INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Project-Team Imedia

Images and Multimedia : Indexing, Retrieval and Navigation

Rocquencourt

THEME 3B

Table of contents

1. Team

Head of project-team
Nozha Boujemaa [Research Director (DR) INRIA]

Administrative assistant
Laurence Bourcier [shared with Eiffel project-team]

INRIA staff
Anne Verroust [Research Scientist (CR1)]
François Fleuret [Research Scientist (CR2) INRIA]
Michel Crucianu [Research Scientist (CR1), civil servant since 1/09/2002]
Jean-Philippe Tarel [Research Scientist (CR2), civil servant since 1/11/2001]
Jean-Paul Chièze [Senior Technical Staff, half-time]

Scientific advisor
Donald Geman [Professor at Johns Hopkins University and ENS Cachan]

Research scientist (partner)
Valérie Gouet [Assistant Professor at CNAM since 01/09/2002]

Invited Professor
Vincent Oria [Research Scientist, visiting since 01/06/2003]

Post-Doctoral fellow
Yuchun Fuang [Post-Doctor since 01/05/2003]

PhD Student
Hichem Houissa [INRIA Rocq grant since 1/10/2003]
Nizar Grira [INRIA Rocq grant since 1/12/2002]
Sabri Boughorbel [INRIA Rocq grant since 1/11/2001]
Marin Ferecatu [INRIA Rocq grant since 1/10/2001]
Julien Fauqueur [INRIA Rocq since 1/03/2000 [1]]
Bertrand Le Saux [INRIA Rocq since 1/11/1999 [2]]
Hichem Sahbi [French-Algerian cooperation grant since 1/10/1999 [3]]

Graduate Student intern
Hichem Houissa [DEA ATS-Paris 11 since April 2003]
Cedric Timsit [DEA IARFA-Paris 6 since April 2003]

Student intern
Akram Hentati [Internship Sup'Com-Tunis since March 2003]

2. 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

[1] PhD defended during 2003
[2] PhD defended during 2003
[3] PhD defended during 2003

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

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

Participants: Nozha Boujemaa, Valérie Gouet, Julien Fauqueur, Hichem Houissa, Sabri Bougorbel, Jean-Philippe Tarel, Nizar Grira, Anne Verroust, Hichem Sahbi, Jean-Paul Chièze.

Glossary

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, 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 [19] [23].

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

Key words: *Statistical learning*, *algorithmic optimization*, *kernel methods*.

Participants: François Fleuret, Donald Geman, Hichem Sahbi, Sabri Boughorbel, Michel Crucianu, 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.

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

3.3.2. Clustering methods

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

Participants: Nozha Boujemaa, Bertrand Le Saux, Nizar Grira, Michel Crucianu.

Unsupervized 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

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

Participants: Marin Ferecatu, Yuchun Fang, Julien Fauqueur, Donald Geman, Nozha Boujemaa, Michel Crucianu, Hichem Houissa.

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.

3.5. Cross-media indexing

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

Participants: Marin Ferecatu, Francois Fleuret, Valerie Gouet, 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 mathods 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

- **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 developped by IMEDIA can be used with this norm.

5. Software

5.1. IKONA/MAESTRO Software

The architecture of this client/server software and several visual signatures were a subject of a deposit to APP.
Key words: *User interface*, *image retrieval by content*, *relevence feed-back*.

Correspondents: Marin Ferecatu, Paul-Paul Chièze
The user interface or "client" is the software used to send the query, to display the pages of results and to handle complex queries (with feedback, keywords, etc...), so it should be intuitive, fast and easy to use.

IKONA is a new architecture for building Content Based Image Retrieval software prototypes, designed and implemented in our team during the last three years [20]. It consists of two independent parts: the server and the client. Each of the two parts communicates with the other through a network protocol which is a set of commands the server understands and a set of answers it returns to the client. The communication protocol is modular and extensible, i.e. it is easy to add new functionality without disturbing the overall architecture.

Global signatures for images databases implemented in the indexer include :

- Generic signatures: Color, Shape and Texture features investigated at the Imedia Group.
- Specific signatures: Faces and signatures for fingerprints.

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 interface of the client:

- **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 [21];
- **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.

A good starting point for exploring the possibilies offered by IKONA is our Web demo, available at http://www-rocq.inria.fr/cgi-bin/imedia/ikona. This 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. More screenshots describing the visual searching capabilities of IKONA are available at http://www-rocq.inria.fr/imedia/ikona.html.

6. New Results

6.1. Construction and organization of the visual feature space

6.1.1. Coarse-to-Fine Face Component Extraction

Key words: *Face analysis*, *coarse-to-fine processing*, *face component extraction*, *support vector machines*.

Participant: Hichem Sahbi.

Face component extraction is a preliminary step for many applications in face processing such as detection and recognition. Several works tackled this issue and the most representative study is [30] that is based on elastic bunch graph matching. The method presented in this abstract uses the same idea of graphs in order to model the geometry of faces and differs in the clustering strategy which is more efficient from the algorithmic point-of-view.

A face is represented by a graph where each node corresponds to the location of a particular face component and it is associated with a support vector machine that responds positively if and only if a given subimage corresponds to its underlying face component (cf. Figure 1, left). Two graph structures are used in order to achieve accurate and efficient extraction of the facial components. First, a coarse graph is applied only into a small sub-lattice of a scene, and makes it possible to carry out invariance to translation and non-linear face deformations due to changes in expression. Then, a fine graph is applied locally, in order to find the exact location of the face components (cf. Figure 1, right). This method, effective and computationally efficient, can be used for face recognition.

6.1.2. A study on fine similarity measures for local descriptors

Key words: *Points of interest*, *local descriptors*, *similarity measures*.

Participants: Valérie Gouet, Jean-Philippe Tarel.

We focus on the local characterization of images by points of interest. The use of points of interest allows us to make queries on parts of images, as well as on objects contained in an image.

In the previous approaches, different point extraction algorithms and different local descriptors of these extracted points are proposed. To compare these different local descriptors, the similarity measure must be carefully chosen, for each local descriptor, to achieve the best performance. It is only when the similarity measure is optimized for each descriptor that a comparison between different kind of descriptors can be performed. Descriptors are subject to different kind of noise and the optimal similarity measure is directly related to the shape of this noise. In particular, descriptor components do not have the same scale of variations. Therefore, the similarity measure must be weighted differently depending of the components. The problem is then to correctly estimate these weights.

Our study is focused on five kinds of local descriptors: jets, differential invariants, edge orientation histograms, Hough transform and algebraic representation of the local contours. They are all computed on color images.

The simplest way to estimate the weights is to use the covariance matrix of the database (i.e. Mahalanobis distance). It turns out that this approach is not optimal, because when the order of the jet is higher than 2, performances are decreasing.

We thus shift to an approach where perturbations in images are simulated. The model consists in considering typical photometric and geometric transformations and perturbations that usually apply to images. Thus, starting from a database of representative images and performing synthetic perturbations, we are able to estimate an average covariance matrix of the local descriptor. This approach enables us to improve retrieval results on jets of high orders, as illustrated in Fig. 2.

