INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Project-Team Imedia

Images and Multimedia : Indexing, Retrieval and Navigation

Rocquencourt

THEME COG

Activity Report

2004

Table of contents

1. Team

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

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

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

Research scientist (partner)
Valérie Gouet [Assistant Professor at CNAM]
Jean-Philippe Tarel [Research Scientist (CR1) LCPC]

Scientific advisor
Donald Geman [Professor at Johns Hopkins University]

Invited Professor
Peter Belhumeur [Professor at Columbia University (NY)]
Vincent Oria [Research Scientist, visiting since 01/06/2004]

Post-Doctoral fellow
Yuchun Fang [Post-Doctor with Sagem since 01/05/2003]
Ithéri Yahiaoui [Post-Doctor since 17/11/2003]

Expert Engineer
Hichem Sahbi [Sagem project since 01/08/2004]

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 Law-To [CIFRE grant with INA since 1/5/2004]
Olfa Besbes [Joint tutorship with Sup'Com, national grant since 1/9/2004]

Student intern
Mohamed Chaouch [Internship Polytechnic School-Tunis since February 2004]
Josselin Chièze [Internship EPITA since June 2004]
Nicolas Hervé [Internship CNAM since August 2004]

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 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, Ithéri Yahiaoui, Hichem Houissa, Sabri Bougorbel, Jean-Philippe Tarel, Nizar Grira, Anne Verroust, 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, 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 [28] [34].

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.

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

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, 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. [7]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

Keywords: *User interface*, *image retrieval by content*, *relevence feed-back*.

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

Correspondents: Marin Ferecatu, Paul-Paul Chièze, Nicolas Hervé

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 [29]. 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 [30];
- **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. Visual saliency

Keywords: *Focus of attention*, *Local descriptors*, *Real-time monitoring*, *Spatio temporal descriptors*, *Visual saliency*.

Participants: Julien Law-To, Valérie Gouet, Nozha Boujemaa, Olivier Buisson, Alexis Joly.

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 application considered is real-time monitoring of huge databases of videos (about 100000 hours), and more particularly video content-based copy detection.

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 content based video copy identification 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 such real-time monitoring, it is necessary to build a very compact description. To do that, we are investigating two kinds of approaches:

- The spatio-temporal information must be exploited jointly, instead of working on the time dimension and the spatial dimension separately.
- To be more compact and more significant, we are working on new signatures that should be inspired by preattentive human vision and the focalization of attention. Such features have been widely studied by the neurophysiological community and can be modelized mathematically.

6.1.2. Combination of descriptors for region-based classification

Keywords: *FastMap*, *adaptive classification*, *feature-space reduction*, *region description*.

Participants: Hichem Houissa, Nozha Boujemaa.

Our objective is to achieve classification of different adaptive color distributions in order to generate a new robust visual thesauri firstly introduced in [4]. Visual thesaurus is a new way to achieve mental image retrieval [1]. We described the extracted regions (by means of adaptive competitive agglomeration algorithm [38]) with their Adaptive Distribution of Color Shades (ADCS) [5]that take into account a fine distribution of pixel colors in the region.

These distributions are, on the one hand, very accurate for representing the color distribution in the region, but, on the other hand, disturb the classification process because of their adaptiveness. Each ADCS has its own dimension which is automatically determined by the pixels color classification. We focused here on the way to classify distributions that were described in variable feature-space dimensions.

Some quantization methods were applied to embed all the data in a fixed dimension subspace, but these approaches have an antagonist effect regarding the ADCS approach: the information we obtained by fine description of the regions is discarded by rough quantization of the ADCS bins, thus other clustering approaches have been investigated in order to handle multi dimensional spaces.

The method we tested is based on spectral clustering [44]and relies on features space reduction using the FastMap algorithm [33]. The main advantage of this approach is that it handles only objects (in our case regions) in a projected subspace that keeps the notion of distance almost unchanged with regard to space distortion and mapping. This is achieved by defining an affinity matrix between data where dimension of each feature is not involved. Data are then mapped into a $k - dimension$ subspace where each object (in our

case each region) is characterized by its k principal bins instead of n variable number of bins. Once principal axis defined [33], we achieve a clustering algorithm on the embedded data. Our method is in progress and preliminary results should be available soon. In our first tests, we focused on the reliability of the spectral clustering- its ability to cluster correctly multi dimensional descriptors- since this approach seems to be the most efficient one to take into account the different visual descriptors provided to us for intelligent search.
This study is important for combining descriptors knowing that our objective is to build a visual thesauri that handles photometric features (colors) and also region texture and granularity. Thus, different descriptions should be used for extracted regions so that the visual thesauri gives to the user a global overview of the database content and let him compose his query according to visual patches strongly representative.

6.1.3. Low-level features for plants description and retrieval

Participants: Ithéri Yahiaoui, Nozha Boujemaa.

This work is involved within Biotim project which concerns the feasibility study of a multi-modal interactive retrieval based on text and image.

We use two image databases, the first one is from IRD. The task consists to plant species recognition for both expert and large public users. The second image database is provided by INRA, it is used to extract quantitative measurements for the study of genetic modification expressions. We focus in the first stage on the INRA application with gene expression.

First, we adopted a global approach which consists in measuring the similarity of the images according to their general appearances. For this we calculate descriptors by using all image pixels. These results are visible on an online web demo on "arabidopsis" collection provided by NASC (Nottingham Arabidopsis Stock Centre http://arabidopsis.info/) see http://www-rocq.inria.fr/cgi-bin/imedia/ikona This demo is also reachable via NASC web site (http://seeds.nottingham.ac.uk/Nasc/action.lasso?-token.user=17112963248&-response=/Nasc/picbook/picture_book.lasso) The areas of image backgrounds is more than 30% of total image area. Thus descriptors are skewed, retrieval performances are decreased. Therefore, we choose to eliminate useless information and to retain only the essential part of the image by using partial queries . We define zones of interest corresponding to the plants on which we compute specific descriptors more adapted to these regions of interest.

In order to separate the plant from the background, image is segmented using a competitive classification algorithm. Knowing that the plants in the INRA database underwent genetic modifications, the color of the leaves was faded. Consequently the fine segmentation that we carried out caused division of the region corresponding to the plant in several regions. Thus the faded zones of the plant which represent the genetic modifications can be separated as shows it the following examples Fig 1 and 2:

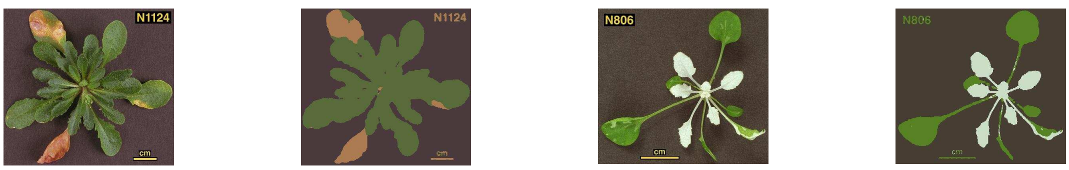


Figure 1.

Fine segmentation results are useful to make quantitative geometric measurements such as the areas of the sheets having undergone modification of color with respect to genes modification. On the other side, to identify a single mask for the plant (the principal zone of interest), we have to achieve a coarse segmentation. Therefore, we model the image appearance by using local color distribution. Similarity measure is then modified to handle appropriate data distribution clustering. Here some examples of plant masks obtained:

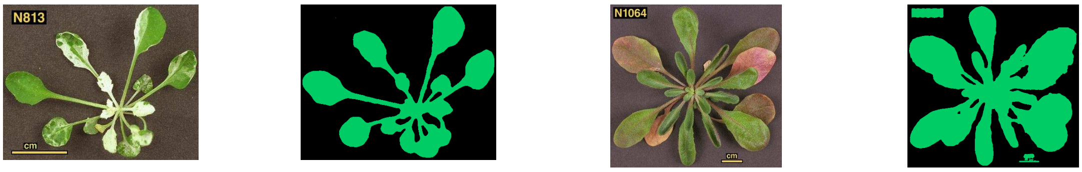


Figure 2.

Once plants masks were built, existing local descriptors in Ikona were adapted and then computed . We designed a family of new shape descriptors which characterize the outline of the mask (external shape of the plant). Fig 3 relates one preliminary result showing retrieval of similar plants appearance with different background.

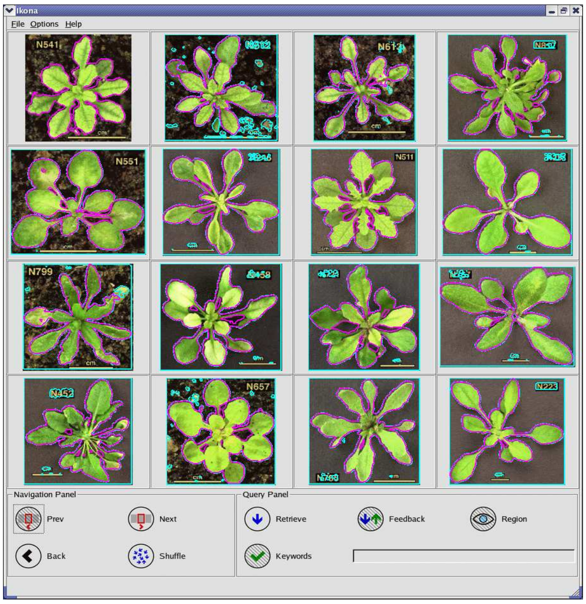


Figure 3. Plants retrieval

6.1.4. A Hough transform-based 3D descriptor

Keywords: *3D indexing*, *anisotropy*, *shape indexing*.

