in a 29-dimensional space.

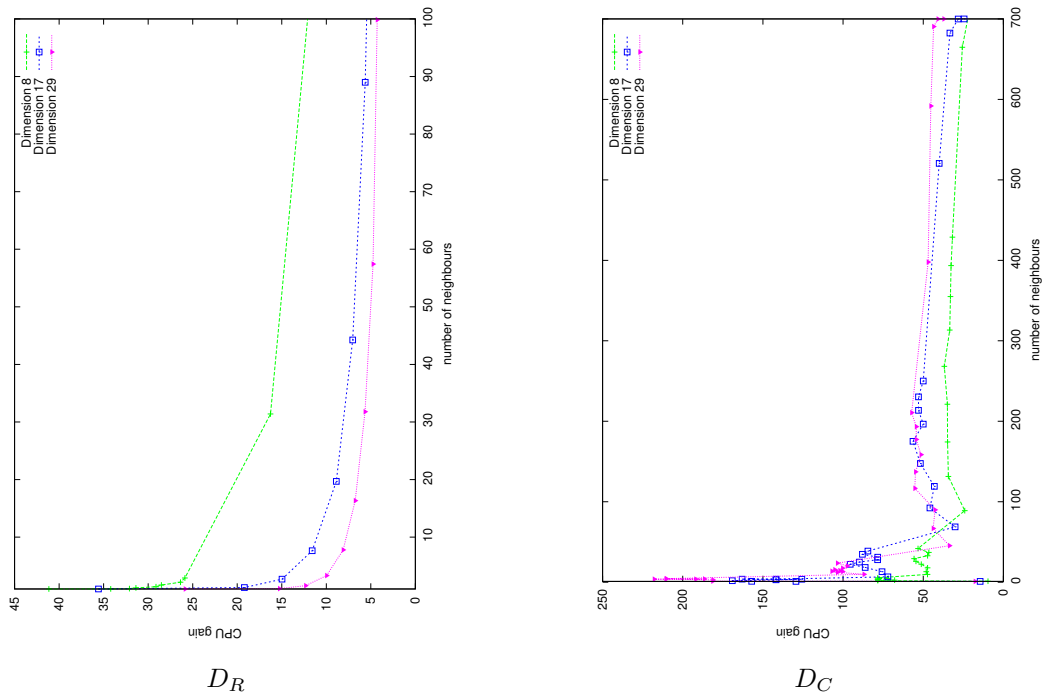


Figure 1. Sequential scan/SR-Tree gain versus the number of neighbors for the datasets  $D_R$  and  $D_C$ .

#### 6.1.1.2. Multiple queries

When considering objects or parts of image described with a set of local descriptors, searching in the feature space is usually done independently and sequentially for each local descriptor. We have studied *multiple queries* approaches [41], [38] existing in the Database community, before applying and adapting them to the retrieval of groups of local descriptors. Two directions have been investigated :

- Reduction of the I/O costs, by studying a new approach for searching in the multidimensional structure. A lot of multidimensional structures exist [4] ; the structure considered is the SR-Tree one, which achieves good performances for feature spaces based on local descriptors (see previous section);
- Reduction of the CPU costs, by considering relations existing on distances computed between several query points and points of the feature space, for instance the triangular inequality. The structure considered is also the SR-tree one, but the proposed improvements could be applied to any tree structure. We have revisited the two lemmas proposed in [41] and proposed three novel ones.

Figure 2 illustrates our approach : it presents CPU time consumed for similarity search with multiple queries, by using or not some of the lemmas. The evaluation has been done within a dataset of 1 million of real points of interest. The results show that the joint use of such lemmas allows to accelerate the search, whatever the dimensions.

All these studies are done in collaboration with the CEDRIC/Vertigo research group, during the thesis of Nouha Bouteldja in the CEDRIC laboratory, and were accepted for publication in an international conference [14].

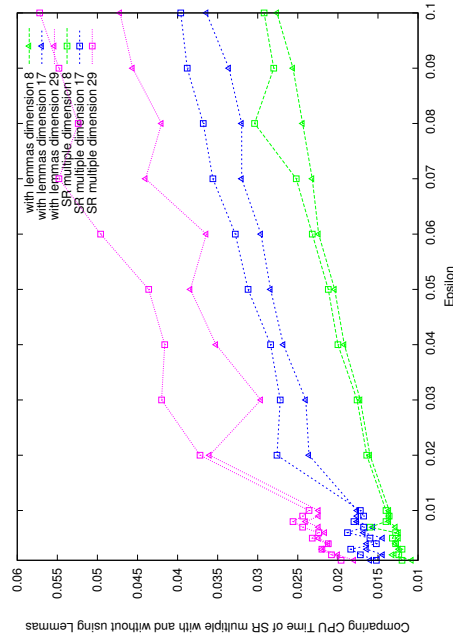


Figure 2. CPU time average without (rectangle lines) and with (triangle lines) some of the lemmas used jointly.

### 6.1.2. Visual saliency

**Keywords:** Local descriptors, real-time monitoring, spatio-temporal descriptors, visual saliency.

**Participants:** Julien Law-To, Olivier Buisson, Valérie Gouet-Brunet, Nozha Boujemaa.

This work is done in collaboration with the INA (French National Institute of Audiovisual), within the scope of the CIFRE thesis of Julien Law-To. The main application considered is real-time monitoring of huge databases of videos (about 100000 hours). At present, we focus more particularly in video content-based copy detection (CBCD).

In order to protect the INA from the piracy of its videos, we have to link the TV broadcast on one side to the video database on the other side (monitoring) in order to find identical sequences. The CBCD system must be robust to common transformations used in the TV post-production such as zooming, cropping, shifting, etc.

Local descriptors have been proved to be very useful for image indexing with application to object or sub-image retrieval. In the Computer Vision community, a number of recent techniques have been proposed to identify points of interest or regions of interest in images. If directly applied to image sequences, one of the drawbacks of such descriptors is its spatio-temporal redundancy. When considering applications like real-time monitoring, it is necessary to build a very compact description. In this context, two kinds of promising approaches can be investigated :

1. The spatio-temporal information can be exploited jointly, instead of working on the time dimension and the spatial dimension separately. See for example the work of Laptev on space-time interest points [58] ;
2. To be more compact and more significant, the image description should involve new signatures that should be inspired by pre-attentive human vision and the focalization of attention. Such features have been widely studied by the neurophysiological community and can be modeled mathematically. See for example the work of Heidemann [53] on focus-of-attention from local color symmetries.

This year, we have investigated the first class of approaches. Studying the trajectories of the points of interest is an efficient solution to avoid temporal redundancy and, moreover an interesting way to strongly fingerprint the video sequence. By modeling and labelling the different kinds of behaviours of those points according to their trajectories, the CBCD system can be improved. Such a content-based video description is richer and more compact than the usual ones. It is also generic then allows to investigate new possibilities for video retrieval, such as :

- Specific queries on video sequences, like object retrieval with a particular behaviour;
- Segmentation of the video based on the points of interest behaviour (camera motion, cut, credits, ...).

Based on this description, the current experiments show an improvement on the quality of the monitoring process and on its flexibility. This work has been submitted to the international conference CVPR 2006.

### 6.1.3. *Silhouette shape descriptor for image retrieval*

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

**Participants:** Itheri Yahiaoui, Nozha Boujemaa, Nicolas Hervé.

Within the BIOTIM project, we have designed a new shape descriptor called Directional Fragment Histogram (DFH) [34], [33]. This shape descriptor is computed using the outline of the plant. Both local and global directional information are coded. Local information are computed from segments taken randomly from the external contour. The accumulation of all possible segments of a specific length provides the global information. Several versions of this descriptor are studied and tested on different image databases. The implemented DFH versions are based respectively on the freeman chain-codes, the angles of the gradient, the curvatures and the combination of both angles of the gradient and curvatures. In our evaluation, the following two kinds of image database are used :

- Botanical databases
  - The Arabidopsis database provided by the INRA Institute. It consists of about 400 images.
  - The Swedish leaves database taken from the web site of the "the Swedish Museum of Natural History" . This database is composed of 1125 images and 15 classes, each classe represents a tree species.
  - The Smithsonian leaves database provided from the collaboration with the NSF projet "An Electronic Field Guide : Plant Exploration in the 21st Century". It contains 134 classes including 1520 images : see <http://www.cfar.umd.edu/gaaga/leaf/leaf.html>.
- Generic Objects databases
  - The Mpeg7 CE-Shape-1database taken from the web page of Professor Latecki : <http://www.cis.temple.edu/latecki/research.html>, it contains 1400 images and 70 classes.
  - The Kimia databases is composed of three subdatasets. They contain respectively 99, 256 and 1032 silhouette images : see <http://www.lems.brown.edu/vision/software/index.html>.
  - The ETH-80 database includes 3280 images classified into 80 objects : see <http://www.vision.ethz.ch/projects/categorization/eth80-db.html>.

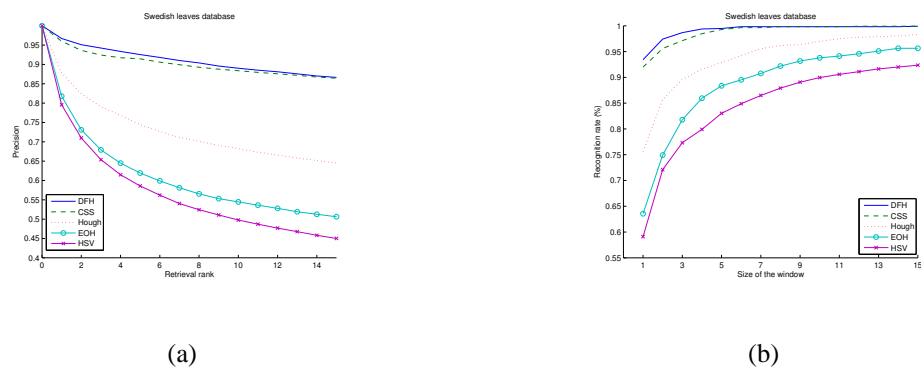


Figure 3. Precision and Roc curves

We evaluate the DFH gradient version as well as some existing shape descriptors on the Swedish database. These shape descriptors are IKONA shape descriptor based on the Hough transform [3], Edge Orientation Histogram (EOH) [57], MPEG-7 CSS descriptor [60] and CCH descriptor [56]. The figures 3 represent respectively precision and ROC graphs.