The proposed approach also allows us to specify more precisely the range where descriptor components are allowed to vary, without enforcing complete invariance, as it is usually performed. For instance, it is possible to constrain invariance only to a range of rotation angles, rather than to use rotation invariants that are less discriminant than jets. Indeed, in our experiments we noticed that it is rarely the case that two images (or parts

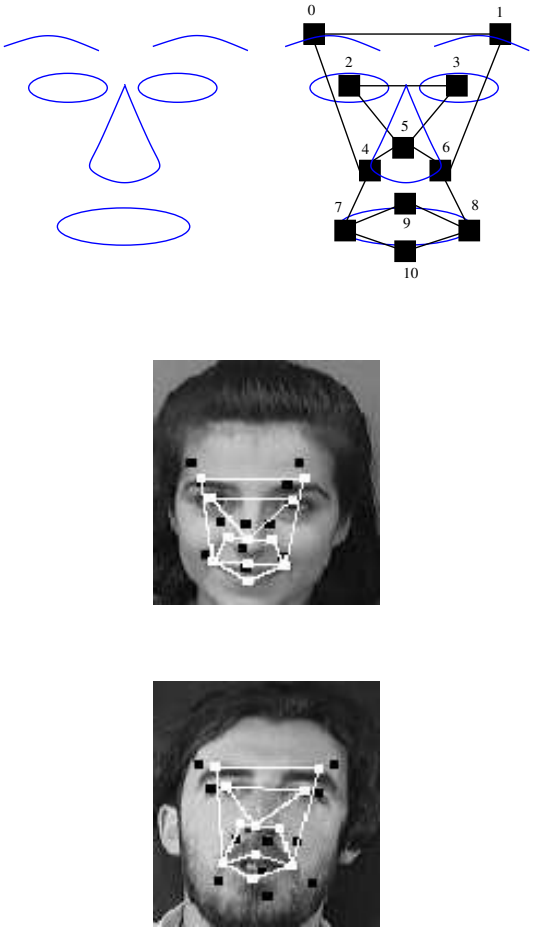

Figure 1. Left: Face components and their underlying graph structure. Right: Face component extraction using the graph structure. The black points correspond to the location of the components using the coarse graph while the white points are those found using the fine graph.

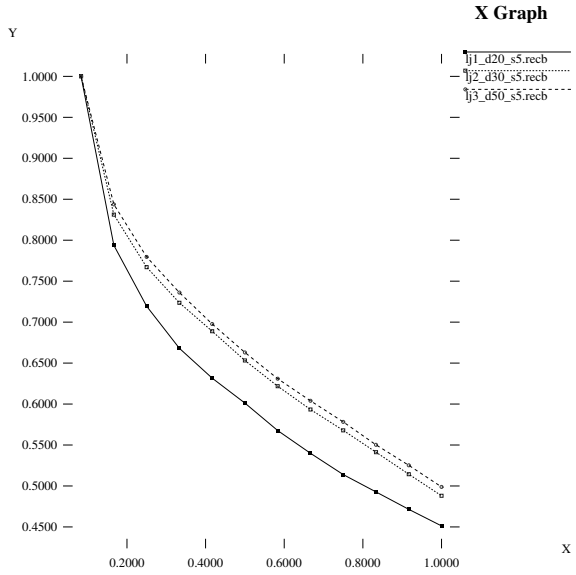

Figure 2. Average precision versus recall, for jets of order 1, 2 and 3. Performance is increasing with the order of the jet.

of images) in correspondence with each other differ by large rotation angles. The proposed approach allows us to choose component weights when the Euclidean distance is used, and must be extended to other kind of distances such as L1. The estimation of the covariance matrix is performed on a databases of representative images. However, it can also be interesting to estimate it from a database of image queries.

6.1.3. Image Segmentation based on ARC clustering algorithm

Key words: *robust clustering*, *adaptive clustering*, *noisy regions*, *coarse segmentation*.

Participants: Hichem Houissa, Nozha Boujemaa.

Our objective is to handle the segmentation shortcomings of the Competitive Agglomeration algorithm. We aim at improving this algorithm by stressing on both shading effects and noise cluster. The Adaptive Robust Clustering **(ARC)** algorithm was used in order to categorize database images when clusters differ significantly in both size and population. We used the same algorithm for region extraction in images.
This segmentation has mainly two advantages. First, when classifying features, a new cluster is taken into account to collect *outliers* defined as aberrant or fuzzy vague features. Such a "noise cluster" contains, among others, pixels located at the frontier of two adjacent clusters in the feature space (non-discriminant u_{ij}) or pixels that are photometrically dissimilar to the detected prototypes. In addition, variable cluster compacities need a specific parameter $\alpha_s(k)$ to avoid merging meaningful but sparsely populated clusters. This shape characteristic is given by the factor $\frac{d^2_{moy}}{d^2_{moy}(s)}$ representing the compactness of a cluster s.

$$\alpha_s(k) = \frac{d^2_{moy}}{d^2_{moy}(s)}\alpha(k) \quad for\, 1 \leq s \leq C \tag{1}$$

We focused on improving the quality of segmentation with the ARC algorithm. We have characterized the noisy regions in images and we have dealt with the gradual shading so that segmented images contain less regions of interest while preserving the semantics.

The figure 4 illustrates the segmented images obtained by the existing CA algorithm.

Figure 3. Original Images

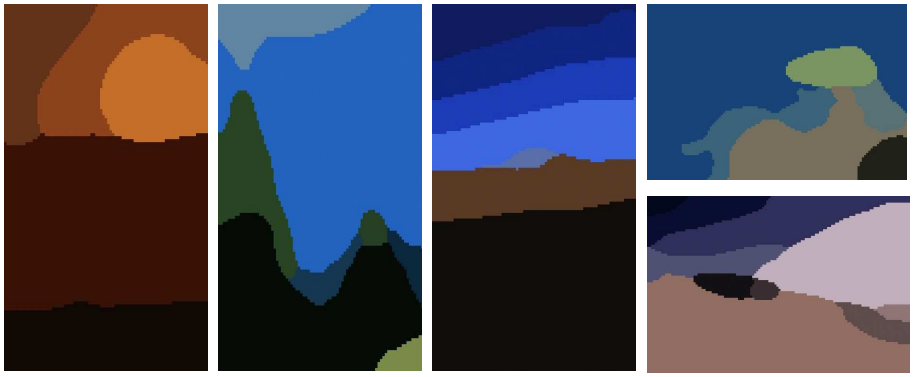

Figure 4. Segmented images with CA

The figure 5 shows the result of the segmentation with the ARC algorithm. Note that the major advantage of such a segmentation is to handle the shading effects mainly due to variable brightness.

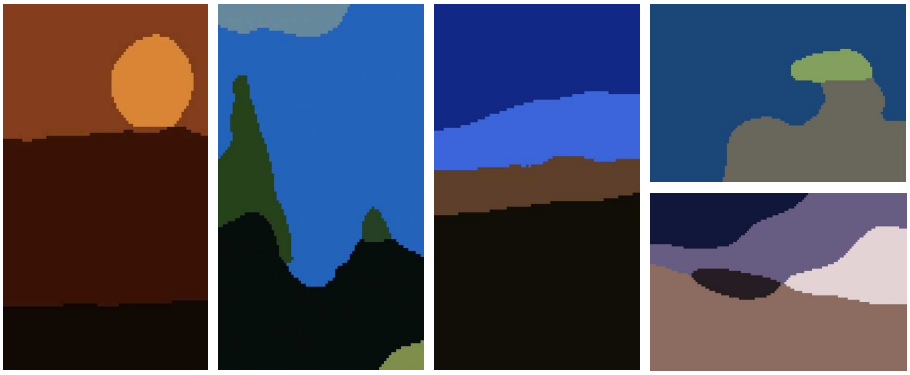

Figure 5. Segmented images with ARC

6.1.4. A skeletal approach for a 3D descriptor

Key words: *3D indexing*, *shape indexing*.

Participants: Cédric Timsit, Anne Verroust.