Participants: Mohamed Chaouch, Anne Verroust-Blondet.

This year, we have implemented and compared the efficiency of several existing shape descriptors on the Princeton Shape Benchmark database [46]. We have improved the efficiency of the descriptor DH3D based on a 3D Hough transform [50]: we have introduced a new descriptor, the "ellipsoidal Hough 3D descriptor" (EDH3D), which is more robust to anisotropic deformations.

This improvement is particularly visible on the "dining chair" class (cf. figure 4 and 5). This work is described in [21].

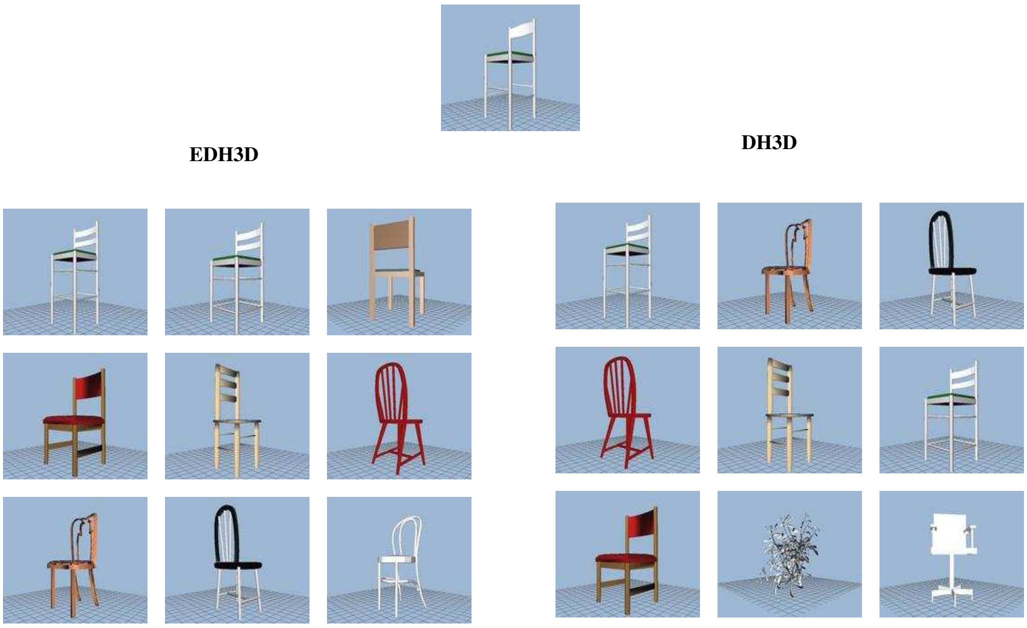


Figure 4. Retrievied models for a query within the "dining chair" class

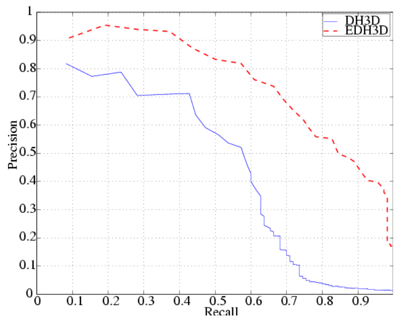


Figure 5. Precision-recall curves for the "dining chair" class

6.1.5. Affine Invariant Shape Description Using the Triangular Kernel

Keywords: *Statistical learning*, *image retrieval*, *kernel principal component analysis*, *scale-invariance*, *shape description*, *triangular kernel*.

Participant: Hichem Sahbi.

Solving the *semantic gap* in content based image retrieval basically requires relevant low level characteristics, also referred to as descriptors, including color, texture and shape. The latter may be efficient to capture the discriminant information for many applications, for instance tracking silhouettes, character recognition, contour matching for medical imaging, $3D$ reconstruction, etc. Many signatures exist in the literature for the purpose of shape and curve description; among them the well studied edge orientation histogram, radon transform, etc. Other methods range from those based on: learning shape statistics [32], geometry and a priori knowledges such as the curvature scale space (CSS) [43], skeleton and axial representations [45], context extraction and alignment [24], algebraic description and invariant moments [42]. A good shape description should be robust to affine transformations and also to the local non-linearities, the noise and the mirror effects. Kernel principal component analysis (KPCA), also known as the non-linear version of PCA, considers a positive definite kernel $k(x, x') = \langle\Phi(x), \Phi(x')\rangle$ where $\langle, \rangle$ stands for the inner product and Φ is a mapping from an input space into a higher (possibly infinite) dimensional space referred to as *the feature space*, where linear PCA can be performed. In contrast to the linear case, this non-linear version is affine invariant in the feature space but not in the input space when using many kernels for instance the well studied Gaussian. In this work [19], we show that the proposed triangular kernel [18] achieves this affine invariance both in the feature and the input spaces. The VC dimension of a family of hypotheses (eigenspaces) trained using this kernel is infinite, it follows that the dimension of span of the underlying principal directions increases with respect to the size of the training set. This increase of dimensionality makes it possible to capture the non-linearities in shape geometry.
We ran our experiments on the SQUID[1] database consisting of 1100 curves. Each curve in the SQUID database containing between 400-1600 points is randomly sampled in order to extract 128 ($2D$) training points which were used to synthesize 4 other curves with random orientations (in $[0^o, 360^o]$), scale factors (in $[0, 2]$) and locations (in ± 20 pixels). At the end, a total of 5500 curves were used in our experiments. The triangular kernel is rotation and translation invariant [35], so under these transformations, the eigenvalues of KPCA remain unchangeable. Only scaling the data with γ, scales the eigenvalues by γ^p. Hence, the eigenvalues $\{\lambda_i^{(\gamma)}\}$ can be normalized with respect to the largest value $\lambda_1^{(\gamma)}$ in order to cancel the factor γ^p. Therefore, $\{\lambda_i^{(\gamma)} \ / \ \lambda_1^{(\gamma)}\}$ will be scale invariant and can be used as an affine description of a given curve.
Figure (6) illustrates some results; for each submitted query, the system finds first the 4 most similar shapes which differ only by affine transformations, then the system finds the other similar curves with an increasing order of the Mahalanobis distance. Figure (7) shows pairs of good matches between points taken from two different curves. Given two curves, the matching criteria is based first on projecting their points in the span of the underlying eigenvectors. Afterwards, each point in one curve is assigned to an other point in the second curve which minimizes an L_2 distance. Notice that we do not consider any smoothness or neighborhood criteria such as: every two neighboring points in one curve should have neighboring matches in the other one; this can reduce false matches.

6.1.6. Comparing local descriptors and fine similarity measures

Keywords: *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 (see for example a survey on points of interest in [11]). The use of points of interest allows us to make queries on parts of images, as well as on objects contained in an image.

[1] www.ee.surrey.ac.uk/Research/VSSP/imagedb/squid.htm

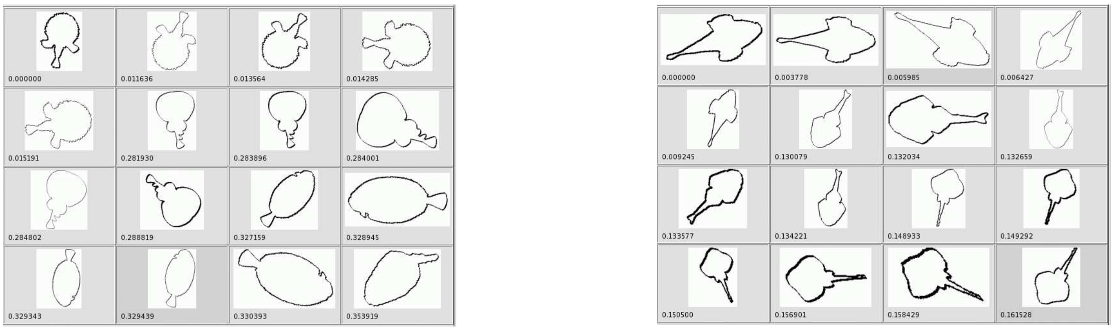


Figure 6. Top-left images (in both sessions) are query shapes while the others are some results sorted from top-left to bottom-right according to their dissimilarity.