As illustrated in these graphs, our DFH shape descriptor is more performant than the others. Noting that the CSS-MPEG7 descriptor gives comparable performances. However, its computation time is about 250 times compared to the DFH. Both CCH and EOH present lower results because they focus only on the global shape without taking into account local distribution of directions. The Figure 4 shows "partial queries" results on two leaves taken from the swedish database.

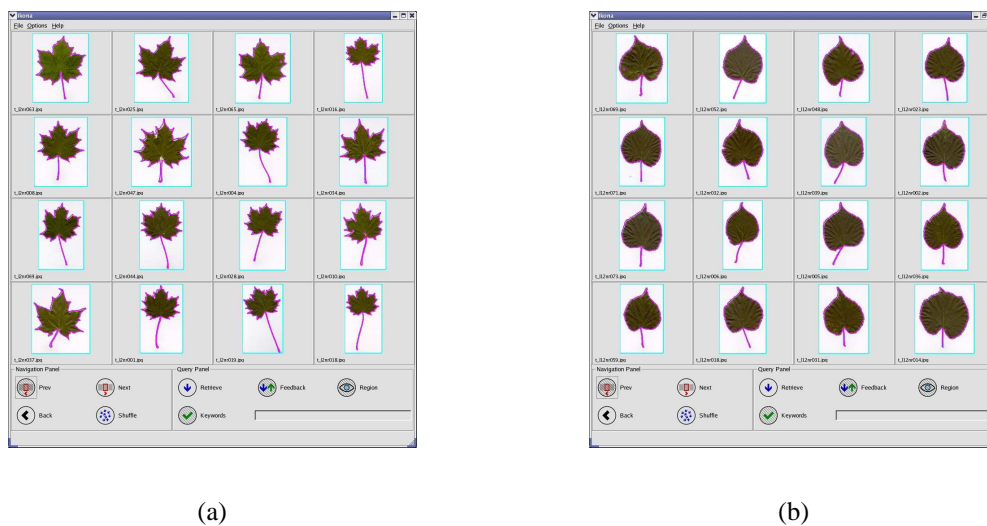


Figure 4. Two Examples of similarity-based retrieval on the swedish database using the proposed shape descriptors

The performances of the different versions of DFH are depending on shape database content. In future work, we intend to improve our DFH descriptor to handle this limitation.

#### 6.1.4. 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 [Unité URISA, École Supérieure des Communications de Tunis (SUP'COM), Cité Technologique des Communications de Tunis, 2088 Tunisie.].

The development and application of various remote sensing platforms result in the production of huge amounts of satellite image data. Therefore, there is an increasing need for effective querying and browsing in these image databases. 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 segmented into a number of regions. Segmentation consists in partitioning the image into non-overlapping regions that are homogeneous with regards to some characteristics such as spectral and texture. Remote sensed images contain both textured and non-textured regions. This is even more true today with high resolution images such as IKONOS, SPOT-5 and QuickBird data. In order to cope with this content heterogeneity, we propose an adaptive variational segmentation algorithm<sup>1</sup> using active curves evolution via level set [35]. It is based on combining cues of spectral and texture according to their discrimination power for each region. The feature vector consists of spectral channels and 4 texture features given by the second order matrix and TV flow based local scale feature where a coupled nonlinear diffusion was applied [42]. Motivated by Fisher-Rao's linear discriminant analysis, we define two region's weights which code respectively the relevance of spectral and texture cues. Therefore, regions with or without texture are processed in the same framework. The obtained segmentation criterion is minimized via curves evolution within an explicit correspondence between the interiors of evolving curves and regions of segmentation. Thus, an unambiguous segmentation to a given arbitrary number of regions is obtained by the multiregion competition algorithm [61]. Experimental results on various panchromatic and multispectral images are shown in figures 5, 6 and 7. In a future work, we intend to use the constructed nonlinear scale-space to provide a multiscale satellite images segmentation. Our aim is to decompose satellite image content into a hierarchy of attributed regions describing semantically topological relations and properties.

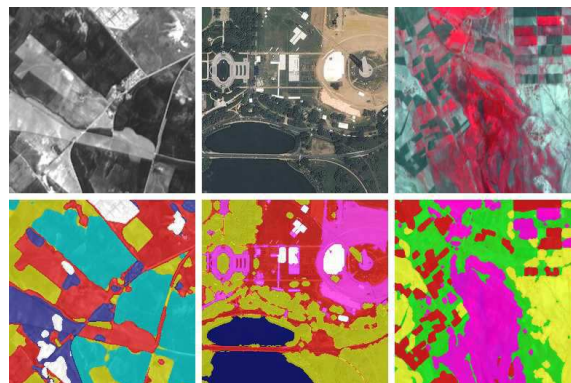


Figure 5. First row : Original images without textured regions. Second row : Segmentation results superimposed on original images.

#### 6.1.5. 3D indexing : 2D/3D shape descriptors

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

<sup>1</sup>This work is partially supported by QuerySat project and INRIA STIC project.



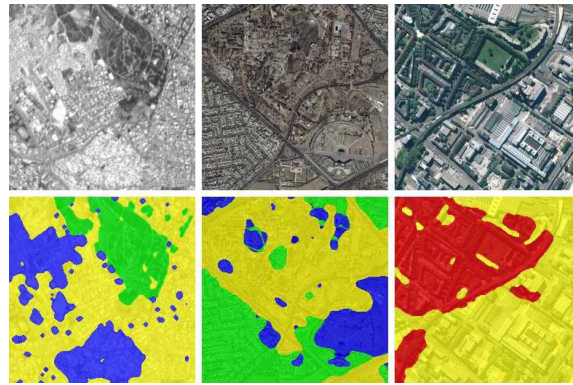


Figure 6. First row : Original images with only textured regions. Second row : Segmentation results superimposed on original images.

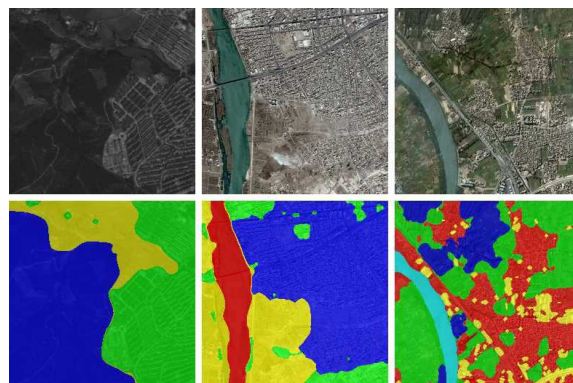


Figure 7. First row : Original images containing both non-textured and textured regions. Second row : Segmentation results superimposed on original images.

**Participants:** Mohamed Chaouch, Anne Verroust-Blondet.

We have implemented and compared the efficiency of several existing 2D/3D shape descriptors, i.e. descriptors built from 2D views of 3D models. We have improved the efficiency of the descriptors based on silhouettes or depth-buffer images. The 2D shape signatures are extracted from classical Fourier Transform (1DFFT for silhouettes and 2DFFT for depth-buffer images). We have introduced a new approach based on relevance index, which takes into account the diversity of information contained in the projection images. It was evaluated on the Princeton database (907 models) and these retrieval results (cf. Figure 8) show its performance and robustness in the 3D-model retrieval process. This work is described in [36] and is supported by the European Network of Excellence DELOS II within "Description, Matching and Retrieval by Content of 3D Objects" Project (Task 3.8 of JPA2).



















Without relevance index			With relevance index		
					
✓ 0.0518	✓ 0.0546	✓ 0.0590	✓ 0.0099	✓ 0.0099	✓ 0.0105
					
✓ 0.0603	✓ 0.0623	✗ 0.0625	✓ 0.0107	✓ 0.0113	✓ 0.0114
					
✗ 0.0666	✗ 0.0672	✗ 0.0673	✗ 0.0116	✗ 0.0117	✓ 0.0118

Figure 8. Retrieval result from the "desk chair" class.

### 6.1.6. Geometric consistency of local descriptors

**Keywords:** *geometric consistency, geometry, image retrieval, local descriptors.*

**Participants:** Alexis Joly, Valérie Guet-Brunet.

There is an increasing variety of content-based image retrieval scenarios involving the use of local descriptors (object class recognition [45], object and scene recognition [59], content-based copy detection [24]). Enhancing the performance of these scenarios by using the geometric distribution or the relative positions of the local descriptors is an active research area. During this year, we have shown that in the copy detection scenario, the robust estimation of a global geometric transformation model after the search is widely profitable to improve the discrimination of the detection (paper submitted to IEEE transactions on

multimedia). However, for other scenarios, using the geometry remains a challenging task : Including the geometric distribution in the descriptor itself often leads to a lack of robustness during the search of similar local descriptors whereas post-processing techniques are generally highly time consuming and thus limited to very small data sets. Moreover, in most of them, the geometric consistency is limited to rigid transformation models which do not allow to enforce the matching when two geometric distributions are dependent but not linearly linked. For a few months, we are investigating the use of non parametric geometric consistency measurements such as mutual information and robust correlation ratio and we plan to combine them with some robust local geometric properties that could be included in the descriptor itself in order to limit the number of matches during the second step.