Computing a 3D shape descriptor is a key issue in shape-based retrieval of 3D models. As in the 2D case, it is used to built indexes and to perform similarity queries. We have choosen to represent the shape of a 3D polygonal model by a skeleton, i.e. a graph composed of 1D skeletal curves, and to take a topological approach based on Reeb Graphs. For this purpose, we have developed a 3D shape descriptor (cf. figure 6) derived from the skeletons of [26] and from the work of M. Hilaga, Y. Shinagawa, T. Kohmura and T. L. Kunii presented at Siggraph'01. Our skeletal shape descriptor is invariant by affine transformations and more robust than the two previous approaches (cf. [16]).

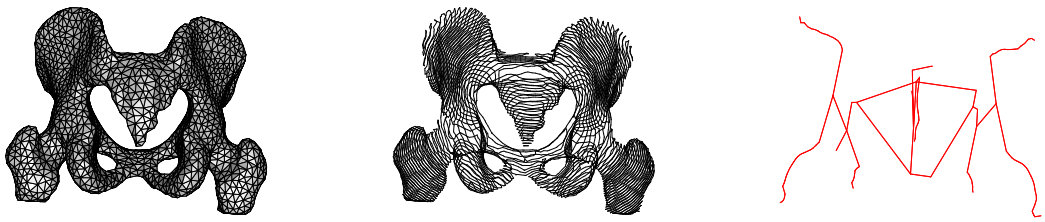

Figure 6. Pelvis bone: 5000 facets, 40 levels, 3 holes / 3 disjoint cycles in the skeleton

6.1.5. Feature spaces structuring

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

Participants: Akram Hentati, Valérie Gouet.

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 our applications. In the continuation of the work started last year, we study the effectiveness of two categories of structuring approaches: one is based on an arborescent structuring of the feature space (SR-Tree like) and the other is based on data filtering (VA-File like). The study already undertaken last year (2002) has been enriched this year jointly with Michel Scholl (CNAM). Indeed, we built representative benchmarks by considering several categories of point distributions, from synthetic data (various distributions of gaussian clusters and uniform distributions) and real data (several databases of various contents were used). Each feature space was generated for several dimensions, in order to perceive the behavior of the algorithms according to the curse of dimensionality problem. Several hardware configurations were also tested, with the aim of not privileging any of the considered approaches. In particular, we studied their performance while varying the memory size of the hardware. The set of bi-processors present on the Rocquencourt site enabled us to consider very high-dimensional distributions (about 10 millions of points). The two approaches were compared by measuring the response times for various point queries (range queries). The results obtained confirmed us that the tree approach is best adapted for the characterization of images by points of interest. The response times obtained with the VA-File approach are lower only for uniform distributions or for very small memory sizes. These configurations do not correspond to real cases of current use of image queries by points of interest. Because the best results on real databases were obtained with the tree approach, we are currently studying an approach that is specific to the point descriptor, based on data clustering.

6.2. Statistical Learning and Object Detection

6.2.1. Feature selection with Conditional Mutual Information

Participant: François Fleuret.

Key words: *feature selection*, *classification*, *information theory*, *mutual information*.

In a context of classification, we propose to use conditional mutual information to select a family of binary features which are individually informative and weakly pairwise dependent. We show that on a task of image classification, despite its simplicity, a naive Bayesian classifier based on features selected with this Conditional Mutual Information Maximization (CMIM) criterion performs as well as a classifier built with AdaBoost. We also show that our method is more robust than boosting when trained on a noisy data set.

By reducing the number of features, one can both reduce overfitting of learning methods and increase the computation speed of prediction. We focus in this paper on the selection of a few tens of binary features among a set of several tens of thousands, in a context of classification. Our approach consists in picking features which maximize their mutual information with the class to predict, conditionally to the response of any feature already picked. This Conditional Mutual Information Maximization criterion (CMIM) does not select a feature similar to already picked ones, even if it is individually powerful, as it does not carry additional information about the class to predict. Thus, it ensures a good tradeoff between independence and discrimination.

Experiments on a face vs. non-face classification task (cf. figure 7) demonstrate that features chosen according to this criterion can be efficiently combined with a naive Bayesian approach and lead to error rates similar to those obtained with AdaBoost. Also, experiments show the robustness of this method when

Figure 7. The two upper rows show examples of background pictures and the two lower rows show examples of face pictures. All those images are grayscale of size 28×28 *pixels, extracted from complete pictures taken on the WWW. Faces are roughly centered and standardized in size.*

challenged by noisy training sets (cf. table 1). It actually achieves better results than regularized AdaBoost, even though it does not require the tuning of a regularization parameter [14].

Beside those performances in term of error rate, this method demonstrates a very high training speed. Our most efficient implementation can perform the training for the classification task described above in a fraction of a second.

Table 1. Error rates with 50 features on a noisy training set whose labels have been flipped randomly with 5% probability.

Classifier	Training error	Test error
CMIM + Bayesian	5.06%	1.95%
AdaBoost$_{reg}$ (optimal)	3.8%	3.06%
AdaBoost	0.58%	6.33%
MIM + Bayesian	9.47%	8.59%

6.2.2. Scale invariance of SVM based on the triangular kernel

Participants: François Fleuret, Hichem Sahbi, Michel Crucianu.

Key words: *support vector machines*, *kernel invariance*, *statistical learning*.

We study the scale-invariance of support vector machines using the triangular kernel, and on their good performance in artificial vision. Our main contribution is the analytical proof that, by using this kernel, if both training and testing data are scaled by the same factor, the response of the classification function remains the same.

For a decade now, Support Vector Machines have proven to be generic and efficient tools for classification and regression. SVMs got their popularity both from a solid theoretical support, and because they clearly untie the specification of the model from the training. The former corresponds to the choice of the underlying kernel, and the later can be done optimally with classical quadratic optimization methods.

We have been using the triangular kernel to build SVM for face detection, which was the most efficient in our experiments despite its poor popularity. We have proven that this kernel makes the training process scale-invariant, and leads to multi-scale classification boundaries (cf. figure 8).

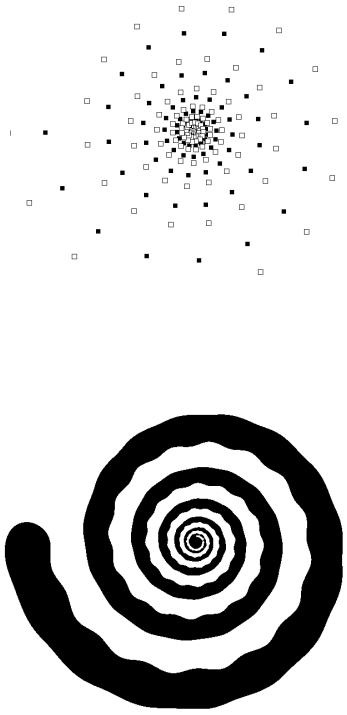

Figure 8. The triangular kernel can separate two populations, even if this requires various scales. Training set is shown on the left and classification with the triangular kernel is shown on the right.

Precisely, if both the training and the testing data are scaled by the same factor, the response of the classification function remains the identical. Formally, if we denote f^γ the classification function learned on a training set scaled by a factor γ, we have

$$\forall \gamma > 0, \forall x, \qquad f^\gamma(\gamma\, x) = f^1(x)$$

The fundamental property of the triangular kernel can be intuitively understood by comparing it with the usual Gaussian kernel at various scales (cf. figure 9). Whereas the former remains identical in shape at all scales, the latter behaves either as a uniform distribution or as a Dirac. These results have been presented in [28] and [10].