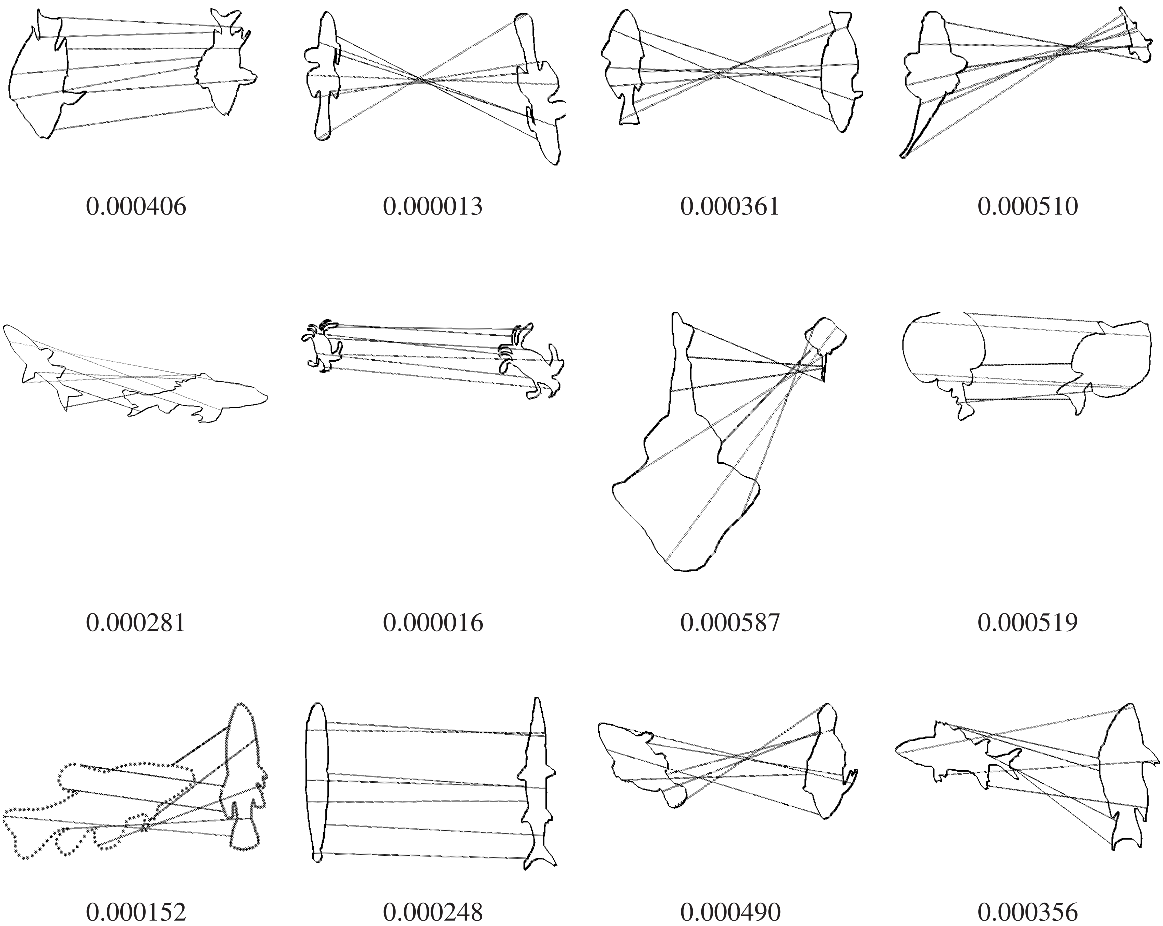


Figure 7. Some matching results and the underlying scores. For ease of visualization only a subset of matches is displayed.

We have compared performances of four kinds of local shape descriptors: edge orientation histogram (EOH), Hough transform (Hough), and two different algebraic fits of the local contours (Fit1 and Fit2). For each signature, parameters were optimized to provide best results on an reference database. Every descriptors are of the same size. The two best local shape descriptors, using Mahalanobis distance, are the Fit2 and Hough signatures as shown in Fig. 8. Compared to jets order 2 which are very localized, the advantage of these shape descriptors is that performances increase when the size of the window of focus increases.

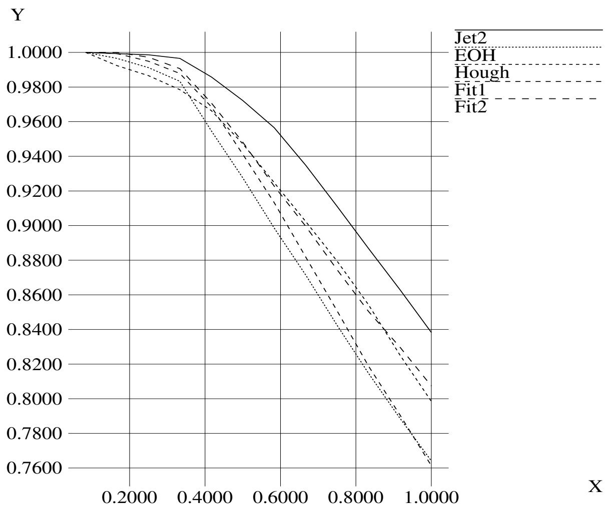


Figure 8. Comparison of average precision versus recall for jets of order 2 (Jets2), edge orientation histogram (EOH), Hough transform (Hough), two algebraic fits of the local contours (Fit1 and Fit2).

To compare two points of interest, the similarity measure must be carefully chosen to achieve the best performance. In particular, descriptor components do not have the same scale of variations. The problem is then to correctly estimate these scale of variations. An first approximate can be obtained with the covariance matrix computed on the whole set of points of the database (i.e. Mahalanobis distance). Better covariance matrices can be obtained using simulated perturbations in images. These perturbations consist in applying typical photometric and geometric transformations, in adding noise. This approach enables us to improve retrieval results on jets of high orders, as illustrated in Fig. 9(a).

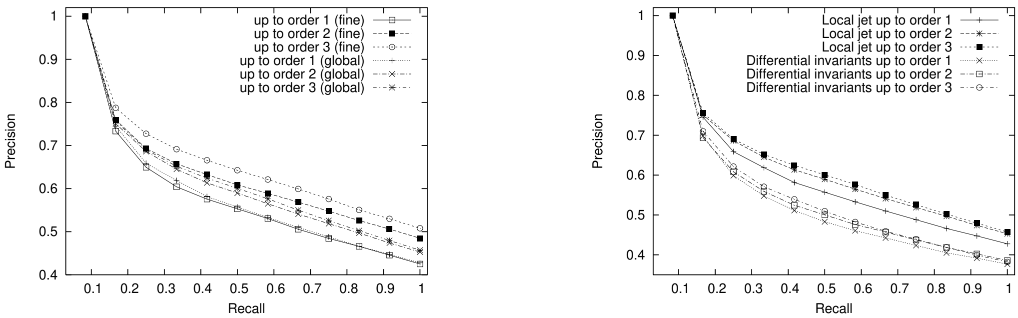
This solution also allows us to specify more precisely the range where descriptor components are allowed to vary, without enforcing complete invariance, as it is usually performed. For instance, it is possible to constrain invariance only to a range of rotation angles. This is illustrated in Fig. 9(b) where the best results are obtained with the jet descriptors, rather than using Hilbert invariants which are invariants to rotation too. This comparison was performed using the Columbia color database, where objects differ only from small image rotation.

6.1.7. Feature spaces structuring

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

Participants: Rochdi Bouchiha, Valérie Gouet, 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



(a) Comparison of descriptors computed with two kinds of weights

(b) Comparison of two kinds of descriptors related to different degrees of freedom

Figure 9. Average precision versus recall obtained on the Columbia coil-100 database. (a) for local jets up to order 3 using a similarity measure with weights estimated from the whole database (global) and with ones estimated from synthetic small perturbations (fine); (b) for local jets and Hilbert's invariants up to order 3.

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 gain scalability for 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.

When considering objects or parts of images described with a set of local descriptors, searching in the feature space is usually done independently and sequentially for each descriptor. This year, we study multiple query approaches [31][27][25] existing in the Database community and we apply them to the retrieval of groups of local descriptors.

We are investigating to axes:

- Reduction of I/O cost, by studying a new approach for searching in a multidimensional structure. The structure currently considered is the SR-Tree one (see for example a survey on index structures in [12]), which is less time consuming than other approaches for feature spaces based on local descriptors we are investigating;
- 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.

This work has been done in collaboration with the CEDRIC/Vertigo research group, during the internship of Rochdi Bouchiha (ENIS, Tunisia) at CNAM [20].

6.2. Interactive Retrieval

6.2.1. Interactive Retrieval of Mental Face Images

Keywords: *face retrieval*, *image retrieval*, *mental retrieval*, *relevance feedback*.

Participants: Yuchun Fang, Donald Geman, Nozha Boujemaa, Jean Paul Chièze, Hichem Sahbi.

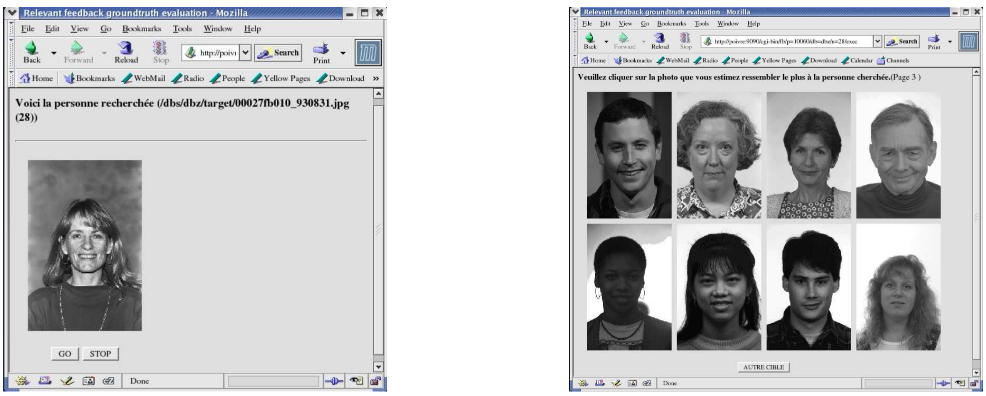
The problem of mental face retrieval is a target search problem. A Bayesian model is constructed to implement the process of query and answer. For realization, special efforts have been paid on three key problems: coherence analysis, answer model design and display model design.

- Answer model

The form of answer model is decided by the specific research object: face images. Two types of answer strategies have been compared: relevant/irrelevant response and comparative response. The latter one is proved to be more realistic by real user experiments.