### 6.1.7. Discriminant local features selection

**Keywords:** *density estimation, discriminant, image retrieval, local features.*

**Participants:** Alexis Joly, Nozha Boujemaa, Shin'ichi Satoh.

This work was done in collaboration with the NII (National Institute of Japan) within the scope of the visit of Alexis Joly at the NII (july 2005). Local features are well-suited to content-based image retrieval because of their locality, their local uniqueness and their high information content [62]. However, as they are selected only according to the local information content in the image, there is no guaranty that they will be distinctive in a large set of images. A local feature corresponding to a high saliency in the image can be highly redundant in some specific databases, such as the TV news database stored at NII in which textual characters are extremely frequent. To overcome this issue, we propose [23] to select relevant local features directly according to their discrimination power in a specific set of images. By computing the density of the local features in a source database with a new fast non parametric density estimation technique, it is indeed possible to select quickly the most *rare* local features in a large set of images. Figure 9 illustrates the difference between the 20 most salient points of an image and the 20 most rare points according to their density in a large image database.



Figure 9. left : 20 most salient points - 20 most rare points

## 6.2. Interactive cross-modal retrieval

### 6.2.1. Region labelling using a point-based coherence criterion

**Keywords:** *Visual thesaurus, region labelling.*

**Participants:** Hichem Houissa, Nozha Boujemaa.

Query By Visual Thesaurus is 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 obtained by means of region categorization [46] into visual patches that stand for the database regions representatives.

We introduce 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 [21]. This dual region/points description strengthens the "page zero" construction process and is not computationally expensive since PCC is evaluated on single coarse regions using Harris color points detector that catches the local photometric variability on small sites (few pixels). The use of points of interest was motivated by their ability to finely describe coarsely segmented region. Thus our approach tends to add semantic knowledge information on low level features by labelling rough regions into homogenous and textured classes. The novelty of our approach is the unsupervised region categorization and the construction of a visual region summary that handles both photometric attributes and point-based description.

Gouet-Brunet and Boujema [52] proposed a color version of the Harris detector for image description. The idea of the Harris detector [54] is that interest points are local maxima related to the second moment matrix  $\mu(x, \sigma_I, \sigma_D)$  (1) and the gradient distribution in a local neighborhood of a point  $x$ .

$$\mu(x, \sigma_I, \sigma_D) = g(\sigma_I) * \begin{bmatrix} L_x^2(x, \sigma_D) & L_x L_y(x, \sigma_D) \\ L_x L_y(x, \sigma_D) & L_y^2(x, \sigma_D) \end{bmatrix} \quad (1)$$

where  $\sigma_I$  is the integration scale,  $\sigma_D$  is the derivation scale,  $g$  the Gaussian and  $L$  the image smoothed by a Gaussian for each channel  $i \in \{R, G, B\}$ .

As far as points of interest are concerned, a homogenous region is more likely to contain less points than a textured one because points catch the local photometric variability around a very small site. As a matter of fact, assumed homogenous patches should be well described by means of classical color moments, whereas textured or non homogenous regions are better characterized by points attributes.

Let  $P = \{P_k, k \in \{1, \dots, i\}\}$  be the set of points of interest detected on the region  $R_j$ . We superimpose a grid  $Cell = \{Cell_{(i,j)}\}$  on the region.

The idea of PCC is to compute the number of effective points contributing to describe a texture from those that are only due to coarse segmentation. To do so, we build a point histogram where bin  $(i, j)$  reflects the number of interest points located in the  $(i, j)$ -th cell (see equation 6.2.1).

$$bin(i, j) = \sum_k \delta_{ijP_k}$$

$$\delta_{ijP_k} = 1 \text{ if } P_k \in Cell_{(i,j)}$$

Once all points visited, the histogram is *binarized* i.e. all filled cells are switched on, empty others are considered switched off (Figure 10). Each cell is visited once so that we keep only those containing at least one point of interest (*effective cell*)(2).

$$bin(i, j) = \begin{cases} 1 & \text{if } bin(i, j) \geq 0 \\ 0 & \text{else} \end{cases} \quad (2)$$

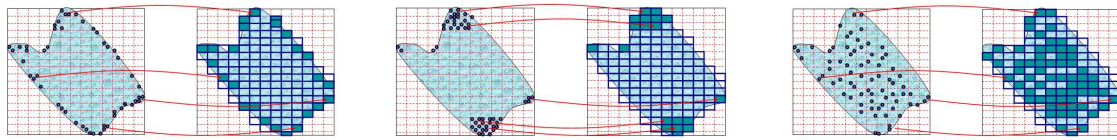
For visual thesaurus construction, we started with regions clustering using mean colors descriptors. The resulting first level of classification brings together all visually similar regions into a set of representatives. This step is following by PCC computation and labelling for all classes. PCC highly discriminates between different regions structure labelling them either with *homogenous* or *textured*. Figure 11 shows, for example, three classes obtained by mean color classification; each class is divided into 2 classes using PCC : homogenous and textured [22].

### 6.2.2. Cross-modal image retrieval with active relevance feedback

**Keywords:** *geometric consistency, geometry, image retrieval, local descriptors.*

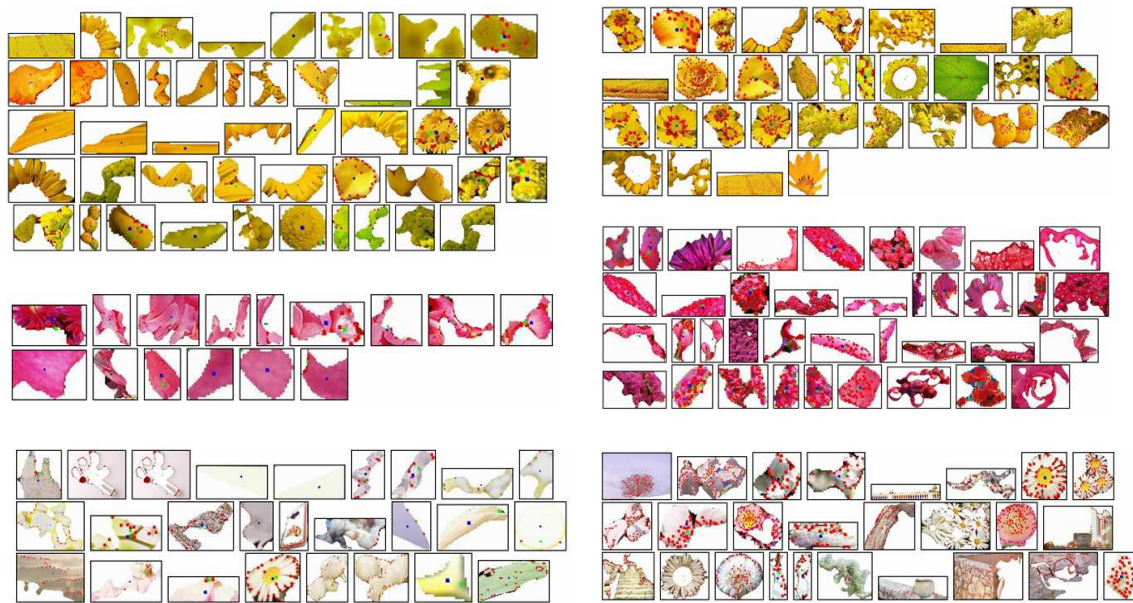
**Participants:** Marin Ferecatu, Nozha Boujema, Michel Crucianu.

Many of the available image databases have keyword annotations associated with the images. As keywords and visual features provide complementary information, using both sources of information is an advantage in



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



(a) Examples of homogenous labelled regions : they encapsulate perceptually homogenous regions as well as those resulting from points accumulation over borders and grouping on small details (b) Examples of textured labelled regions : they gather textured patches and fully covered regions denoting of an important photometric variation

Figure 11. Region subclasses labelling using PCC over mean-color clustered patches

many applications. We address the challenge of semantic gap reduction using a hybrid visual and conceptual representation of the content within an active relevance feedback context. In [17] we introduce a new feature vector, based on the keyword annotations available for the images, which makes use of conceptual information extracted from an external lexical database<sup>2</sup>. To obtain a scalable solution for representing sets of keywords as reliable and comprehensible feature vectors, we suggest to select a limited set of *core concepts*, and to associate to every such concept a dimension in the feature vector.

We rely on an external ontology, defining semantic relations between concepts, to find good candidates for the core concepts and to define the feature vectors for sets of keywords. WordNet is a well-known general purpose ontology that organizes nouns, verbs, adjectives and adverbs into synsets (set of words having similar meaning), each representing one underlying lexical concept. The concepts are linked by semantic relations of various types, such as synonymy, hypernymy, hyponymy, etc.

The *core concepts* we need for building the conceptual feature vectors should allow us to evaluate the conceptual similarity between keywords  $w$  that are mapped to different concepts  $c(w)$  in the ontology. We must then rely on the hypernymy/hyponymy subgraph in WordNet linking the concepts associated to all the keywords in the database to the most generic concepts. For every concept corresponding to a keyword annotating an image, we find all the paths in the ontology that lead to the most generic concepts. The paths obtained for all the keywords in the database define the hypernyms graph. A small set (compared to the number of different keywords) of core concepts is then selected; good candidates are super-concepts of several  $c(w)$  concepts that are relatively close to these; also, the core concepts must be balanced among all the branches containing  $c(w)$  concepts.