We have also shown that the kernel principal component analysis and the kernel discriminant analysis display a scale-invariance property when the triangular kernel is employed (cf. figure 10).

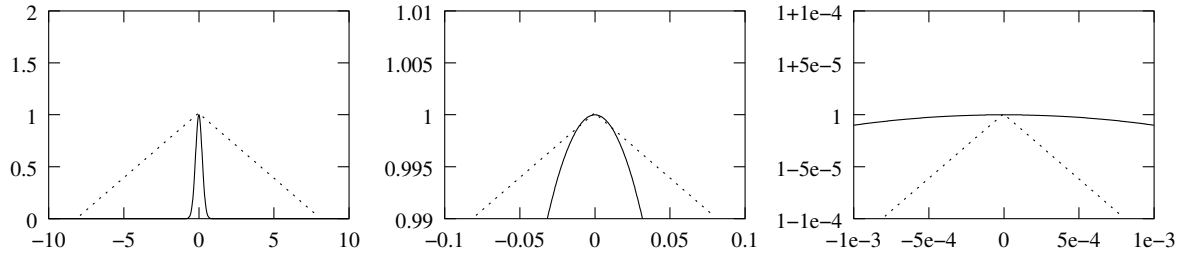

Figure 9. Triangular kernel (dash line) and Gaussian kernel (continuous line) at various scales (left to right, respectively $\times 10^0$, $\times 10^2$ and $\times 10^4$). Intuitively, whereas the triangular kernel is identical in shape at all scales, the Gaussian kernel has different shapes, from a Dirac-like to a uniform weighting of the neighborhood.

Figure 10. KFD with the Gaussian kernel (left) and triangular kernel (right)

6.2.3. Non-Mercer Kernels for SVM Recognition

Key words: *SVM*, *Recognition*, *Point of Interest*, *Positive Kernels*.

Participants: Sabri Boughorbel, Jean-Philippe Tarel, François Fleuret.

One one hand, Support Vector Machines have been employed with significant success for solving difficult pattern recognition problems with global feature representations. On another hand, local features in images have shown to be suitable representations for efficient object recognition. Therefore, it is natural to try to combine the SVM approach with local feature representations to gain advantages on both sides.

The problem of designing kernels for *sets* of local features leads to a space of non-fixed dimension as it is for global feature representations. Moreover, by choosing a kernel, we select how two sets of local features are compared. This comparison is usually performed using matching algorithms to tackle the occluding problem. Therefore, it is desirable that the kernels employed are able, at least, to mimic matching algorithms. When the kernel is a Mercer kernel, the convergence of the SVM algorithm toward a unique optimum is proven. This uniqueness property is one of the main advantages of the SVM compared to other learning approaches. In the case of sets of local features, we give several counterexamples that suggest that matching algorithms are not in general Mercer kernels. Nevertheless, in object recognition experiments, the use of simple matching algorithms as kernels gives good results in practice.

This experimental observation leads us to put forward a new criterion for keeping the uniqueness property of the SVM even for kernels that are not Mercer kernels. This criterion enforces conditions on the kernel as well as on the size of the training set. It insures, in probability, the positivity of the Gram matrix so that matching algorithms can be used as kernels. This criterion can also be used with advantages in many other applications where the SVM approach applies, with local and global representations as well.

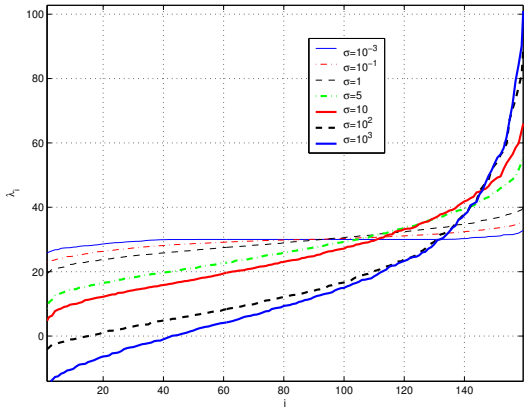

Figure 11. Eigenvalues of the Gram matrix for different values of the kernel hyperparameter.

6.3. Clustering Methods

6.3.1. Unsupervised clustering by adaptive agglomeration

Participants: Bertrand Lesaux, Nozha Boujemaa.

Our objective was to be able to cope with the problem of the "page zero", i.e. how to handle an image collection that the user doesn't know when formulating a visual query. We developed a fuzzy clustering method that is able to categorize the image feature space so as to group visually similar images.

Competitive Agglomeration methods estimate the number of clusters in the data. Such algorithms aim at minimizing an objective function, in which two terms are competing. The first term allows to control the shape and the size of the clusters, while the second one is a regularization term the purpose of which is to control the number of clusters.

We defined a new objective function [2] that formalizes our previous work [12] to cope with clusters of different densities. One competition factor is defined for each cluster, instead of one for the whole data-set. This leads to new optimal update expressions for the membership of points to classes. Then, by defining a competition factor that is proportional to the empirical density of the corresponding cluster, we are able to retrieve both compact and scattered classes.

6.3.2. Semi-Supervised Clustering in complex feature spaces

Participants: Nizar Grira, Michel Crucianu, Nozha Boujemaa.

Feature spaces for image indexing are very complex and high-dimensional. The navigation and the retrieval in these spaces are completely dependent on how well we are able to organize and access image signatures. We have already addressed this problem in our previous work with the ARC algorithm [22] for global image signatures. For point-based retrieval, we deal with local descriptors that have different meaning and dimension. We are interested in exploring other clustering-based methods, more precisely the semi-supervised ones and the kernel-based ones. Our challenging objective remains dealing with the overlap between clusters and with a given amount of outliers present in natural visual feature spaces.
We selected and improved number of them: competitive agglomeration [25], EM, and kernel-based methods (SVC [17]: Support Vector Clustering). After performing tests on synthetic data and real image databases, it can be noticed that these algorithms are unable to offer a configuration of classes that handle the various problems encountered because:

- Several images belonging to different categories may be similar in some features so the influence of the features is generally not equally important in the definition of the category to which similar patterns belong.
- Most similarity measures do not behave well in high-dimensional spaces.

Thus we propose a feature selection and weighting to determine the degree of relevance of each attribute not for the entire dataset but for each cluster.
On the other hand, since we need to partition data automatically, find the optimal number of clusters, identify outliers and learn relevant features, we also invetigate semi-supervised categorization by using a set of labelled samples that can guide the clustering process and generate a better partition of the database.
To deal with clusters having complex (non-elliptical) shapes, a kernel version of the fuzzy c-means clustering algorithm is also proposed (cf. figure 12) and is currently being evaluated.

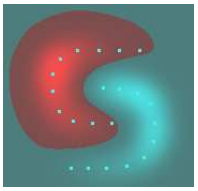

Figure 12. Kernel fuzzy c-means: toy example using the Gaussian kernel

6.4. Interactive Retrieval

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

Key words: *data visualization*, *Euler diagrams*, *graph planarity*.

Participants: Anne Verroust, Marie-Luce Viaud.