- Display model
  For the aim of quick retrieval, each iteration should catch as much as possible information about target from user. Such idea is formulated by maximizing mutual information between target and answer. Stimulated by such formulation, two heuristic solutions have been proposed and compared as display model.
- Coherence analysis
  The coherence in the proposed work refers to the coherence of image similarity/dissimilarity between machine and human. Semantic gap and subjectivity in human responses are two major origins of incoherence. In the proposed work, we make an average evaluation of coherence among many kinds of candidate signatures under the setting of comparative response.

For adapting the model to real user response, a web interface is designed both to collect user response and to verify the performance of relevance feedback model as shown in Fig.10. The first page shows the target as in Fig.10 (a), and several groups of display are shown in preceding pages as in Fig.10 (b).

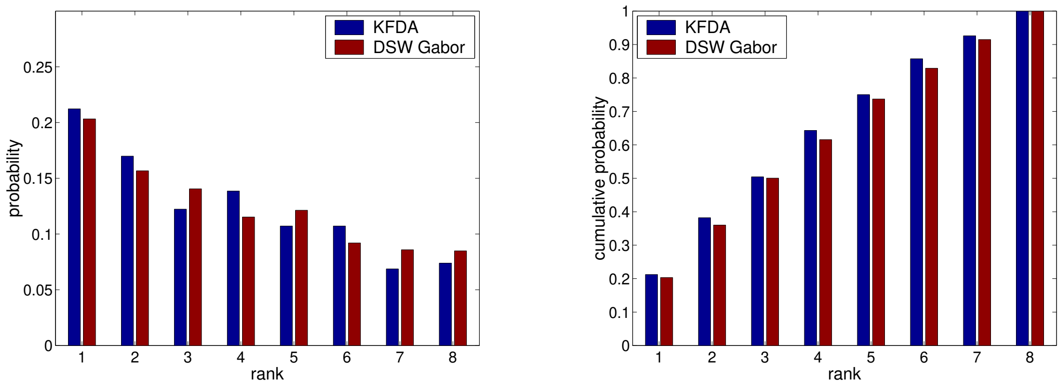


(a) (b)

Figure 10. Web interface for comparative answer scheme: (a) the page of target; (b) the page of display.

Coherence is evaluated with both probability and cumulative probability of users' choice of each rank of display: the 1-st closest, 2-nd closest,...,8-th closest to target on data set of real user responses. Examples are shown in Fig.11, which are the results with two signatures Kernel Fisher's Discriminant Analysis (KFDA) and Dynamic Space Warping on Gabor features (DSW-Gabor). The experiment shows that various signatures possess similar coherence.

Several groups of real user tests have been done to evaluate the Bayesian relevance feedback model. The result is evaluated by the curve of cumulative probability with respect to iteration numbers. In Fig.(12), the experiment is done by 22 INRIA researchers, in total 78 tests have been performed. In comparison, we also show the curve of simulation and random display. The result proves the feasibility of mental face retrieval with the proposed model.



(a) (b)

Figure 11. Coherence analysis: (a) Probability with respect to rank; (b) Cumulative probability with respect to rank.

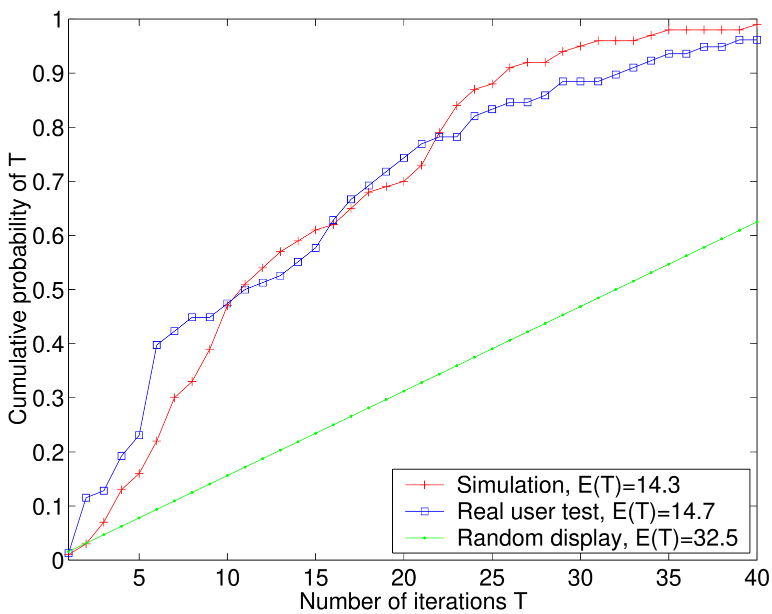


Figure 12. Real mental retrieval test with familiar faces as target

6.2.2. 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, we want to a tool for automatically computing a "Euler like" diagrammatic representation of the result. The usability of such an interface in a traditional library environment is presented in [16]. After having studied the feasibility of our approach (we have showned in [17] that a query involving at most eight different terms can be represented by a planar extended Euler diagram), we are currently implementing a system drawing automatically a planar extended Euler diagram from any configuration involving less than nine sets.

6.2.3. A Comparison of User Strategies in Image Retrieval with Relevance Feedback

Keywords: *angular kernel*, *labelling strategy*, *performance measure*.

Participants: Michel Crucianu, Jean-Philippe Tarel, Marin Ferecatu.

Since large-scale experiments with real users are costly and difficult to set up, the evaluation and the comparison of relevance feedback (RF) algorithms usually rely on the use of image ground-truth databases and on the emulation of the user by a strategy. In general, the used strategy consists in labeling all returned images as either "relevant" or "irrelevant" without any mistake. But the behavior of *real* users can not be expected to comply to strict specifications. Moreover, it seems reasonable to expect that the choice of a strategy has an impact on the quality of the RF results. How general, and thus meaningful, are then the conclusions drawn from comparisons performed with ground-truth databases? To study this issue, we also emulate the user behavior and rely on ground-truth databases. However, in our evaluation, we use multiple strategies as well as several image databases of different complexity. The comparison we focus on is between the MP (most positives) selection criterion and the MAO (most ambiguous and orthogonal) criterion introduced in [8]. We use SVM-based RF with both the RBF kernel and the angular (or not compact triangular) kernel. Among the labeling strategies, we consider various partial labeling methods, as well as making mistakes while labeling.

We first find that the ranking between MP and MAO does not depend much on the selected strategy. Second, the angular kernel performs constantly better than the RBF one, and the difference increases with the complexity of the classes. Third, the ranking between strategies appears to be relatively independent of the complexity (roughly corresponding to the semantic level) of the ground-truth classes. This robustness property with respect to variations in the user strategy and in the complexity of the database is a very desirable property when designing relevance feedback systems that should be effective for most of the users. We suggest that comparisons between RF algorithms should be conducted with several different strategies to evaluate the stability of the results with respect to changes in the user strategy.

Since the ranking of user strategies is relatively stable with respect to class complexity and rather similar with the two selection criteria, a user strategy *can* be advised to real users. We notice that the best strategy goes against conventional wisdom, in that it requires to avoid labeling too many negative examples.

6.2.4. Samples selection with redundancy reduction for SVM-based relevance feedback.

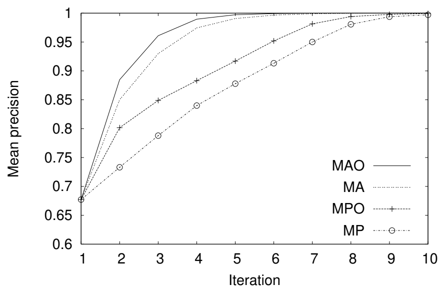
Keywords: *active learning*, *image retrieval*, *reduction of redundancy*, *relevance feedback*, *sample selection*.

Participants: Marin Ferecatu, Michel Crucianu, Nozha Boujemaa.

Active learning (see [49] [48]) is widely used with relevance feedback (RF) in image retrieval to help user refine queries or to find complex/semantic classes of images. In this case, the images selected by the RF machine to be presented to the user are closest to the separating hyperplane defined by the SVM, i.e. the most ambiguous (MA) in terms of the decision criterion. While this principle works well for the selection of one example to be presented to the user, it is significantly suboptimal for selecting several examples.

In [8] we put forward a new active learning selection strategy that minimizes redundancy between the images presented to the user and takes into account assumptions that are specific to the retrieval setting. If

x_i and x_j are the input space representations of two candidate images, then we require a low value for the value taken by the kernel for this pair of images, $K(x_i, x_j)$. If the kernel K is inducing a Hilbert structure on the feature space, if $\varphi(x_i)$, $\varphi(x_j)$ are the images of x_i, x_j in this feature space and if all the images of vectors in the input space have constant norm, then this additional condition corresponds to a requirement of orthogonality between $\varphi(x_i)$ and $\varphi(x_j)$. We call this criterion the selection of the "most ambiguous and orthogonal" (MAO) candidates. This criterion can also be applied, with a different account, for conditionally positive definite kernels.



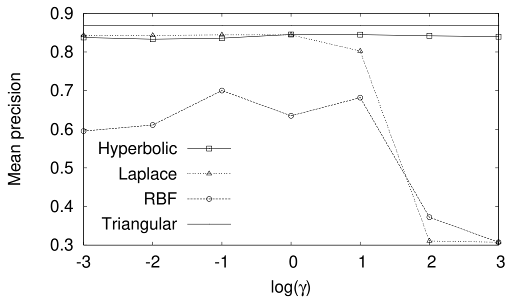


Figure 13. (a) Comparison of several selection strategies and (b) sensitivity of the SVM to the scale parameter for the different kernels (MAO selection criterion).