For any image in the database, we project the keywords associated to it to the set of core concepts, through the use of several semantic similarity functions. The resulting feature vectors have the advantage of being compact, comprehensible (each dimension corresponds to a core concept) and easy to integrate in any CBIR system. Experimental evidence shows that the joint use of the visual descriptor and the new conceptual descriptor dramatically improves the quality of the results, both in a Query By Example (QBE) context and with Relevance Feedback (RF), as illustrated in the following figures. A detailed presentation of our method can be found in [3]. Our relevance feedback framework was described in [47]



Figure 12. QBE illustration in a database of 3700 images. The query image is in the top left corner, the results returned by the system are in decreasing order of visual similarity from left to right and top to bottom. When using only the visual descriptor (left image), the effects of the semantic gap are clearly visible. When using only the conceptual descriptor (middle image), the content of the returned images is in many cases too abstract and does not represent well user's intention. In the case of the joint use of visual and conceptual descriptors (right image) the returned images clearly correspond better to the intention of the user (in this case the user is looking for "outdoor statues of man riding a horse"). The image database used in our experiments was kindly provided by Andrea de Polo from Alinari Picture Library (<http://www.alinari.com>)

<sup>2</sup>This work was partially supported by funding from EC Network of Excellence MUSCLE (FP6-507752)

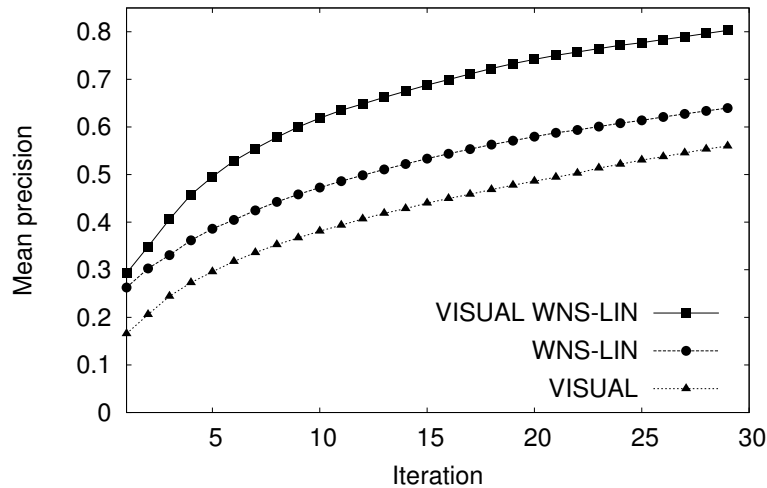


Figure 13. Mean precision vs. iterations diagram for relevance feedback on a ground truth database of 3700 images. The joint use of visual (VISUAL) and conceptual descriptors (WNS-LIN) can drastically improve the quality of the returned results compared to employing the visual or conceptual features alone.

### 6.2.3. Semantic cartography of a database from a user's query

**Keywords:** Euler diagrams, data visualization, hypergraph planarity.

**Participants:** Anne Verroust-Blondet, Marie-Luce Viaud.

We have pursued the work performed with Marie-Luce Viaud from INA (Institut National de l'Audiovisuel) within a collaboration agreement between INRIA and INA : in order to propose a graphic interface allowing the user to elaborate new strategies while searching in a database, as presented in [7], a tool automatically computing a "Euler like" diagrammatic representation of the result is under construction. This year, we have implemented a system drawing automatically a planar extended Euler diagram from any configuration involving less than nine sets. Each set corresponds to the database documents having a given term in its associated description. Our results [31] have been presented at Euler Diagrams 2005, workshop co-organized with INA (cf. examples in figure 14).

## 6.3. Clustering and learning

### 6.3.1. Active semi-supervised clustering

**Keywords:** active learning, clustering, pairwise-constraints, semi-supervised clustering.

**Participants:** Nizar Grira, Michel Crucianu, Nozha Boujemaa.

Content-based image retrieval (CBIR) can be much improved by providing a relevant organization of image collections. While it is easy to apply standard unsupervised clustering algorithms to the descriptors of the images in a database, the results of this fully automatic categorization are rarely satisfactory. The problem we are interested in may be enlightened as follows. Simple semantic information in the form of pairwise (must-link or cannot-link) constraints between data items or class labels for some items is available and the structure of the entire data has to be discovered with respect to this semantic knowledge.

We assume that users can easily evaluate whether two images should be in the same category or rather in different categories, so they can easily define the constraints mentioned above. Following previous work

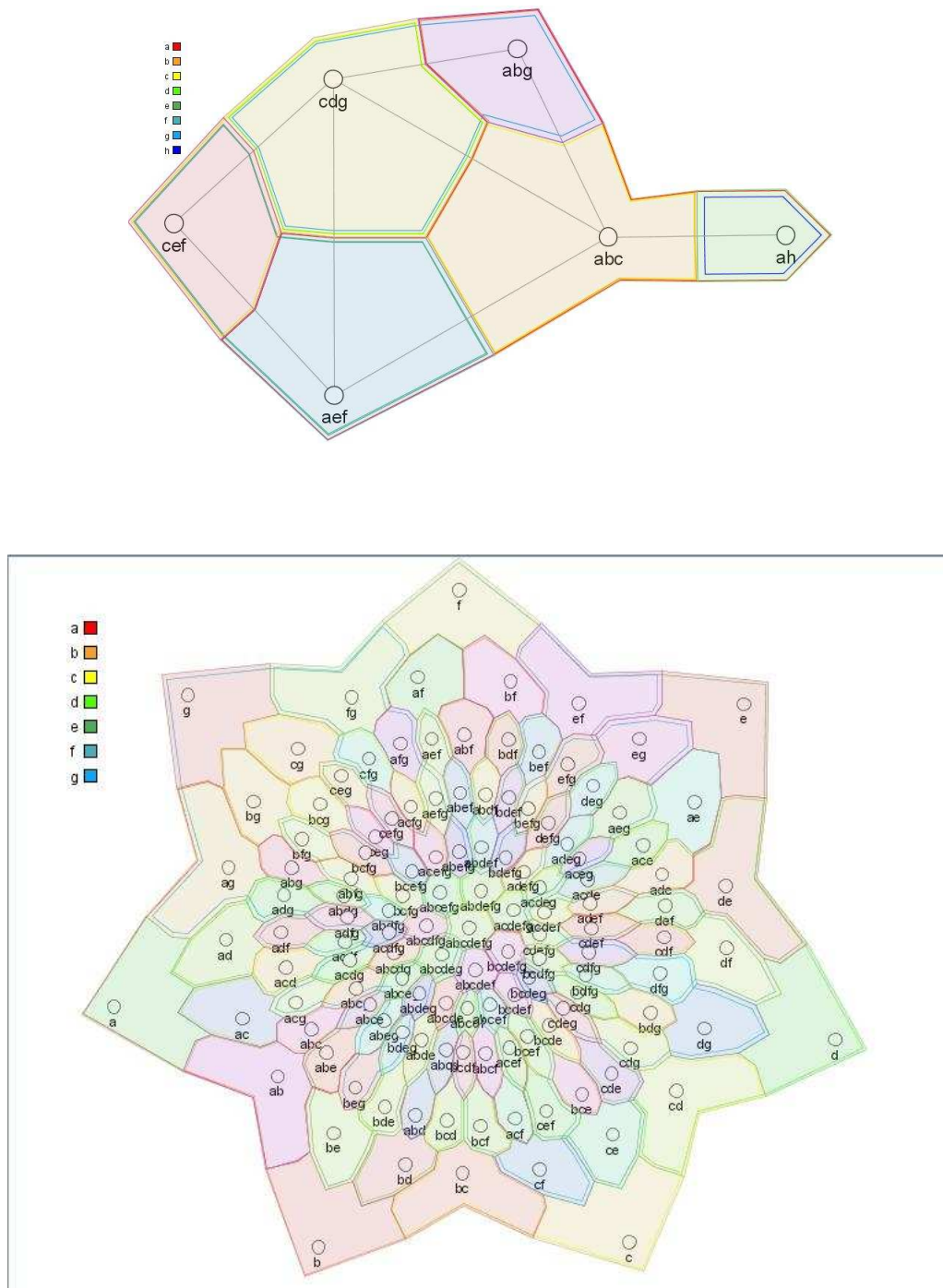


Figure 14. On the left : an Euler diagram built from the sets  $A, B, C, D, E, F, G, H$ , such that only  $A \cap H$ ,  $A \cap (B \cap C)$ ,  $A \cap (B \cap G)$ ,  $A \cap (E \cap F)$ ,  $C \cap (D \cap G)$ ,  $C \cap (E \cap F)$  are non empty. On the right : another Euler diagram built from 7 sets.



by Demiriz et al. [44], Wagstaff et al. [65] or Basu et al. [37], in [20] we introduced Pairwise-Constrained Competitive Agglomeration (PCCA), a fuzzy semi-supervised clustering algorithm. PCCA is reminded in the next paragraph. In the original version of PCCA [20] we did not make further assumptions regarding the data, so the pairs of items for which the user is required to define constraints are randomly selected. But in many cases, such assumptions regarding the data *are* available. We argue in [18] that quite general assumptions let us perform a more adequate, *active* selection of the pairs of items and thus significantly reduce the number of constraints required for achieving a desired level of performance.

In [20], we proposed an algorithm named PCCA that belongs to the family of search-based semi-supervised clustering methods. It is based on the Competitive Agglomeration (CA) algorithm [51], a fuzzy partitional algorithm that does not require the user to specify the number of clusters to be found. Using the same notations as in [20], the PCCA objective function minimize :

$$\begin{aligned} \mathcal{J}(\mathbf{V}, \mathbf{U}) &= \sum_{k=1}^C \sum_{i=1}^N (u_{ik})^2 d^2(\mathbf{x}_i, \mu_k) - \beta \sum_{k=1}^C \\ &+ \alpha \left( \sum_{(\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{M}} \sum_{k=1}^C \sum_{l=1, l \neq k}^C u_{ik} u_{jl} + \sum_{(\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{C}} \sum_{k=1}^C u_{ik} u_{jk} \right) \end{aligned} \quad (3)$$

under the constraint  $\sum_{k=1}^C u_{ik} = 1$ , for  $i \in \{1, \dots, N\}$ .