In order to propose to the user a semantically structured map of the result of a complex query on a database, we want to develop a tool for automatically computing a diagrammatic representation of the result, derived from the Euler diagrams.
This year we have studied the feasibility of the approach.
We have defined a diagrammatic representation called *extended Euler diagrams* adapted for representing set inclusions and intersections of a collection of sets $X_1, ..., X_n$: each set X_i and each non empty intersection of a sub-collection of $X_1, ..., X_n$ is represented by a unique connected region of the plane. Starting with an abstract description of the diagram, we have defined the dual graph G and reasoned with the properties of this graph to build a planar representation of the $X_1, ..., X_n$. We have shown by a constructive method in [15] that any collection of $n < 9$ sets can be represented by extended Euler diagrams. This result can be interpreted

using hypergraph notions of planarity: any hypergraph having at most eight hyperedges is vertex-planar. This result is optimal: we have shown that when $n = 9$ some collections of $X_1, ..., X_n$ lead to non planar dual graphs.
This work has been performed with Marie-Luce Viaud from INA (Institut National de l'Audiovisuel) within a collaboration agreement between INRIA and INA.

6.4.2. Relevance Feedback for Face Retrieval

Key words: *relevance feedback*, *face retrieval*, *information theory*.

Participants: Yuchun Fang, Donald Geman, François Fleuret, Nozha Boujemaa, Hichem Sahbi.

The proposed work aims at using relevance feedback for the retrieval of mental images of faces. Presently, we concentrate on modelling the process of relevance feedback based on information theory. The final target of the project is a prototype of face retrieval with relevance feedback.

Face retrieval is a specific application in multimedia information indexing. In this project, we dedicate to a special scenario according to which the user has a "mental image" of the face of a target and is looking for that face in a large face database. Our objective is to minimize the number of query-and-answer iterations of relevance feedback during retrieval. The framework of the system is illustrated in Fig. 13.

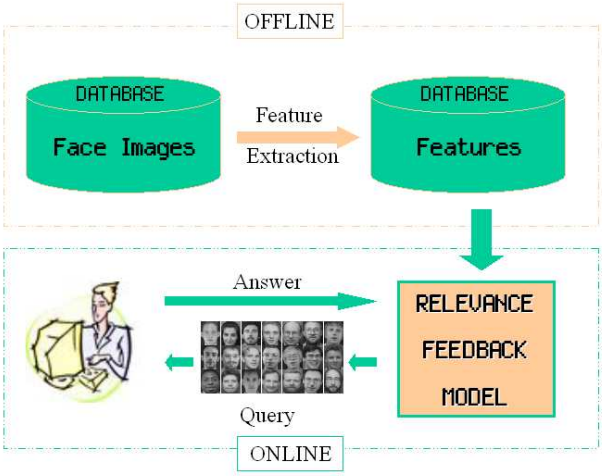

Figure 13. The framework of relevance feedback for face retrieval

It contains two parts: off-line feature extraction and online feedback-based retrieval. The off-line stage aims at constructing feature databases. In the online stage, the relevance feedback machine serves to provide users with a group of face images, and users label them as "relevant" or "irrelevant" by comparing their mental images with the images shown. This process of query and answer continues iteratively until the occurence of the target or until other pre-set terminating conditions are satisfied.

The project contains three main tasks: performing feature analysis, establishing the relevance feedback model and implementing the prototype. For feature analysis, both global and local features are extracted for face images. Several popular representation schemes are compared: principal component analysis (PCA), linear discriminant analysis (LDA), independent component analysis (ICA), kernel-PCA, kernel-DA and wavelet analysis. A relevance feedback model is established based on information theory. The principle is to select images by minimizing mutual information. Since those candidates that minimize mutual information will provide more information about target, it is expected that the target can be searched out with a smaller number of iterations. The proposed model has been tested on a database of simulated features, with simulated users. We have adjusted the parameters for the model. The evaluations are based on statistics of iteration

numbers in each test. Model deteriorates slowly with the decrease of the size of the sampling database and with the increase of the size of the database. An example of average number of iterations with respect to the size of database is shown in Fig. 14. To summarize, the model is stable to various parameters. It works fast according to either the number of iterations or the time required. We currently began experiments on a face database. The first version of the interface is realized on the ORL face database with PCA features of global faces.

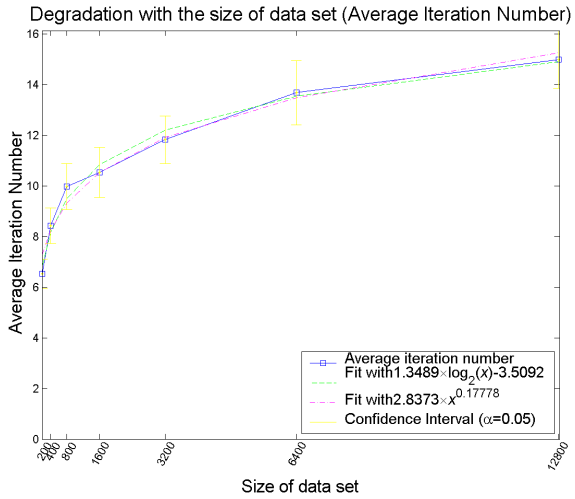

Figure 14. The dependency between the average number of iterations and the size of the database

In the near future, more detailed feature analysis will be performed so that the proposed model can work on large face databases. A web site is under construction for building up a ground-truth database. The interface will be developed after the model passes the test on this database.

6.4.3. Region retrieval with active feedback

Key words: *SVM*, *kernel selection*, *active learning*.

Participants: Marin Ferecatu, Michel Crucianu, Nozha Boujemaa.

Our objective is to let the user express his target in a precise way by using visual parts of images and not only on the entire visual appearance of images. We adapted several signatures to the description or image regions and we implemented SVM-based relevance feedback (RF) mechanisms. As we identified several drawbacks for RF mechanisms that were put forward in the litterature, we realized that we needed a deeper understanding of the relations between some general prior assumptions, the characteristics of the data and the RF mechanisms.

An RF system is composed of a learner and a selection strategy that decides what images are returned to the user at every RF round. Existing approaches use 2-class SVMs as learners and return either the most positive images (MP strategy) or the most ambiguous images (i.e. those that are closest to the decision frontier; this is an active learning strategy, denoted here as MA).

The selection of the kernel is usually performed by a set of experiments, without taking into account the characteristics of the data. On a groudtruth database, by studying the mean distance between the members of each class, we found that the size of the various classes covers a change in scale of 1 to 8. This strongly supports the use of the kernels that reduce the sensitivity of the SVMs to a scale parameter (such as the triangular, power or even hyperbolic kernel). The experimental comparison of several kernels, shown in figure 15 confirmed this observation.

An important general assumption in retrieval is that the images that are relevant to a given user at a given moment are only a small share of the entire collection. Both the direct use of learners that maximize the

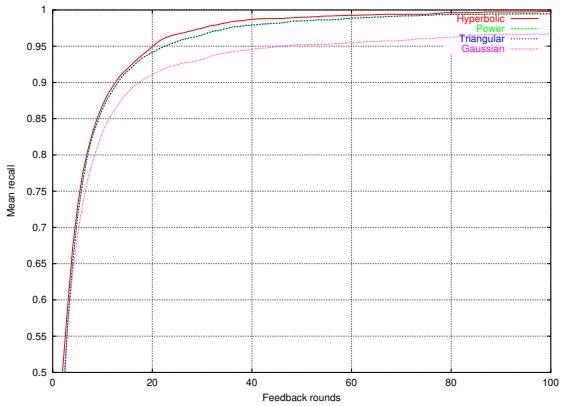

Figure 15. Comparison of several kernels for relevance feedback

margin and the use of an MA strategy appear to ignore this important assumption. The existing MA strategy also ignores two other important facts: (i) in the first (if not all) rounds of feedback the estimation of the frontier is highly unreliable; (ii) efficiency is significantly reduced by redundancies between the MA images selected. In our work we attempt to use more appropriate learners and to replace the MA strategy by a truly "most informative" active learning strategy that takes these observations into account.