In Fig.13(a) we see the results of comparing several selection criteria: MA, MAO, MP (most positive) and MPO (most positive and ortogonal). The groundtruth database we used here has 52 classes, each class having 72 images and the comparison measure employed is the mean precision taken over a window of size equal to the class size. The reduction of the redundancy between the images selected for labeling improves the results, both for MAO with respect to MA and for MPO with respect to MP. Also, the MA and MAO selection criteria compare favorably to the MP and MPO criteria. Similar results on several groundtruth databases confirm the attractiveness of our selection strategy. Moreover, when the target of RF is the batch classification of images, the MAO selection criterium also performs better in terms of classification error and false positives rate compared to MP and MPO criteria (see [10]). The same conclusions remains true when using RF with databases of image regions (see [9]).

6.2.5. Kernels reducing sensitivity to scale for relevance feedback.

Keywords: *image retrieval*, *kernel function*, *relevance feedback*, *scale invariance*.

Participants: Marin Ferecatu, Michel Crucianu, Nozha Boujemaa.

During the study of several groundtruth databases [3], we found that the size of the various classes often covers an important range of different scales in the space of visual features (1 to 7 in our experiments). We expect yet more significant changes in scale to occur from one database to another or from one user-defined class to another class within a large database. We have shown that kernels that depend on a scale parameter does not perform well on generalist databases and we propose the use of specific kernels producing scale invariance with respect to the representation of images in the space of visual descriptors (see [9], [8]).

We compared the behavior of the Gaussian kernel ($K(x_i, x_j) = \exp\left(-\gamma \|x_i - x_j\|^2\right)$), the Laplace kernel ($K(x_i, x_j) = \exp\left(-\gamma \|x_i - x_j\|\right)$), the hyperbolic ($K(x_i, x_j) = 1/\left(\varepsilon + \gamma \|x_i - x_j\|\right)$) and the triangular kernel ($K(x_i, x_j) = -\|x_i - x_j\|$) on several databases with SVM-based relevance feedback. The triangular kernel, $K(x_i, x_j) = -\|x_i - x_j\|$, is a *conditionally* positive definite kernel, but the convergence of SVMs remains guaranteed with this kernel. In [36] the triangular kernel was shown to have a very interesting property: it makes the frontier found by SVMs invariant to the scale of the data.

We considered the MAO selection criterion (see [8]) and we use as precision measure the percentage of positive examples included in the first N images returned by the RF machine at each feedback round (N is the class size). In Fig.13(b) we present the mean precision computed over the first 30 feedback iterations as a function of the scale parameter of the kernels (γ).

Since the classes present in a database often have significantly different scales, any value for the scale parameter of the Gaussian kernel will be àinadequate for many classes, so the results obtained with this kernel cannot be very good. Comparatively, the use of the Laplace kernel reduces the sensitivity of the SVM to scale: an increase of γ beyond 1 has a strong negative impact on the results, while a reduction of γ does not have significant consequences. Also, the hyperbolic kernel produces scale invariance for a large spectrum of values of the parameter γ.

In real applications, the scales of the user-defined classes cannot be known *a priori* and the scale parameter of a kernel cannot be adjusted online, so important variations can be expected for the performance of RF-based retrieval. The scale-invariance obtained by the use of the triangular kernel or the hyperbolic kernel becomes then a highly desirable feature and makes this kernel a very good alternative.

6.3. Clustering methods in complex feature spaces

6.3.1. Semi-supervised clustering with pairwise-constraints

Keywords: *Clustering*, *Pairwise-constraints*, *Semi-supervised Clustering*.

Participants: Nizar Grira, Michel Crucianu, Nozha Boujemaa.

Effective access to the content of an image collection requires a meaningful categorization of the images. Such a categorization can rely on clustering methods working on image features, but should greatly benefit from the semantic information the user can provide, related to the visual content.

Existing unsupervised clustering algorithms do not incorporate available semantic information and the resulting categories often do not reflect user expectations unless further feedback [15]. Consequently, *semi-supervised clustering*—letting "knowledge" provide a limited form of supervision—has recently become a topic of significant interest. More specifically, to help unsupervised clustering, a small amount of knowledge concerning either class labels for some items or pairwise constraints between data items can be used; the constraints specify whether two data items should be in the same cluster or not. However, the few existing semi-supervised clustering algorithms, such as Pairwise Constrained K-means (PCKmeans) [23], rely on parameters that are difficult to choose (such as the number of clusters) and require a high number of constraints to reach good results. The new semi-supervised clustering algorithm we put forward in the following, Pairwise Constrained Competitive Agglomeration (PCCA) [13], provides solutions to these problems.

Our method is based on a minimization of the fuzzy objective function (1) which combines feature-based similarity between data points and knowledge of the pairwise constraints. Let $\mathcal{M}$ be the set of must-link pairs such that $(x_i, x_j) \in \mathcal{M}$ implies x_i and x_j should be assigned to the same cluster, and $\mathcal{C}$ be the set of cannot-link pairs such that $(x_i, x_j) \in \mathcal{C}$ implies x_i and x_j should be assigned to the different cluster. Let $X = \{x_i | \ i = 1, \cdots, N\}$ be the set of N vectors and $V = \{\mu_k | \ k = 1, \cdots, C\}$ the set of prototypes of the C clusters.

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

With the constraint $\sum_{k=1}^{C} u_{ik} = 1, for\ i \ \in \{1, .., N\}$. In (1), $d^2(x_i, \mu_k)$ represents the distance from a feature point x_i to a cluster prototype μ_k and u_{ik} is the membership of x_i to a cluster k.

The first term in (1) is the sum of squared distances to the prototypes weighted by constrained memberships (Fuzzy C-Means objective function). This term reinforces the compactness of the clusters. The second term is composed of

- The cost of violating the pairwise *Must-link* constraints. The penalty corresponding to the presence of two such points in different clusters is weighted by their membership values.
- The cost of violating the pairwise *Cannot-link* constraints. The penalty corresponding to the presence of two such points in a same cluster is weighted by the membership values.

This term is weighted by α, which is a way to specify the relative importance of the supervision.
The third component is the sum of the squares of the cardinalities of the clusters (Competitive Agglomeration) and controls the number of clusters.

Figure 14. A sample of the 4 classes of the image database

Fig.15 presents the dependence between the percentage of well-categorized data points and the number of pairwise constraints considered for a ground-truth image dataset (a sample of which is shwon in Fig.15).

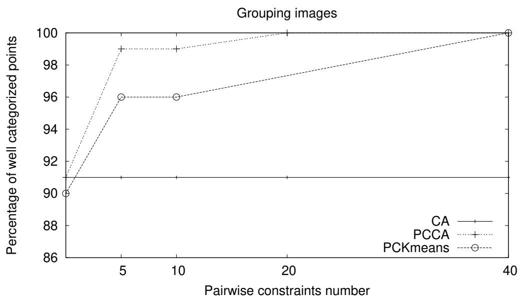


Figure 15. Results on the ground-truth image database

The graph for the basic CA [37] algorithm (ignoring the constraints) is only given as a reference. We can first notice that, by providing simple semantic information in the form of pairwise constraints, the user can significantly improve the quality of the categories obtained. The number of pairwise constraints required for reaching such an improvement is relatively low with respect to the number of items in the dataset. Also, with a similar number of constraints, PCCA performs significantly better than PCKmeans by making a better use of the available constraints.

6.3.2. Consistency of entropy regularization in fuzzy clustering

Keywords: *Fuzzy C-Means*, *Fuzzy Clustering*, *Image Retrieval*, *Image Segmentation*, *Regularization*.

Participants: Hichem Sahbi, Nozha Boujemaa.

One of the main issues in the existing clustering methods remains setting the appropriate number of classes for a given problem. The well-known fuzzy C-mean (FCM) algorithm [26] has proven to perform well when the application allows us to know a priori the number of clusters or when the user sets it manually. Of course, the estimation of this number is application-dependent, for instance in image segmentation it can be set a priori to the number of targeted regions. Unfortunately, for some applications such as database categorization, it is not always possible to predict automatically and even manually the appropriate number of classes.

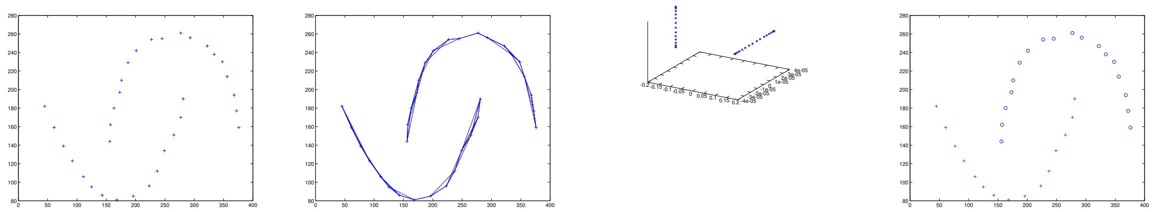


Figure 16. From left to right (1) original non-linearly separable data (2) the underlying adjacency graph. The arity M is set to 3. (3) The underlying isomap embedding makes the two classes linearly separable. (4) the result of the clustering process. The two clusters are shown with different colors.