It can be shown (see [20]) that the equation for updating memberships is

$$u_{rs} = u_{rs}^{FCM} + u_{rs}^{Constraints} + u_{rs}^{Bias} \quad (4)$$

The first term in (4),  $u_{rs}^{FCM}$ , comes from FCM and only focusses on the distances between data items and prototypes. The second term,  $u_{rs}^{Constraints}$ , takes into account the supervision : memberships are reinforced or deprecated according to the pairwise constraints given by the user. The third term,  $u_{rs}^{Bias}$ , leads to a reduction of the cardinality of spurious clusters, which are discarded when their cardinality drops below a threshold.

To make this semi-supervised clustering approach attractive for the user, it is important to minimize the number of constraints he has to provide for reaching some given level of quality. This can be done by asking the user to define must-link or cannot-link constraints for the pairs of data items that are expected to have the strongest corrective effect on the clustering algorithm (i.e. that are *maximally informative*).

When using PCCA we consider that the similarities between data items provide relatively reliable information regarding the target categorization and the constraints only help finding the most relevant clusters. There is then little uncertainty in identifying well-separated compact clusters. To be maximally informative, supervision effort (i.e. constraints) should rather be spent for defining those clusters that are neither compact, nor well-separated from their neighbours. One can note that this is consistent with the findings in [55] regarding unsupervised clustering.

Following this remarks, we consider that the least well-defined cluster at iteration  $t$  is the one with the smallest density at that iteration.

When the least well-defined cluster is found, we need to identify the data items near its boundary. In the fuzzy setting, one can consider that a data item represented by the vector  $\mathbf{x}_r$  is assigned to cluster  $s$  if  $u_{rs}$  is the highest among its membership degrees. The data items at the boundary are those having the lowest membership values to this cluster among all the items assigned to it.

### 6.3.1.1. Ambiguousness criterion

As we already mentioned that in real configurations, there's very often no sharp boundary between clusters so that a fuzzy partition is often better suited for the determination of membership-based boundaries. To this end, after finding the least well-defined cluster with the upper criterion, we consider a virtual boundary that is only defined by a membership threshold and will usually be larger than the true one (this is why we call it "extended" boundary). The items whose membership values are closest to this threshold are considered to be on the boundary and the user is asked to provide constraints directly between these items. We should note that the extended boundary will probably contain ambiguous points shared with near clusters.

### 6.3.1.2. Non-redundancy criterion

Non-redundancy is complementary to ambiguousness in maximizing the amount of information provided by the constraints. The non-redundancy criterion iteratively chooses from an augmented set of feature points lying on the boundary of the least defined cluster vectors  $\mathbf{x}_j$  that maximize the lowest of the values of the distances  $d(\mathbf{x}_i, \mathbf{x}_j)$  for all the  $\mathbf{x}_i$  items already included in the selected non-redundant set. This can be expressed as :

$$\mathbf{x}_j = \operatorname{argmax}_{\mathbf{x} \in S} \min_i d(\mathbf{x}, \mathbf{x}_i) \quad (5)$$

where  $S$  is the augmented set of points and  $\mathbf{x}_i$  are the already chosen points.

In the following, we use the name AFCC (Active Fuzzy Competitive Clustering) for the resulting algorithm.

### 6.3.1.3. Experimental evaluation

We compared AFCC to original PCCA, to the CA algorithm [51] (unsupervised clustering) and to PCKmeans [37] (semi-supervised clustering).

The comparison shown here is performed on a ground truth database composed of images of different phenotypes of *Arabidopsis thaliana*, corresponding to slightly different genotypes. This scientific image database is issued from studies of gene expression. A sample of the images is shown in [18].

Figure 15 presents the dependence between the percentage of well-categorized data points and the number of pairwise constraints considered. We provide as a reference the graphs for the CA algorithm and for K-means, both ignoring the constraints (unsupervised learning). The correct number of classes was directly provided to K-means and PCKmeans. CA, PCCA and AFCC were initialized with a significantly larger number of classes and found the appropriate number by themselves.

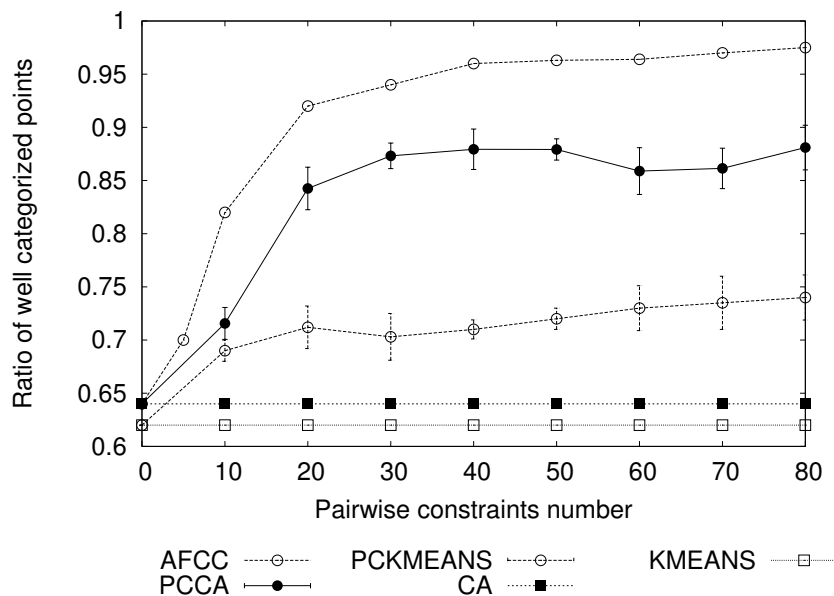


Figure 15. Results obtained by the different clustering algorithms on the Arabidopsis database

These experimental results clearly show that the user can significantly improve the quality of the categories obtained by providing a simple form of supervision, the pairwise constraints. With a similar number of constraints, PCCA performs significantly better than PCKmeans by making a better use of the available constraints. The fuzzy clustering process directly takes into account the pairwise constraints thanks to the

signed constraint terms in the equation for updating the memberships. The active selection of constraints (AFCC) further reduces the number of constraints required for reaching such an improvement. The number of constraints becomes very low with respect to the number of items in the dataset.

### 6.3.2. Kernels for similarity learning

#### 6.3.2.1. The intermediate matching kernel

**Keywords:** image classification, interest points, kernels, matching.

**Participants:** Sabri Boughorbel, Jean Philippe Tarel, Nozha Boujemaa.

Object classification with images represented by sets of local features is a challenging task for kernel based methods. We introduce a new kernel which operates on sets of vectors, named the intermediate matching kernel [12], [2]. It is based on a new and flexible matching approach which improves results over GCS kernel [10]. The matching procedure is guided by an intermediate set of vectors. Indeed, an explicit mapping  $\Phi_{\mathbf{m}}$  depending on a vector  $\mathbf{m} \in \mathbb{R}^d$  defines the core of the proposed kernel. One possible choice, but not the only one, for  $\Phi_{\mathbf{m}}$  is to associate, for each element of the input space  $\mathcal{X}$  (a set of vectors from  $\mathbb{R}^d$ ), the nearest vector from  $\mathcal{X}$  to  $\mathbf{m}$  :

$$\begin{aligned} \Phi_{\mathbf{m}} : \mathcal{P}(\mathbb{R}^d) &\rightarrow \mathbb{R}^d \\ \mathcal{X} &\mapsto \mathbf{x}_{\mathbf{m}}^* = \arg \min_{\mathbf{x} \in \mathcal{X}} \|\mathbf{x} - \mathbf{m}\| \end{aligned}$$

This mapping is performed for  $\mathcal{X}$  and  $\mathcal{X}'$  which are the two sets of vectors to be compared, then the matched pair of vectors is compared using any positive definite kernel  $k$  operating on vectors such as the RBF kernel. A sequence of parametric mappings  $\{\Phi_{\mathbf{m}_1}, \dots, \Phi_{\mathbf{m}_p}\}$  are separately applied for the two sets of vectors. The intermediate matching kernel is finally obtained as a sum over the matched pairs. More formally, for any set of vectors  $\mathcal{M} = \{\mathbf{m}_1, \dots, \mathbf{m}_p\}$ ,  $\mathbf{m}_i \in \mathbb{R}^d$ , the intermediate matching kernel can be defined as :

$$K_{\mathcal{M}}(\mathcal{X}, \mathcal{X}') = \sum_{i=1}^p k(\mathbf{x}_{\mathbf{m}_i}^*, \mathbf{x}'_{\mathbf{m}_i}) = \sum_{i=1}^p k(\Phi_{\mathbf{m}_i}(\mathcal{X}), \Phi_{\mathbf{m}_i}(\mathcal{X}')) \quad (6)$$

By construction,  $K_{\mathcal{M}}$  is positive definite. An important fact to notice is that the positive definiteness of the intermediate matching kernel is insured when the matching set  $\mathcal{M}$  is chosen independently from  $\mathcal{X}$  and  $\mathcal{X}'$ . We define  $\mathcal{M}$  as the set formed with the center classes obtained by clustering all vectors of the training sample. However, other approaches can be used.

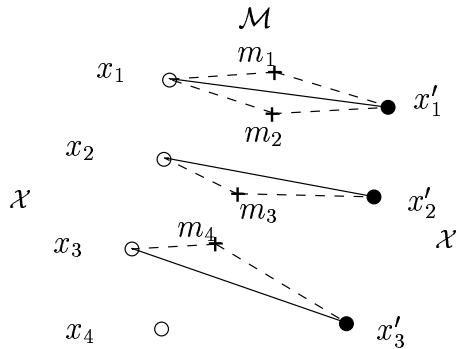


Figure 16.