6.4.4. Category refinement by relevance feedback

Participants: Bertrand Lesaux, Nozha Boujemaa.

After clustering a database 6.3.1, we provide the user with a tool for refining categories and for personalizing the organization of the image collection. For this, he only has to label a few images as positive or negative to define the class he considers as visually relevant.

The problem of learning a decision rule of the relevance of the images from given samples can be considered as a classification task. Support-vector machines (SVM) are a commonly used classifier in relevance feedback problems. The standard scenario is to iterate clusterings according to the annotated images and presentations of new ambiguous images until the definition of an exact frontier. Such a process is highly time-consuming. However, to learn the classification rule in one step is hazardous, since only a few training examples are available. Moreover, as the few negative examples may not represent the complete distribution of the irrelevant images, there is asymmetry in the training sample. Hence, when training an SVM according to the samples, some images are mis-classified by using the standard decision rule. These ones fall near the frontier of the SVM, in the uncertainty zone between the support vectors.

We propose to consider all the test data to choose a classification rule, and not only the training ones [11]. Images are ranked using the decision function of the SVM. It happens that images from the same real class are usually correctly grouped, so we just have to shift the bias to adjust the frontier to the data. We notice that the decision function presents discontinuities between two distinct classes. We aim to retrieve those "jumps" to be able to detect the class changes. For that purpose, we compute the discrete derivate of the decision function, and look for peaks of this derivate. Finally, the frontier is chosen as the rank of the first peak after the last positive support vector.

This method allows to have a better definition of the class of relevant images with no need to annotate many images. If compared with the mere SVM, it allows to reduce the number of iterations required to refine a given category (cf. Fig. 16).

6.4.5. Retrieve of Multi-modal categories of Images

Key words: *Reconnaissance des formes*, *descripteur de formes*, *comparaison d'images*, *modélisation statistique*, *distance entre images*.

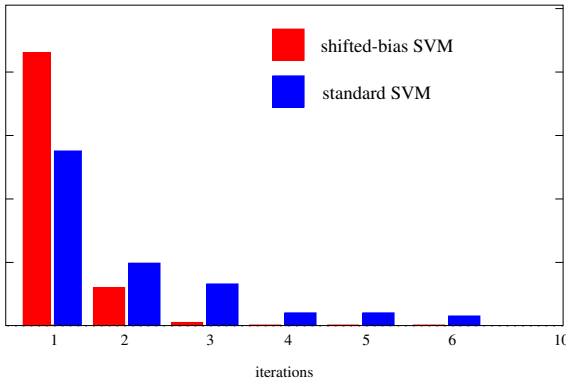

Figure 16. Comparison of the number of iterations required to eliminate all the misclassified images from one category.

Participants: Jean-Philippe Tarel, Hichem Sahbi, Michel Crucianu, Nozha Boujemaa.

By using non-local kernels such as the power kernel, we are able to better retrieve categories of images with several modes, using relevance feedback based on a Support Vector Machine.

The retrieve of categories of images is a learning task that can be solved with advantages using SVM and relevance feedback, in particular to tackle the problem of the search of of a category with several modes, and of the search of a first really relevant image.

The use of a local kernel, such as the classical Gaussian kernel, leads to difficulties in the search of far away modes in the category. First, only one scale is selected by the Gaussian kernel when each mode may have its own scale. Second, when the most relevant images are returned to the user, the Gaussian kernel usually selects images too close to the positive examples, due to its locality.

Therefore, several local and non-local kernels: Gaussian, Laplace, Generalized T-Student and Power kernels were compared. Generalized T-Student kernel (or Hyperbolic) $K(x,y) = \frac{1}{1+\|\frac{x-y}{s}\|^\alpha}$ was prove to be a Mercer kernel. Power kernel $K(x,y) = \|\frac{x-y}{s}\|^\alpha$ is not Mercer, but is usable when $0 < \alpha < 2$ [18]. It turns out that to better retrieve multi-modal categories, non-local kernels such as the power kernel must be preferred over local kernels.

We have also developed an interface prototype for quick validation of SVM relevance feedback that allows the user to better retrieve a first really relevant image in the category he has in mind, as well as to better grow it. One originality is that the user can continuously increase or reduce the set of positive examples. This gives him more freedom to develop different kind of strategies. As an illustration, Fig. 17 shows 30 images of citrus fruits interactively retrieved with our system.

6.4.6. Spatial layout for partial query

Key words: *spatial layout*, *region adjacency*, *force histogram*.

Participants: Hichem Houissa, Nozha Boujemaa, Vincent Oria.

The aim of this work is to model and to represent objects contained in images together with the spatial relationships between the objects in order to enhance image querying. Objects refer to regions obtained by segmenting the images. The method consists in associating the geometry of the objects with the spatial relationships of pairs of objects in addition to the relative distances between them. This is achieved through a *force histogram* that combines both the distance and the orientation from one object to another. For any direction θ and 2 image objects A and B, the Force Histogram computes a function $F^{AB}(\theta)$ over a set of lines

Figure 17. 30 images of citrus fruits obtained by feedback in 62 image clicks from a database of 5601 images. The power kernel with $\alpha = 1$ is used on a color histogram signature.

on A and B sharing the same angle θ. The value of $F^{AB}(\theta)$ is the total weight to support the proposition "A is in direction θ of B". Fuzzy spatial predicates and operators are further defined on the force histograms to express spatial relationships, distances and orientations of pairs of objects. The predicates and the operators are necessary in the definition of expressive image query languages where the user query is expressed as a formula that combines objects, predicates (including spatial predicates) and logic connectives. The main idea behind this is to use the techniques of Geographic Information Systems **(GIS)** for handling the topological aspects and to benefit from the logic languages defined for spatial database queries. This work is based on both contributions of query models for spatial objects [4] and database index representations [5].

6.4.7. New visual query paradigm beyond query by global visual example

Participants: Julien Fauqueur, Nozha Boujemaa, Gouet Valérie.

In early research years in the field of CBIR, the concept of query by visual example (QBVE) has been proposed and shown to be relevant for visual information retrieval. It is obvious that QBVE is not able to satisfy the multiple visual search usage requirements.

First, we have investigated the partial visual query that ignores the background of the images and allows the user to directly express his visual interest without any relevance feedback mechanism. Partial visual selection is allowed through local descriptors by regions or by points of interest, depending on the precision and search time requirements. Discussion of these two different partial image description/selection was published this year in a book chapter [4] (in press) and in a research report [13]. Thanks to this approach the user can explicitly point out the image patch that is relevant for his query.

The second retrieval paradigm allows query by mental image to retrieve images without an image example to start from. This is performed with the new proposed approach: image retrieval query by logical composition of region categories [9]. The user can directly specify the logical composition of visual patches that are the parts of his target mental image. After performing clustering in the image regions feature space, the query interface consists of a visual summary of the image regions available in the database that constitutes a visual thesaurus. A new symbolic indexing and querying approach based on inverted file principle is presented, which relates closely to that of text retrieval. This approach could be easily combined with high-level semantic labelling and querying. This consistuted our on-going work. Simple and generic, it is extendable to multimedia document retrieval indexed by other physical content descriptors.

Work on coarse region segmentation and fine description was published this year in two journals (JVLC [7] and TSI [6]). Julien Fauqueur's PhD thesis [1] has been defended in november 2003).

6.5. Cross-media indexing

6.5.1. Automatic textual face annotation from visual content analysis

Participants: Hichem Sahbi, Nozha Boujemaa, François Fleuret.