Several methods exist in the literature in order to automatically find the number of classes for clustering; among them competitive agglomeration (CA) [38]. Solving such a problem implies finding the membership degrees of each training example to the different clusters and assigning the former to the cluster which maximizes its membership. Nevertheless, the constraints on the bounds and the consistency of the membership degrees are difficult to carry. In this work [14], we introduce a new formulation of the regularized fuzzy C-means (FCM) algorithm which allows us to find automatically the actual number of clusters. The approach is based on the minimization of an objective function which mixes, via a particular parameter, a classical FCM term and an entropy regularizer. The method uses a new exponential form of the fuzzy memberships which ensures the consistency of their bounds and makes it possible to interpret the mixing parameter as the variance (or scale) of the clusters. For some applications, it turns out that setting the cluster variance is more intuitive and easier than finding the number of classes mainly for large datasets living in high dimensional spaces.

The metric used for clustering makes it possible to handle manifolds with particular shapes, for instance, the Mahalanobis distance is suitable to carry out Gaussian distributed data. In the second part of this contribution, we adapt the metric, used for clustering, to the topology of the clusters. We use isomap [47], which allows us to embed a training set from an *input space* into an *embedding space* where the underlying classes become linearly separable. The whole objective is to make classes linearly separable and to increase the ratio between the inter and the intra class variances (scales). Therefore arbitrary and non-linearly separable classes will map into Gaussian separable classes (cf. figure 16), so setting the optimal scale parameter becomes easier (cf. figure 17).

6.4. Object Recognition

6.4.1. Alternative Kernels for Image Recognition

Keywords: *SVM*, *pattern recognition*, *positive definite kernel*, *statistical learning*, *visual similarity*.

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

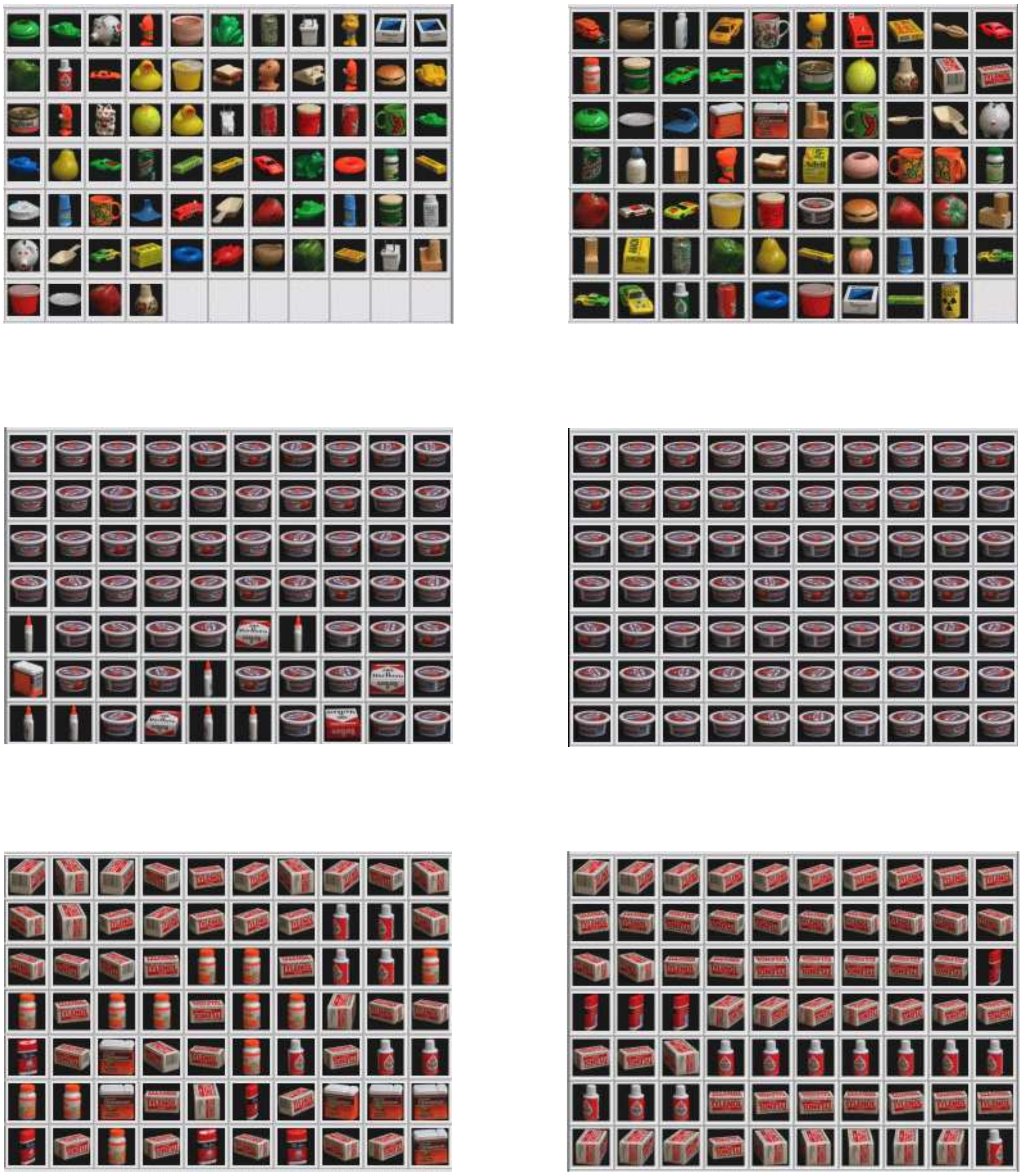


Figure 17. Top: cluster prototypes resulting from our clustering method on the Columbia set and using (on the left) laplacian color space and (on the right) isomap. Middle and bottom: images from two different clusters after the application of our algorithm where the scale σ is set to 0.075 *(resp.* $\sigma = 0.0055$ *using isomap). Images in the left-hand side are related to the laplacian color space while the others are related to isomap.*

Support Vector Machines have been proved to be a successful approach for image recognition. One of the main advantage of SVM compared to other learning algorithms is the uniqueness property of its solution. To insure such property, the used kernel must be positive definite or more generally conditionally positive definite.

Following our last year investigations on the Power kernel $K(x,y) = -\|x_y\|^\beta$, with $0 < \beta < 2$, and the Generalized T-Student kernel $K(x,y) = \frac{1}{1+\|x_y\|^\beta}$, we introduce the new kernel $K(x,y) = -log(1 + \|x_y\|^\beta)$, we named Log kernel. This kernel is interesting since, it has intermediate shape between Power and Generalized T-Student kernels. We prove that Log kernel is conditionally positive definite, and thus suitable for SVM. In our comparison experiments on Corel Database, the log kernel is better performing than other kernels such as polynomial, RBF, and power kernels.

Usually, the kernel hyper-parameters such as the scale parameter are tuned using cross-validation procedure. In practice is is not realistic to assume that data are distributed uniformly on the feature space. This explains why kernels leading to invariant SVM clustering are of main importance, such as the Power kernel. We thus propose a new kernel $K(\mathbf{x},\mathbf{y}) = \sum_{i=1}^{m} min(|x_i|^\alpha, |y_i|^\beta)$, we named Generalized Histogram Intersection kernel which we prove that it is positive definite. This kernel leads to invariant SVM clustering and to good performances in our experiments.

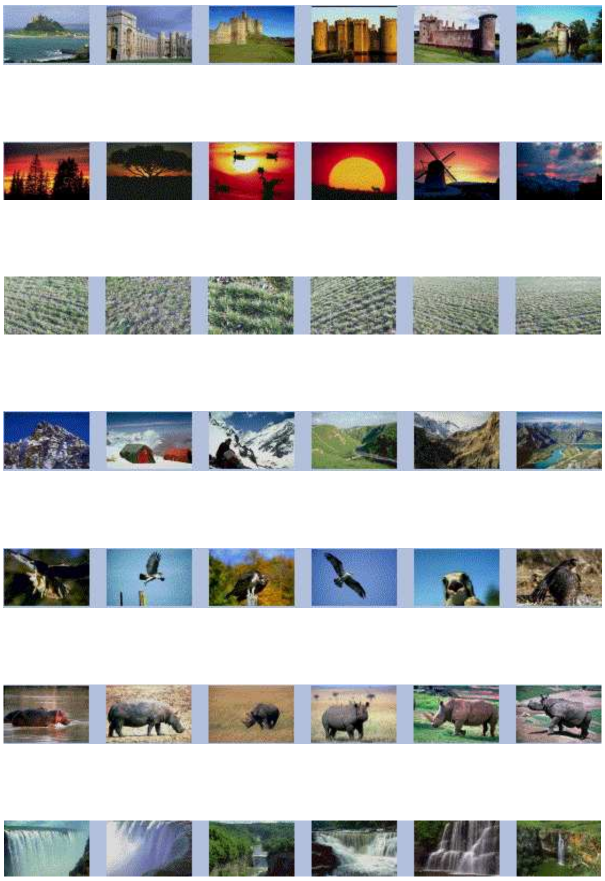


Figure 18. The 7 classes of 100 images used for experiments (6 classes from Corel database).