An example of the intermediate matching kernel construction. The two points  $\mathbf{m}_1$  and  $\mathbf{m}_2$  select the same pair  $(\mathbf{x}_1, \mathbf{x}'_1)$ . The point  $\mathbf{x}_4$  is far from  $\mathcal{M}$ , then it is discarded from the intermediate matching kernel expression.

Table 1. Validation and test errors comparisons for the different kernels on a database of 1000 images from ETH80.

	Local jets	
Kernels	valid. error	test error
Matching kernel (loser-take-nothing)	19.66±0.20	19.66±0.40
Matching kernel (winner-take-all)	13.46±0.87	13.1±0.69
GCS-based kernel	8.93±0.53	9.33±1.00
Intermediate matching kernel	8.33±0.54	<b>8.93±0.16</b>

Table 1 summarizes performance comparisons with other kernels that operate on sets of vectors. The task is image classification with images represented by set of jets around interest point. The Intermediate Matching kernel yielded the best test errors with 8.33%. In this experiment, the size of the matching set is 40.

#### 6.3.2.2. The LCCP for the selection of kernel parameters

**Keywords:** *gradient descent, kernel parameters, model selection, svm.*

**Participants:** Sabri Boughorbel, Jean Philippe Tarel, Nozha Boujemaa.

Tuning hyper-parameters is a necessary step to improve learning algorithm performances. For SVM, adjusting kernel parameters may increases results drastically. Parameter tuning is usually performed using cross-validation by sweeping the parameter space. When the number of parameters increases, such as in the case of multiple kernel parameter, the complexity of such grid search is exponential with respect to the number of optimized parameters. The gradient descent approach introduced in [43] reduces significantly search steps of optimal parameters. We define the LCCP (Log Convex Concave Procedure) derived from the CCCP optimization framework (Convex ConCave Procedure) [67] for optimizing kernel parameters, by minimizing the radius-margin bound. To apply the LCCP, we prove, for a good choice of kernels, that the radius is log convex and the margin is log concave [13]. The LCCP is more efficient than gradient descent technique since it insures that the minimized criterion decreases monotonically and converges to a local minimum without searching the size step. We apply the LCCP for the optimization of parameters  $\theta_i$  in the following kernel :

$$K_{L_1, \theta}(\mathbf{x}, \mathbf{x}') = - \sum_{i=1}^n \frac{|x_i - x'_i|}{\theta_i}, \quad \theta = (\theta_i), 1 \leq i \leq n \quad (7)$$

which is an extension of the  $L_1$ -distance kernel :

$$K_{L_1}(\mathbf{x}, \mathbf{x}') = - \sum_{i=1}^n |x_i - x'_i|. \quad (8)$$

We also recall the  $L_2$ -distance kernel :

$$K_{L_2}(\mathbf{x}, \mathbf{x}') = - \sum_{i=1}^n (x_i - x'_i)^2, \quad (9)$$

The above two kernels can be shown to be conditionally positive definite [9], [11].

Table 2. Test error's comparison of the single parameter L 1 -distance kernel ( $K_{L_1}$ ), L 2 -distance kernel ( $K_{L_2}$ ) and L 1 -distance kernel with multiple parameters ( $K_{L_1, \theta}$ ).  $n$  denotes the number of parameters for multi-parameter's kernel. LCCP is used for optimizing the radius margin bound.

	Thyroid	Titanic	Heart	Breast-cancer
$K_{L_1}$ (8)	<b>5.77%</b>	22.68%	20.65%	28.97%
$K_{L_2}$ (9)	11.21%	22.56%	18.23%	29.77%
$K_{L_1, \theta}$ (7)	6.20%	<b>22.08%</b>	<b>17.34%</b>	<b>27.12%</b>
$n$	5	3	13	9

Tab. 2 summarizes the average test errors for different data sets and shows clearly the interest of the multiple kernel parameters, despite a preliminary weighting of the databases.

### 6.3.2.3. Kernel PCA for similarity invariant shape recognition and its application to leaf species

**Keywords:** invariant shape, kernels.

**Participant:** Hichem Sahbi.

There is a more and more interest in shape recognition with the growing need for retrieval applications such as finding leaf species. In this work, we introduced a new shape description method targeted for 2D shape contours. A contour is a manifold  $\mathcal{L} = \{(x(s), y(s)) \in \mathbb{R}^2, s \in [0, 1]\}$ , where  $s$  is referred to as the curvilinear abscissa and  $\mathcal{S}$  as a uniform sampling of  $\mathcal{L}$ , according to  $s$  (i.e.,  $s = k\Delta, k \in \mathbb{N}^+, \Delta = 1/|\mathcal{L}|$ ). It is shown in our work [49], [28] that the triangular kernel, defined as  $k(x, x') = -\|x - x'\|^p, p \in ]0, 2[$  achieves similarity and particularly scale invariance for many kernel methods including support vector machine and kernel principal component analysis (KPCA). Given a shape contour  $\mathcal{S}$ , the top  $d$  eigenvalues, of KPCA on  $\mathcal{S}$ , are used as its shape description. These eigenvalues are rotation and translation invariant when using the Gaussian kernel and can be normalized to be also scale invariant when using the triangular kernel. Notice that computing this description does not necessitate sampling a curve according to an *ordered* curvilinear abscissa which might be unavailable and difficult to find mainly for complex contours.

This shape description is also scale invariant when using the linear kernel, nevertheless the dimension of the underlying eigenspace will not exceed two<sup>3</sup>, so the eigenvalues of this eigenspace will not be sufficient to discriminate different and complex shapes especially those containing fine details as shown in the experiments (cf. Fig 18, right). While this problem might be overcome when using the Gaussian kernel, KPCA is not similarity invariant, when using this kernel and it necessitates finding different scales for different curves, so it is meaningless to compare the underlying eigenvalues. The triangular kernel gathers two properties : achieving scale (and similarity) invariance while ensuring a "rich" description as the number of eigenvalues increases when the cardinality of a contour  $\mathcal{S}$  increases.

We ran our experiments on Smithsonian and Swedish [64] databases consisting respectively of 1525 and 1125 images of leaves. The Smithsonian dataset has 135 categories with different cardinalities while the Swedish set has 15 classes, each one contains 75 images. Notice that only the external contours are used for shape description even though this is not sufficient to make the "best" predictions as color and internal structures are also prominent. Figure (17) shows the precision-recall curves for different values of the regularizer parameter  $p$  of the triangular kernel. The precision reaches the highest value for  $p = 1.9$  and drops dramatically when  $p$  is close to its upper bound (i.e.,  $p = 2$ ). Figure (18) shows the precision-recall on the Swedish dataset for  $d = 100$  dimensions using our shape description. Comparisons, with the representative state of the art work, are reported using linear PCA ( $d = 2$ ) and other shape descriptors including the edge orientation histogram (EOH) [50], Hough [3] and Curvature Scale Space (CSS) [63]. When the recall is less than 30%, the precision of KPCA is better than the other descriptors including CSS.

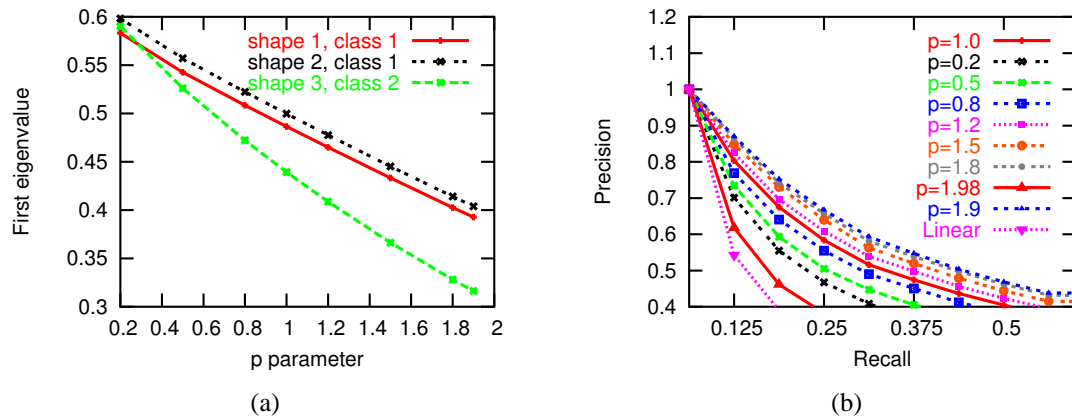


Figure 17. (a) Variation of the first eigenvalue of three shapes. We can clearly see that the first eigenvalues of shape 1 and 2 (from the same class) are close while the first eigenvalue of shape 3 diverges as  $p$  increases. (b) Variation of precision-recall on the Smithsonian dataset with respect to the kernel parameter  $p$ . ( $d = 100$ ).

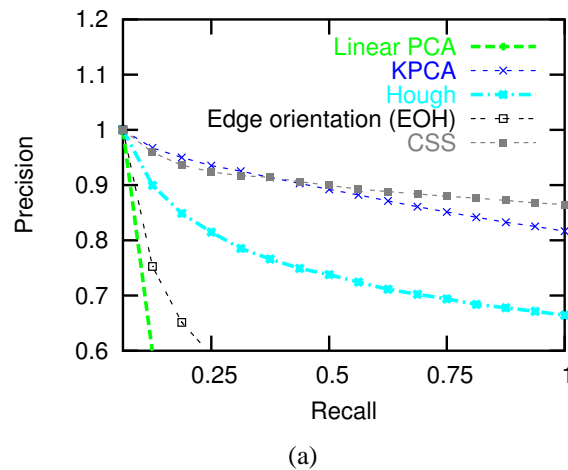


Figure 18. Precision versus recall on the Swedish dataset using the eigenvalue description (with linear and triangular kernels,  $p = 1$ ), Hough, EOH and CSS [63]. When using the triangular kernel, 100 eigenvalues are considered for our shape description while (at most) 2 eigenvalues are considered when using the linear kernel.