The work presented in this section has served for face's information annotation in the TF1 (the French TV channel) video news archives within Mediaworks project [8]. This task has been addressed by coupling a rapid face-detector with a novel face recognition algorithm. A thesaurus of 300 well-extracted face images is used corresponding to 15 persons from the French government. This thesaurus approach is well suited for VIP recognition. This allows a central and updated people (textual and visual) reference for all archivists of TF1. The textual information can be more consistent than only names but also more complete information on function. The video set consists on a news stream of 50 minutes, which was broadcasted by TF1 on May 5th 2002. We sampled the video at one frame each 4 (s), resulting into 750 images containing 1077 faces and we run our face detector on the extracted frames.

The face detection algorithm run on the video frames is based on a hierarchy of support vector classifiers (SVMs) which serves as a platform for a coarse-to-fine search for faces where most of the image is quickly

[4] J-P. Cheiney, V. Oria. Spatial Database Querying with Logic Languages, DASFAA, 1995
[5] P. Rigaux, M. Scholl, A. Voisard, Spatial Databases with application to GIS, Morgan Kaufmann, 2002

rejected as "background" and the processing naturally concentrates on regions containing faces and face-like structures [3] [29][24]. Beside the information about face location, we also get an automatic labeling of the type of shots. Depending on the distance between the eyes, relatively to the scene size, each face is qualified as Long Shot, Medium Shot, or Close Up [8]. From that information, we can provide the user with new requests based on the number of visible faces, their locations in the scene, and the type of shot. In order to identify faces in the thesaurus, a method based on combining a new image representation referred to as "the entropy map" which is invariant to photometric transformations with a flexible pseudo distance based on dynamic programming [27] . This allows us to handle faces with localization errors, occlusion and non-linear deformations. Each detected face image is processed by normalizing its pose and making it upright. Then, we estimate its entropy map, and we compute its matching pseudo distances to the 300 faces belonging to the thesaurus. The identity of the face is inferred using the nearest neighbor (cf. Figures 18).

Figure (a) shows a french politician correctly annotated. Figure (b) shows robustness of recognition and annotation even when there is a transition effect. We also have a false positive detection which is annotated as "non-recognized". This false detection will be filtered by face tracking performed by Vista INRIA Project.

(a) (b)

Figure 18. Example of automatic face annotation after real-time face detection and recognition

6.5.2. Exploiting text-image ressources in biodiversity

Participants: Nozha Boujemaa, Michel Crucianu, Itheri Yahiaoui, Valerie Gouet, Marin Ferecatu, Nizar Grira, Hichem Houissa.

The BIOTIM project (exploiting Text-IMage ressources in BIOdiversity), part of the national initiative "Masses of data", began in October 2004. The web site of the project is http://www-rocq.inria.fr/imedia/biotim/ (French only). The overall goal of this project is to conceive generic methods for the automatic analysis of large amounts of texts and images in order to acquire a common semantic layer and, building upon this initial result, to develop generic methods for a multi-modal examination of the structured data obtained. These methods will be developed and evaluated in two different contexts, the first being part of a recent effort to make the most of the botanical collections available (partner: Institute of Research for Development, IRD in the following) and the second concerning the study of gene expression data issuing from large scale experiments (partner: National Institute for Agronomic Research, INRA in the following). The linguistic component is taken in charge by the INRIA project ATOLL, while the image-related activities are performed by IMEDIA. Both IMEDIA and ATOLL must contribute to the development of the common semantic layer.

Our work on this project started with a careful analysis of the existing text and image corpora. After the extraction by ATOLL of a first vocabulary from the text corpora provided by the IRD, we shall select, together with several botanists, a first set of the visual characteristics (together with the corresponding species) that are mentioned in the texts and can also be identified in botanical pictures. The development of the common semantic layer between images and text will begin with these visual characteristics and the corresponding species.

We also tested our methods on the current image corpora; a demo of interactive retrieval on the first corpus provided by the INRA is available at http://www-rocq.inria.fr/cgi-bin/imedia/ikona (select the `arabidopsis` database).

8. Other Grants and Activities

8.1. National Initiatives

8.1.1. Industrial contract with Sagem

The Project is entitled "Interactive face retrieval" and is a two-stage cooperation over 18 months.

8.1.2. BIOTIM Project (exploiting Text-IMage ressources in BIOdiversity) within the national initiative "Masses of data"

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).

8.1.3. RIAMM Project "MediaWorks"

This project concerns the conception and developpement of a generic platform for multimedia document retrieval by content. This work have been done jointly with linguistic team of LIMSI (CNRS), TF1 content provider and AEGIS entreprise. The originality of this project is the implementation of cross-media indexing and retrieval system by image and text.

8.2. European Initiatives

8.2.1. Integrated European Project "AceMedia"

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

8.2.2. European Network of Excellence "MUSCLE"

"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.

8.2.3. European Network of Excellence "DELOS2"

"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.

8.3. International Initiatives

8.3.1. STIC Project INRIA-Tunisian universities "INISAT"

This project involves the MASC 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.

8.3.2. "Working-Group" NSF-Delos

IMEDIA is involved in the workgroup on digital librairies.

9. Dissemination

9.1. Leadership with scientific sommunity

Nozha Boujemaa :

- Invited Plenary talk for : the Dagstuhl seminar, international conferences(ICISP'03, Taima'03) and professional seminar for archival community.
- Scientific Editor of French Journal "Techniques et Sciences informatiques" Special Issue on Visual Information Retrieval, to appear in 2004
- Conference Programme committee member of
  - SPIE Conference on Storage and Retrieval Methods and Applications for Multimedia 04, IEEE Fuzzy Systems03, CBMI 03, ORASIS 03, ICME 04, ICPR 04, MDDE 04 in conjunction with CVPR 04.
- Jury member:
  - Habilitation for "Applied Computer Science"of Andreas Rauber from University of technology, Vienna - Austria,
  - PhD Itheri Yahiaoui (Eurecom - Sophia Antipolis).
- Scientific coordinator of "Single Modality" WP in Muscle NoE (Network of Exellence FP6), Deputy Scientific coordinator of the Muscle NoE
- Journals reviewer:
  - Multimedia Tools and Applications, IEEE Trans. on Multimedia, IEEE Trans. on PAMI, IEEE Trans. Image Processing, IEEE Trans. on CSVT (Circuits and Syst. for Video Technology).
- Coordinator of the "Imagerie, audio-visuel et applications" panel within French-Tunisian Seminar about "Chaine de l'innovation dans le domaine des STIC: de la recherche vers l'industrie". This seminar is co-orgonized by the Tunisian ministry of scientific research and technology, french ministry of research. French partners include: INRIA, CNRS and the GET.
- Invited presentation of Muscle NoE at the "Centre Français Du Commerce Extérieur" also presented at the European Commission in Luxemburg
- Scientific expert for:
  - "Science & Engineering Research Council (SERC)" - "Agency for Science, Technology & Research (A*STAR)", Republic of Singapore
  - Netherlands Organisation for Scientific Research (NWO) , within Innovational Research Incentives Scheme -VICI - projects, Netherlands
  - French research program ACI (Actions Concertées Incitatives) "Masses of data"
  - Regional Research project funds "Region Basse Normandie"

- In Charge of International and European Relations and member of “Bureau du Comité des Projets” of INRIA Rocquencourt Research Unit

Jean-Philippe Tarel :

- Member of review committie for a special issue of the journal *Techniques et Sciences Informatiques*, 2003.
- Member of programm committie of *RoboCup 2003 International Symposium*, Padova, Italy, July 2003.
- 3h Seminar at DEA ESTC, Filère Vision par Ordinateur, CNAM Paris, February 2003.