Table 1. Error rates for different classes and kernels with optimal hyper-parameters. Hyper-parameters (d for polynomial kernel, β for Generalized Histogram Intersection (GHI) kernel, and σ for RBF kernel) are given between parenthesis. The Histogram Intersection (HI) which have been introduced in is a particular case of the GHI kernel for β=1

Kernels	Poly	GHI	RBF	HI
	(d)	(β)	(σ)	($\beta = 1$)
Castles	8.94	**8.07**	13.55	9.05
	(5)	(0.58)	(0.04)	-
Sunset	3.27	**2.91**	9.46	3.56
	(5)	(0.34)	(0.069)	-
Grass	0.31	**0.14**	0.49	0.14
	(5)	(0.58)	(0.014)	-
Mountains	10.92	**9.88**	14.28	11.28
	(5)	(0.34)	(0.001)	-
Birds	8.14	**6.44**	14.19	8.43
	(3)	(0.34)	(0.34)	-
Rhinos	7.58	**4.99**	13.34	7.06
	(7)	(0.20)	(0.04)	-
Waterfalls	9.16	**8.93**	11.86	10.04
	(7)	(0.34)	(0.04)	-
Mean	6.90	**5.91**	11.02	7.08

6.4.2. GCS Kernel for SVM-Based Object Recognition

Keywords: *kernel methods*, *matching*, *pattern recognition*, *statistical learning*, *visual similarity*.

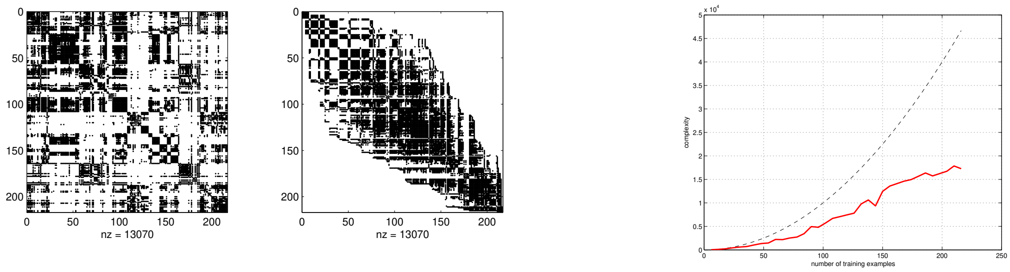
Participants: Sabri Boughorbel, Jean Philippe Tarel, François Fleuret, Nozha Boujemaa.

We introduce a new compactly supported kernel for object recognition. This kernel which we called Geometric Compactly Supported (GCS), is useful for both global and local feature image representations. The compactness property of the GCS kernel leads, in the case of global features, to a sparse Gram matrix which enhances computation efficiency by using sparse linear algebra algorithms. In the case of a set of local features, the compactness of the GCS kernel mimics matching algorithms while remaining computationally efficient. The construction of the GCS kernel is based on a geometric approach where the intersection volume of two n-dimensional balls is used [6].

The GCS kernel $\Phi_n(x,y)$ can be computed recursively. For a dimension n, we define $\phi_{n,k}(x,y)$ as the value of $\Phi_n(x,y)$ at kth iteration. So we have $\Phi_n(x,y) = \phi_{n,n}(x,y)$. We prove for $\|x-y\| < 2r$ that

$$\begin{aligned} \phi_{n,k}(x,y) = \tfrac{k-1}{k}\phi_{n,k-2}(x,y) - \tfrac{1}{k}\tfrac{\|x-y\|}{2r}\left(1-\left(\tfrac{\|x-y\|}{2r}\right)^2\right)^{\frac{k-1}{2}} \\ \phi_{n,2}(x,y) = \arccos(\tfrac{\|x-y\|}{2r}) - \tfrac{\|x-y\|}{2r}\sqrt{1-\left(\tfrac{\|x-y\|}{2r}\right)^2} \\ \phi_{n,1}(x,y) = 1 - \tfrac{\|x-y\|}{2r} \end{aligned} \tag{2}$$

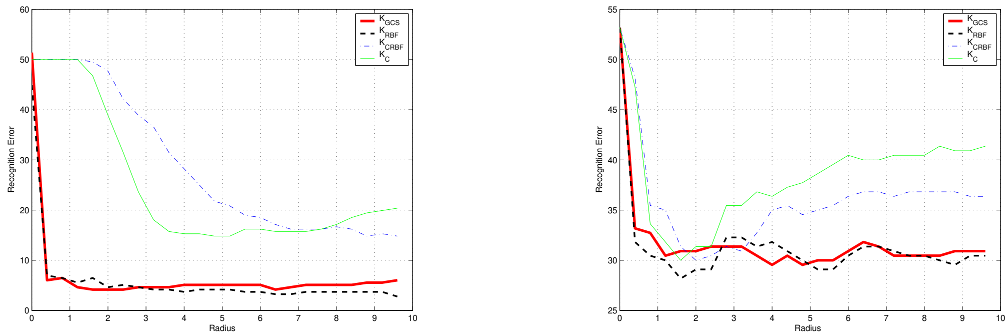
For $\|x-y\| \geq 2r$, $\phi_{n,k}(x,y) = 0$. Φ_1 on $\mathbb{R}$ is the triangular kernel, Φ_2 on $\mathbb{R}^2$ is the circular kernel and Φ_3 on $\mathbb{R}^3$ is the spherical kernel [40][39][41]. Φ_n on $\mathbb{R}^n$ is the GCS kernel, it is positive definite kernel. It can be seen as the generalization of the spherical kernel to high dimensions.



(a) (b) (c)

Figure 19. (a)-(b) Symmetric reverse Cuthill-McKee permutation of sparse Gram matrix. (c) Complexity reduction of SVM dual problem using banded matrix representation.

$$\begin{aligned} K_{RBF}(x,y) &= e^{-\frac{\|x-y\|^2}{r^2}} \\ K_C(x,y) &= \max\left\{\left(1-\tfrac{\|x-y\|}{r}\right)^{\nu},0\right\} \\ K_{CRBF}(x,y) &= K_C(x,y).K_{RBF}(x,y) \end{aligned} \tag{3}$$



(a) (b)

Figure 20. Comparison of kernels K_{GCS}, K_{RBF}, K_{CRBF} and K_C for two different objects.

6.5. IKONA/MAESTRO Software

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

Participants: Nicolas Hervé, Josselin Chièze, Marin Ferecatu, Jean-Paul Chièze.

This year, the Ikona/Maestro software is being re-engineered to improve its modularity. Most of the changes are done on the server part. They relate to the internal architecture. In this new version, the interfaces of all the main components will be unified. This changes will make easier the integration of new descriptors, new query paradigms and new functionnalities. Among others, one can think of :

- the possibility of loading / unloading dynamically the descriptors and the databases
- the visual summary and visual navigation in the clusters of a database
- the combination of different kinds of queries (eg. region-based with point-based query)
- the use of multidimensional indexes (the indexes type may be choosen according to the descriptors)
- the queries involving topological relationships between parts of a picture
- let the user choose the weights of the descriptors when more than one are involved in a query
- the creation of a query language

To give users access over the Web to our IKONA search engine, we developed Ikona-CGI, a user interface employing web-based technologies. Once installed on a web server, the software components associated to this interface allow any user,with a simple web browser, to easily search for images in big image databases. With this interface, queries consist in either entire images, keywords, image regions or points of interest, and the query modes available are query by example, search by composition of categories from a visual thesaurus and refinement with relevance feedback. The interface also provides access to visual summaries of the databases. A tool associated to this interface was also developed: freeman2img is a program that automatically transforms a database of segmented images so that it can be used with Ikona-CGI for region-based search.

7. Other Grants and Activities

7.1. National Initiatives

http://www-rocq.inria.fr/imedia/projects.html

7.1.1. Industrial contract with Sagem [2003-2004]

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

7.1.2. 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 monitiring of INA video collections.

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

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

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

This project concerns the conception and developpement 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.

7.2. European Initiatives

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

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

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

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

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

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

7.3. 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.
http://www-rocq.inria.fr/imedia/vimining/index.html
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.

8. Dissemination

8.1. Leadership with scientific community

Nozha Boujemaa :

- Member of Task Force in Biometrics and Multimedia, EETC (Emergent Technologies Technical Committee), IEEE CIS (Computational Intelligence Society)
- Member of Scientific Council for "Biometry" exhibition at "Cité des Sciences et de l’Industrie".
- Invited Plenary talk for "Université de Tous Les Savoirs". The talk was broadcasted on France Culture (FM radio) and France5 (TV) and is also available on internet.
- Scientific Editor of a Special Issue on Visual Information Retrieval within French Journal “Techniques et Sciences informatiques”, Vol.22, Hermès, 156p [2]. (http://www.lavoisier.fr/fr/livres/index.asp?texte=274620839
- Special session co-organizer with Pr Shin’ichi Satoh on the topic "Visual Content Mining in Multimedia Documents" at PCM’04 Conference (http://www.pcm2004.org/)
- Conference Programme committee member of
  - SPIE Conference on Storage and Retrieval Methods and Applications for Multimedia 04, IEEE Fuzzy Systems04, CBMI 04, ICME 04, ICPR 04, PCM 04, CVDB 04, MDDE 04 in conjunction with CVPR 04.
- Journals reviewer:
  - Multimedia Tools and Applications, IEEE Trans. on Multimedia, IEEE Trans. on PAMI, IEEE Trans. Image Processing, IEE Proc. Vision, Image & Signal Processing, TS, I3.
- Scientific coordinator of “Single Modality” WP5 in Muscle NoE (Network of Exellence FP6), Deputy Scientific coordinator of the Muscle NoE (http://www-rocq.inria.fr/imedia/Muscle/WP5/) and (http://www.muscle-noe.org/)
- Scientific expert for:
  - National Assembly Commission for the law project "LEN" (Loi sur l’Economie Numérique)
  - the French research program ACI (Actions Concertées Incitatives) “Masses of data”
- In Charge of International Relations of INRIA Rocquencourt Research Unit and member of “Bureau du Comité des Projets”
- Member of recruitment commission of INRIA Researcher (CR2)
- Member of researcher mobility commission (détachement/délégation)

Jean-Philippe Tarel :

- Journal Reviewer: IEEE Trans. on PAMI, Machine Graphics & Vision Pattern Recognition Letters.