## 6.4. Software

### 6.4.1. IKONA/MAESTRO software

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

**Participants:** Nicolas Hervé, Mathieu Coutaud, Jean-Paul Chièze.

This year, the Ikona/Maestro software has been fully re-engineered to improve its modularity. Most of the changes have been done on the server part. They relate to the internal architecture. In this new version, the interfaces of all the main components have been unified. This changes will make easier the integration of new descriptors, new query paradigms and new fonctionnalités.

### 6.4.2. aceMedia software integration

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

**Participant:** Nicolas Hervé.

In the aceMedia european integrated project, IMEDIA is in charge of the developement 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. The first version has been delivered in June. It has been integrated in a global client-side application : the PCS User Interface. Further improvements on this module, as well as its integration in a web-based application are planned for the next year.

## 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 resources 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 developement 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 Thales.

### 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

<sup>3</sup>As the training examples live in  $2D$ , we have at most two non null eigenvalues when solving KPCA using the linear kernel.

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.

## 7.3. 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](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 developping 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*

- Keynote speaker for the international conference ESA-EUSC'05: Image Information Mining - Theory and Application to Earth Observation ;
- Scientific expert for the European commission, invited to the consultation workshop "Challenges of Future Search Engines" held at Brussels during September 2005 for the preparation of FP7 calls ;
- Scientific expert for NWO Research agency (Netherlands - CATCH projects) ;
- Organizer of a special session "Machine Learning for Visual Information Retrieval" for the 7th ACM SIGMM International Workshop on Multimedia Information Retrieval (ACM-MIR'05) ;
- "Tutorials Chair" for ECDL'05 besides member of several TPC among them ICME'05, MIR'05...
- Deputy scientific coordinator of Muscle NoE (Network of Excellence FP6), member of steering committee and scientific coordinator of WP5 "Single Modality". Co-organizer (with Valerie Gouet) of several scientific meeting for WP5. ;
- In charge of international relation at INRIA Rocquencourt and member of "Bureau du Comité des Projets" till october 2005 ;
- Since September 2005, member of the "National Evaluation Commission" of INRIA.
- PhD Jury member :
  - Chairman : INSA Lyon.
  - Reviewer : Ecole Centrale Lyon, Université de Nice, Université De Reims.
- HDR Jury member :
  - Reviewer : Antoine Tabbone - Université de Nancy.



### 8.1.2. Valérie Gouet-Brunet

- Member of the scientific commission (section CNU 27) of CNAM ;
- Member of the CNAM/CEDRIC laboratory council ;
- Scientific expert for the french research program ARA "Masses of Data" ;
- Deputy scientific leader of the work package WP5 in the Muscle Network of Excellence ;
- Leader of the WP5/task 2 in the Muscle Network of Excellence ;
- Co-organiser (with Nozha Boujemaa) of the WP5 Third Scientific meeting in the Muscle Network of Excellence, 27-29 April 2005, Paris ;
- Co-organiser (with Nozha Boujemaa) of the WP5 First focus meeting in the Muscle Network of Excellence, 1-2 December 2005, Rocquencourt ;
- Conference program committee member of the International Workshop on Computer Vision meets Databases (CVDB'05), in conjunction with ACM Sigmod, June 17, 2005, Baltimore, MD, USA ;
- Jury member of several engineer diplomas at CNAM ;
- Reviewer for conference papers : ICME'05, CVDB'05, Acivs'05 ;
- Reviewer for a journal paper : IVC.

### 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. Jean-Philippe Tarel

- Journal Reviewer : Journal of Mathematical Imaging and Vision ;
- Jury member :
  - PhD Sabri Boughorbel, Paris Orsay University, 13 juillet, 2005.
- Member of program commity of 11th International Conference on Computer Analysis of Images and Patterns (CAIP'2005). Member of reading commity of Colloque du Groupement de Recherche En Traitement du Signal et de l'Image (GRETSI'2005).

### 8.1.5. Anne Verroust-Blondet

- AFIG President (Association Française d'informatique Graphique) ;
- Member of the Executive Committee (Conseil d'Administration) of the French chapter of Eurographics ;
- Co-organizer of Euler Diagrams 2005 workshop (<http://www-rocq.inria.fr/imedia/euler2005/>) with Marie-Luce Viaud of INA.

## 8.2. Teaching

### 8.2.1. Sabri Boughorbel

- 30h TPs on Java, 1st year, IUT Velizy, May 2005.

### 8.2.2. 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 Paris 6 (7,5 hours).

### 8.2.3. Jean-Philippe Tarel

- 6h courses on "Object Detection", Master STIC, CNAM, november 2005.
- 24h TP on "Computer Vision", Master Sciences de l'ingénieur, Pierre and Marie Curie University, october 2005.
- 3h Course on "Object Detection", Master Sciences de l'ingénieur, Pierre and Marie Curie University, november 2005.

### 8.2.4. Anne Verroust-Blondet

- Course (9H) on Computational Geometry in the option "Computer Vision" in last year of the engineering degree course of the ENSTA school (École Nationale Supérieure de Techniques Avancées, Paris).

## 9. Bibliography

### Major publications by the team in recent years

- [1] 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.

### Doctoral dissertations and Habilitation theses

- [2] S. BOUGHORBEL. *Kernels for Image Classification with Support Vector Machines*, Ph. D. Thesis, Faculté d'Orsay, 2005.
- [3] M. FERECATU. *Image retrieval with active relevance feedback using both visual and keyword-based descriptors*, Ph. D. Thesis, Université de Versailles Saint-Quentin-en-Yvelines, July 2005.

### Articles in refereed journals and book chapters

- [4] V. GOUET-BRUNET. *Encyclopédie de l'Informatique*, chap. Recherche par contenu visuel dans les grandes collections d'images, Vuibert, 2005.
- [5] N. GRIRA, M. CRUCIANU, N. BOUJEMAA. *Fuzzy Clustering With Pairwise Constraints for Knowledge-Driven Image Categorization*, in "IEEE Proceedings - Vision, Image and Signal Processing", to appear, 2006.
- [6] H. SAHBI, D. GEMAN. *A Hierarchy of Support Vector Machines for Pattern Detection*, in "Machine Learning Research", 2005.

- [7] J. THIÈVRE, M.-L. VIAUD, A. VERRONST-BLONDET. *Using Euler Diagrams in Traditional Library Environment*, in "Electronic Notes in Computer Science", vol. 134, June 2005, p. 189–202.
- [8] A. VERRONST-BLONDET, M.-L. VIAUD. *Encyclopédie de l'Informatique*, chap. Introduction à la synthèse d'images, Collectif Vuibert, to appear, 2005.

## Publications in Conferences and Workshops

- [9] S. BOUGHORBEL, J. TAREL, N. BOUJEMAA. *Conditionally Positive Definite Kernels For SVM Based Image Recognition*, in "International Conference on Multimedia and Expo (ICME'05), The Netherlands", 2005.
- [10] S. BOUGHORBEL, J. TAREL, N. BOUJEMAA. *The GCS Kernel For SVM Based Image Recognition*, in "The International Conference on Artificial Neural Networks (ICANN'05), Poland", 2005.
- [11] S. BOUGHORBEL, J. TAREL, N. BOUJEMAA. *The Generalized Histogram Intersection For Image Recognition*, in "The International Conference on Image Processing (ICIP'05), Italy", 2005.
- [12] S. BOUGHORBEL, J. TAREL, N. BOUJEMAA. *The Intermediate Matching Kernel For Local Image Features*, in "The Proceedings of International Joint Conference on Neural Networks (IJCNN'04), Canada, Montreal", 2005.
- [13] S. BOUGHORBEL, J. TAREL, N. BOUJEMAA. *The LCCP For Optimizing Kernel Parameters*, in "The International Conference on Artificial Neural Networks (ICANN'05), Poland", 2005.
- [14] N. BOUTELDJA, V. GOUET-BRUNET, M. SCHOLL. *Evaluation of strategies for multiple sphere queries with local image descriptors*, in "Accepted to the IS and T/SPIE Conference on Multimedia Content Analysis, Management, and Retrieval, San Jose CA, USA", to appear, January 2006.
- [15] 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.
- [16] Y. FANG, D. GEMAN. *Experiments in Mental Face Retrieval*, in "Audio- and Video-based Biometric Person Authentication (AVBA'2005), New York", July 20–22 2005.
- [17] M. FERECATU, N. BOUJEMAA, M. CRUCIANU. *Hybrid Visual and Conceptual Image Representation within Active Relevance Feedback Context*, in "Proceedings of the 7th ACM SIGMM International Workshop on Multimedia Information Retrieval", to appear, November 2005.
- [18] N. GRIRA, M. CRUCIANU, N. BOUJEMAA. *Active semi-supervised clustering for image database categorization*, in "Content-Based Multimedia Indexing (CBMI'05)", June 2005.
- [19] 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)", to appear, November 2005.