Anne Verroust :

- AFIG President (Association Française d’informatique Graphique) ;
- member of the Executive Committee (Conseil d’Administration) of the French chapter of Eurographics;
- in charge of the Computer Graphics part (pôle “informatique graphique”) in the “GdR ALP (Algorithmique, Langage et Programmation)”;
- member of the “commission de spécialistes” of Lille 1 University (27^{th} section).

Michel Crucianu :

- Reviewer for IEEE Trans. on Neural Networks, Neurocomputing, IEEE Trans. on Systems Man and Cybernetics, IEEE Trans. on Pattern Analysis and Machine Intelligence.
- Scientific expert for the French national network RNTL.
- Coordinator of the BIOTIM project (ACI “Masses of data”).

9.2. Teaching

Nozha Boujemaa: course on "Image-based retrieval in Multimedia Databases" at Sup’Com (10h - Tunis).
Anne Verroust: 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).
Valérie Gouet:

- 230 HTD in the Computer Science Department of CNAM ;
- In charge of the option "Computer Vision" of the master ESTC (CNAM - Paris VIII - ENS Fontenay St-Cloud) ;
- Course on Computer Vision in the option "Computer Science" of the Alès School of Mines (27 HTD) ;
- Course "Multimedia databases" in last year of Sup’Com-Tunis (15 HTD).

10. Bibliography

Doctoral dissertations and "Habilitation" theses

[1] J. FAUQUEUR. *Contributions pour la Recherche d'Images par Composantes Visuelles.* Ph. D. Thesis, Université de Versailles, 2003.

[2] B. LE SAUX. *Classification non exclusive et personnalisation par apprentissage : Application à la navigation dans les bases d'images.* Ph. D. Thesis, Université de Versailles, 2003.

[3] H. SAHBI. *Coarse-to-Fine Support Vector Machines for Hierarchical Face Detection.* Ph. D. Thesis, Versailles University, 2003.

Articles in referred journals and book chapters

[4] N. BOUJEMAA, J. FAUQUEUR, V. GOUET. *What's beyond query by example?*. Trends and Advances in Content-Based Image and Video Retrieval, Shapiro, H.P. Kriegel, R. Veltkamp (ed.) LNCS, Springer Verlag, 2003, to appear.

[5] N. BOUJEMAA, M. FERECATU. *Evaluation des Systèmes de traitement de l'information.* Hermes, 2003, chapter Evaluation des systèmes de recherche par le contenu visuel : pertinence et critères, to appear.

[6] J. FAUQUEUR, N. BOUJEMAA. *Recherche d'images par Régions d'intérêt: Segmentation Grossière Rapide et Description Couleur Fine.* 2003.

[7] J. FAUQUEUR, N. BOUJEMAA. *Region-Based Image Retrieval: Fast Coarse Segmentation and Fine Color Description.* volume 15:1, 2003, pages 69-95, to appear.

Publications in Conferences and Workshops

[8] N. BOUJEMAA, F. FLEURET, V. GOUET, H. SAHBI. *Visual content extraction for automatic semantic annotation of video news.* in « IS&T/SPIE Conference on Storage and Retrieval Methods and Applications for Multimedia, part of Electronic Imaging symposium », 2004, to appear.

[9] J. FAUQUEUR, N. BOUJEMAA. *New Image Retrieval Paradigm : Logical Composition of Region Categories.* in « ICIP », 2003.

[10] F. FLEURET, H. SAHBI. *Scale invariance of Support Vector Machines based on the Triangular Ker nel.* in « ICCV2003 workshop SCTV 2003 on Statistical and Computational Theori es of Vision », 2003.

[11] B. LE SAUX, N. BOUJEMAA. *Image database clustering with SVM-based class personalization.* in « IS&T/SPIE Conference on Storage and Retrieval Methods and Applications for Multimedia », january, 2004, to appear.

[12] B. LE SAUX, N. GRIRA, N. BOUJEMAA. *Adaptive Robust Clustering with Proximity-Based Merging for Video-Summary.* in « IEEE International Conference on Fuzzy Systems (FUZZ-IEEE'2003) », may, 2003.

Internal Reports

[13] N. BOUJEMAA, J. FAUQUEUR, V. GOUET. *What's beyond query by example.* Technical report, number RR-5068, INRIA, 2003, http://www.inria.fr/rrrt/rr-5068.html.

[14] F. FLEURET. *Binary Feature Selection with Conditional Mutual Information.* Technical report, number RR-4941, INRIA, october, 2003, http://www.inria.fr/rrrt/rr-4941.html.

[15] A. VERROUST, M.-L. VIAUD. *Ensuring the Drawability of Extended Euler Diagrams for up to 8 Sets.* Technical report, number RR-4973, INRIA, October, 2003, http://www.inria.fr/rrrt/rr-4973.html.

Miscellaneous

[16] C. TIMSIT. *Descripteur d'objets 3D utilisant les squelettes.* Rapport de DEA, DEA IARFA, Paris, September, 2003.

Bibliography in notes

[17] A. BEN-HUR, D. HORN, H. T. SIEGELMANN, V. VAPNIK. *Support Vector Clustering.* volume 2, JMLR and MIT Press, March, 2002.

[18] C. BERG, J. CHRISTENSEN, P. RESSE. *Harmonic analysis on semigroups: theory of positive definit e and related functions.* Springer Verlag, 1984.

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

[20] N. BOUJEMAA, J. FAUQUEUR, M. FERECATU, F. FLEURET, V. GOUET, 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.

[21] N. BOUJEMAA, M. FERECATU, V. GOUET. *Approximate search vs. precise search by visual content in cultural heritage image databases.* in « Invited paper in MIR workshop in conjunction with ACM Multimedia », 2002.

[22] N. BOUJEMAA, B. LE SAUX. *Unsupervised Robust Clustering for Image Database Categorization.* in « IEEE-IAPR International Conference on Pattern Recognition (ICPR'2002) », Quebec, Canada, August, 2002.

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

[24] F. FLEURET, D. GEMAN. *Coarse-to-fine visual selection.* in « In International Journal of Computer Vision », number 2, volume 41, 2001, pages 85–107.

[25] H. FRIGUI, R. KRISHNAPURAM. *A Robust Competitive Clustering Algorithm with Applications in Computer Vision.* in « PAMI », number 5, volume 21, Mai, 1999, pages 450-465.

[26] F. LAZARUS, A. VERROUST. *Level Set Diagrams of polyhedral objets.* in « SMA'99 (Fifth ACM Symposium on Solid Modeling and Applications) », ACM Press, Ann Arbor, June, 1999.

[27] H. SAHBI, N. BOUJEMAA. *Robust face recognition using Dynamic Space Warping..* in « In Proceedings of Springer Verlag Lecture Notes In Computer Science. ECCV's Workshop on Biometric Authentication. », 2002, pages 121–132.

[28] H. SAHBI, F. FLEURET. *Scale invariance of Support Vector Machines based on the Triangular Ker nel.* Technical report, number RR-4601, INRIA, october, 2002, http://www.inria.fr/rrrt/rr-4601.html.

[29] H. SAHBI, D. GEMAN, N. BOUJEMAA. *Face detection using Coarse-to-Fine Support Vector Classifiers..* in « In Proceedings of the IEEE International Conference on Image Processing. », 2002, pages 925–928.

[30] L. WISKOTT, J. FELLOUS, N. KRUGER, C. MALSBURG. *Face Recognition by Elastic Bunch Graph Matching.* in « In Pattern Analysis and Machine Intelligence », 1997, pages 775–779.