- 
- [20] N. GRIRA, M. CRUCIANU, N. BOUJEMAA. *Semi-supervised fuzzy clustering with pairwise-constrained competitive agglomeration*, in "IEEE International Conference on Fuzzy Systems (Fuzz'IEEE 2005)", May 2005.
- [21] H. HOUISSA, N. BOUJEMAA. *Coherence criterion for region labelling and description*, in "Proceedings of the International Conference on Machine Intelligence (ICMI'05), Tozeur, Tunisia", to appear, 2005.
- [22] H. HOUISSA, N. BOUJEMAA. *Region labelling using a Point-based Coherence Criterion*, in "IS and T/SPIE Conference on Multimedia Content Analysis, Management, and Retrieval 2006, San-Jose, CA, USA", to appear, 2006.
- [23] A. JOLY, O. BUISSON. *Discriminant Local Features Selection using Efficient Density Estimation in a Large Database*, in "Int. Workshop on Multimedia Information Retrieval", 2005.
- [24] A. JOLY, C. FRÉLICOT, O. BUISSON. *Content-based Video Copy Detection in Large Databases: A Local Fingerprints Statistical Similarity Search Approach*, in "Int. Conf. on Image Processing", 2005.
- [25] H. LIU, H. SAHBI, L. CROCE-FERRI, M. STEINEBACH. *Advanced Semantic Authentication of Face Images*, in "WIAMIS05, 6th International Workshop on Image Analysis for Multimedia Interactive Services, Montreux, Switzerland", April 2005.
- [26] H. SAHBI, N. BOUJEMAA. *Validity of Fuzzy Clustering using Entropy Regularization*, in "IEEE International Conference on Fuzzy Systems (Fuzz'IEEE 2005), Reno, USA", May 2005.
- [27] H. SAHBI. *Affine Invariant Shape Description Using the Triangular Kernel and its Application to Leaf Recognition*, in "Proceedings of the 4th International Workshop on Content-Based Multimedia Indexing, Riga, Latvia", June 2005.
- [28] H. SAHBI. *Invariant Shape Description and Its Application for Leaf Recognition*, in "Proceedings of the IEEE Workshop on Content Based Image Retrieval", 2005.
- [29] S. THIEMERT, H. SAHBI, M. STEINEBACH. *Applying Interest Operators in Semi-Fragile Video Watermarking*, in "IS and T/SPIE Conference on Security, Steganography and Watermarking of Multimedia Contents (SPIE), San Jose, California", January 2005.
- [30] S. THIEMERT, H. SAHBI, M. STEINEBACH. *Using Entropy for Image and Video Authentication Watermarks*, in "IS and T/SPIE Conference on Security, Steganography and Watermarking of Multimedia Contents (SPIE), San Jose, California", to appear, January 2006.
- [31] J. THIÈVRE, A. VERRONST-BLONDET, M.-L. VIAUD. *Drawing diagrams from labelled graphs*, in "Euler Diagrams'2005, Paris", September 2005.
- [32] A. VERRONST-BLONDET. *Indexation 3D : un tour d'horizon*, in "Journées AFIG 2005, Strasbourg", 2005.
- [33] I. YAHIAOUI, N. BOUJEMAA, N. HERVÉ. *Visual similarity search in botanical collections*, in "International Conference on Machine Intelligence (ICMI'05), Tozeur, Tunisia", to appear, November 2005.

- [34] I. YAHIAOUI, N. BOUJEMAA. *Content-based image retrieval in botanical collections for gene expression studies*, in "IEEE International Conference on Image Processing (ICIP'05)", September 2005.

## Internal Reports

- [35] O. BESBES, Z. BELHADJ, N. BOUJEMAA. *Adaptive Satellite Images Segmentation by Level Set Multiregion Competition.*, Research Report, INRIA, 2006, <http://www.inria.fr/rrrt/rr-5855.html>.

## Miscellaneous

- [36] M. CHAOUCH. *Indexation 3D : Descripteurs 2D/3D basés sur les images de profondeur*, September 2005, Rapport de Master, Master Recherche IAD, Université Paris VI.

## Bibliography in notes

- [37] S. BASU, A. BANERJEE, R. J. MOONEY. *Semi-Supervised Clustering by Seeding*, in "Proceedings of 19th International Conference on Machine Learning (ICML-2002)", 2002, p. 19–26.
- [38] C. BOHM, B. BRAUNMULLER, H. KRIEGEL. *The Pruning Power: Theory and Heuristics for Mining Databases with Multiple kNearest-Neighbor Queries*, in "Proceedings of the International Conference on Data Warehousing and Knowledge Discovery (DaWaK 2000), Greenwich, United Kingdom", 2000.
- [39] N. BOUJEMAA. *"Sur la classification non-exclusive en analyse d'images"*, Habilitation à diriger des recherches, Université de Versailles-Saint-Quentin, 2000.
- [40] 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.
- [41] B. BRAUNMULLER, M. ESTER, H. KRIEGEL, J. SANDER. *Efficiently supporting multiple similarity queries for mining in metric databases*, in "16th Int. Conf. on Data Engineering, San Diego, CA", 2000, p. 256-270.
- [42] T. BROX, M. ROUSSON, R. DERICHE, J. WEICKERT. *Unsupervised Segmentation Incorporating Colour, Texture, and Motion.*, Research Report, n° 4760, INRIA, March 2003.
- [43] O. CHAPELLE, V. VAPNIK, O. BOUSQUET, S. MUKHERJEE. *Choosing Multiple Parameters for Support Vector Machines*, in "Machine Learning", vol. 46, n° 1-3, 2002, p. 131–159.
- [44] A. DEMIRIZ, K. BENNETT, M. EMBRECHTS. *Semi-supervised clustering using genetic algorithms*, in "Intelligent Engineering Systems Through Artificial Neural Networks 9", C. H. DAGLI, ET AL. (editors). , ASME Press, 1999, p. 809–814.
- [45] G. DORKÓ, C. SCHMID. *Selection of Scale-Invariant Parts for Object Class Recognition*, in "International Conference on Computer Vision", vol. 1, 2003, p. 634–640.
- [46] J. FAUQUEUR, N. BOUJEMAA. *New Image Retrieval Paradigm: Logical Composition of Region Categories*, in "In IEEE International Conference on Image Processing", 2003.

- 
- [47] M. FERECATU, M. CRUCIANU, N. BOUJEMAA. *Retrieval of difficult image classes using SVM-based relevance feedback*, in "Proceedings of the 6th ACM SIGMM International Workshop on Multimedia Information Retrieval", October 2004, p. 23 – 30.
- [48] F. FLEURET. *Détection hiérarchique de visages par apprentissage statistique*, Ph. D. Thesis, Université Paris-VI, Paris, 2000.
- [49] F. FLEURET, H. SAHBI. *Scale-Invariance of Support Vector Machines Based on the Triangular Kernel*, in "Third International Workshop on Statistical and Computational Theories of Vision (part of ICCV)", 2003.
- [50] W. FREEMAN, M. ROTH. *Orientation histograms for hand gesture recognition*, in "International Workshop on Automatic Face and Gesture Recognition", 1995, p. 296–301.
- [51] H. FRIGUI, R. KRISHNAPURAM. *Clustering by competitive agglomeration*, in "Pattern Recognition", vol. 30, n° 7, 1997, p. 1109–1119.
- [52] V. GOUET-BRUNET, N. BOUJEMAA. *Object-based queries using color points of interest*, in "In Workshop on Content-Based Access of Image and Video Libraries (CVPR)", 2001.
- [53] H. GUNTHER. *Focus-of-Attention from Local Color Symmetries.*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", vol. 26, n° 7, 2004, p. 817-830.
- [54] C. HARRIS, M. STEPHENS. *A combined corner and edge detector*, in "A combined corner and edge detector", 1988, p. 147-151.
- [55] T. HOFMANN, J. M. BUHMANN. *Active data clustering*, in "Advances in Neural Information Processing Systems (NIPS) 10", 1997, p. 528–534.
- [56] J. IIVARINEN, A. VISA. *Shape recognition of irregular objects*, in "Intelligent Robots and Computer Vision XV: Algorithms, Techniques, Active Vision, and Materials Handling, SPIE", 1996.
- [57] A. K. JAIN, A. VAILAYA. *Image retrieval using color and shape*, in "Pattern Recognition", vol. 29, n° 8, August 1996, p. 1233-1244.
- [58] I. LAPTEV, T. LINDBERG. *Space-Time Interest Points*, in "Proceedings of the International Conference on Computer Vision", 2003, p. 432-439.
- [59] D. G. LOWE. *Distinctive Image Features from Scale-Invariant Keypoints*, in "ijcv", vol. 60, n° 2, 2004, p. 91–110.
- [60] B. MANJUNATH, P. SALEMBIER, T. SIKORA. *Introduction to MPEG-7*, Wiley, 2002.
- [61] A.-R. MANSOURI, A. MITICHE, C. VAZQUEZ. *Image Partitioning by Level Set Multiregion Competition*, in "ICIP", October 2004, p. 2721–2724.

- 
- [62] K. MIKOLAJCZYK, C. SCHMID. *A performance evaluation of local descriptors*, in "IEEE Trans. on Pattern Analysis and Machine Intelligence", vol. 27, n° 10, 2005, p. 1615–1630.
- [63] F. MOKHTARIAN, S. ABBASI, J. KITTLER. *Robust and Efficient Shape Indexing through Curvature Scale Space*, in "Proceedings of British Machine Vision Conference", 1996, p. 53–62.
- [64] O. SODERKVIST. *Computer Vision Classification of Leaves from Swedish Trees*, 2001, Rapport de Master, Linköping University.
- [65] K. WAGSTAFF, C. CARDIE. *Clustering with instance-level constraints*, in "Proceedings of the 17th International Conference on Machine Learning", 2000, p. 1103–1110.
- [66] R. WEBER, H.-J. SCHEK, S. BLOTT. *A Quantitative Analysis and Performance Study for Similarity-Search Methods in High-Dimensional Spaces.*, in "VLDB", 1998, p. 194-205.
- [67] A. L. YUILLE, A. RANGARAJAN. *The Concave-Convex Procedure*, in "Neural Computation", vol. 15, n° 4, 2003, p. 915–936.