- Jury member:
  - PhD Kongbin Kang, Brown University, april 27, 2004
  - PhD Sio-Song Ieng, Paris XI University, novembre 23, 2004
  - PhD Alban Foulonneau, Strasbourg University, december 15, 2004.

Anne Verroust-Blondet :

- 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)";
- Jury member:
  - PhD Jean Combaz (INRIA - Grenoble), PhD Aurelien Barbier (LIRIS - Lyon).
- Journals reviewer:
  - Computer Aided Geometric Design, IEEE Transactions on Systems, Man, and Cybernetics Part B).

Valérie Gouet:

- Conference Program committee member of the International Workshop on Computer Vision meets Databases (CVDB 2004), in cooperation with ACM SIGMOD, June 13, 2004, Maison de la Chimie, Paris, France.
- Jury member of the engineer diploma of O. Guilloux, CNAM, October 2004.
- Reviewer for PAA, ICME, ICPR, TSI, MTA, EURASIP JASP.

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").

8.2. Teaching

Nozha Boujemaa:

- course on "Image-based retrieval in Multimedia Databases" at Sup'Com (10h - Tunis).
- in charge of course at Telecom-Paris "Image indexind and retrieval"
- Coordinator of a fall school on "Analyse d'images: de la théorie aux applications". About 250 students have attended this school (http://www-rocq.inria.fr/imedia/inisat/manifestations.html)

Jean-Philippe Tarel

- - 12h Seminars in Object Detection, Master STIC, CNAM, february and december 2004.

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).

Valérie Gouet:

- 192 HTD in the Computer Science Department of CNAM ;
- In charge of the course "Computer Vision" of the Master in Computer Science of CNAM (6 ECTS) ;
- Course "Multimedia databases" in last year of Sup'Com-Tunis (15 HTD).

9. Bibliography

Articles in referred journals and book chapters

[1] N. BOUJEMAA, J. FAUQUEUR, V. GOUET. *What's beyond query by example?*, L. SHAPIRO, H. KRIEGEL, R. VELTKAMP (editors)., Springer Verlag, 2004.

[2] N. BOUJEMAA, F. JURIE. *Recherche d'information par le contenu visuel*, vol. 22, 2004.

[3] N. BOUJEMAAA, M. FERECATU. *Evaluation des systèmes de recherche par le contenu visuel : pertinence et critères*, S. CHAUDIRON (editor)., Hermès, 2004.

[4] J. FAUQUEUR, N. BOUJEMAA. *Mental Image Search by Boolean Composition of Region Categories*, 2004.

[5] J. FAUQUEUR, N. BOUJEMAA. *Region-based image retrieval: fast coarse segmentation and fine color description.*, in "J. Vis. Lang. Comput.", vol. 15, n° 1, 2004, p. 69-95.

Publications in Conferences and Workshops

[6] S. BOUGHORBEL, J.-P. TAREL, F. FLEURET. *Non-Mercer Kernels for SVM Object Recognition*, in "Proceedings of British Machine Vision Conference (BMVC'04), London, England", http://www-rocq.inria.fr/ tarel/bmvc04.html, 2004, p. 137 - 146.

[7] N. BOUJEMAA, F. FLEURET, V. GOUET, H. SAHBI. *Visual content extraction for automatic semantic annotation of video news*, in "IS&T/SPIE Conference on Storaghe and Retrieval Methods and Applications for Multimedia, part of Electronic Imaging Symposium, San José, CA, USA", January 2004.

[8] 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.

[9] M. FERECATU, M. CRUCIANU, N. BOUJEMAA. *Sample selection strategies for relevance feedback in region-based image retrieval*, in "Proceedings of the Pacific-Rim Conference on Multimedia 2004", December 2004, p. 497 – 504.

[10] M. FERECATU, M. CRUCIANU, N. BOUJEMAA. *Tuning SVM-based relevance feedback for the interactive classification of images*, in "Proceedings of the European Workshop on the Integration of Knowledge, Semantics and Digital Media Technology", November 2004.

[11] V. GOUET. *Content-based image indexing and retrieval with local descriptors: a survey*, in "A state of the art on Image and Video Processing, European Network of Excellence MUSCLE (FP 6), report WP5", 2004, p. 6-12.

[12] V. GOUET. *Index structures for image search by content: a survey*, in "A state of the art on Computation Intensive Methods, European Network of Excellence MUSCLE (FP 6), report WP7", 2004, p. 68-70.

[13] N. GRIRA, M. CRUCIANU, N. BOUJEMAA. *Fuzzy clustering with pairwise constraints for knowledge-driven image categorization*, in "European Workshop on the Integration of Knowledge, Semantics and Digital Media Technology, London, U.K.", The Royal Statistical Society, November 2004.

[14] H. SAHBI, N. BOUJEMAA. *Fuzzy Clustering: Consistency of Entropy Regularization*, in "International Conference on Computational Intelligence (Special Session on Fuzzy Clustering)", 2004.

[15] B. L. SAUX, N. BOUJEMAA. *Image database clustering with SVM-based class personalization*, in "IS&T/SPIE Conference on Storaghe and Retrieval Methods and Applications for Multimedia, part of Electronic Imaging Symposium, San José, CA, USA", jan 2004.

[16] J. THIÈVRE, M. VIAUD, A. VERROUST-BLONDET. *Using Euler Diagrams in Traditional Library Environment*, in "Euler Diagrams 2004, Brighton, U.K.", September 2004.

[17] A. VERROUST, M. VIAUD. *Ensuring the drawability of Extended Euler Diagrams for up to 8 sets*, in "Diagrammatic Representation and Inference, Diagrams 2004, Cambridge, U.K.", Lecture Notes in Artificial Intelligence, n° 2980, mar 2004, p. 128-141.

Internal Reports

[18] H. SAHBI, F. FLEURET. *Kernel Methods and Scale Invariance Using the Triangular Kernel*, Technical report, INRIA-Research Report Number 5143, March 2004, http://www.inria.fr/rrrt/rr-5143.html.

[19] H. SAHBI. *Affine Invariant Shape Description Using the Triangular Kernel*, Technical report, INRIA-Research Report Number 5308, September 2004, http://www.inria.fr/rrrt/rr-5308.html.

Miscellaneous

[20] R. BOUCHIHA. *Indexation et recherche d'images par descripteurs locaux*, 2004.

[21] M. CHAOUCH. *Etude et comparaison de descripteurs de formes pour l'indexation 3D*, june 2004, Projet de fin d'études, Ecole Polytechnique de Tunisie.

Bibliography in notes

[22] A. BARLA, F. ODONE, A. VERRI. *Histogram intersection kernel for image classification*, in "Proceedings of the International Conference on Image Processing, Barcelona, Spain", 2003, p. 513-516.

[23] 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.

[24] S. BELONGIE, J. MALIK, J. PUZICHA. *Shape Matching and Object Recognition Using Shape Contexts*, in "IEEE Trans. on Pattern Analysis and Machine Intelligence", vol. 24, n° 4, 2002, p. 509–522.

[25] S. BERRANI. *Recherche approximative de plus proches voisins avec contrôle probabiliste de la précision ; application la recherche d'images par le contenu*, Ph. D. Thesis, Université de Rennes 1, France, feb 2004.

[26] J. BEZDEK. *Pattern Recognition with Fuzzy Objective Function Algorithms. New York*, Kluwer Academic Norwell, 1981.

[27] 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, http://citeseer.ist.psu.edu/christian00pruning.html.

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

[29] 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.

[30] 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.

[31] BRAUNMULLER, B., ESTER, K. M., H. P., SANDER, J.. *Efficiently supporting multiple similarity queries for mining in metric databases*, in "16th Int. Conf. on Data Engineering, San Diego, CA", 2000, p. 256-270.

[32] D. CREMERS, T. KOHLBERGER, C. SCHN?RR. *Nonlinear Shape Statistics via Kernel Spaces*, in "German National Conference on Pattern Recognition (DAGM)", 2001, p. 269–2763.

[33] C. FALOUSTOS, K.-I. LIN. *Fast Map: A Fast Algorithm for Indexing, Data-Mining and visualization of Traditional and Multimedia Datasets*, Technical report, ISR-Technical Report Number 94-80, 1995.

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

[35] F. FLEURET, H. SAHBI. *Scale-Invariance of Support Vector Machines Based on the Triangular Kernel*, 2003.

[36] F. FLEURET, H. SAHBI. *Scale-invariance of support vector machines based on the triangular kernel*, in "3rd International Workshop on Statistical and Computational Theories of Vision", October 2003.

[37] H. FRIGUI, R. KRISHNAPURAM. *Clustering by competitive agglomeration*, in "Pattern Recognition", vol. 30, n° 7, 1997, p. 1109–1119.

[38] H. FRIGUI, R. KRISHNAPURAM. *A Robust Competitive Clustering Algorithm With Applications in Computer Vision*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", vol. 21, n° 5, 1999, p. 450-465.

[39] M. GENTON. *Classes of kernels for Machine Learning: a Statistics perspective*, in "Journal of Machine Learning Research", vol. 2, 2001, p. 299-312.

[40] T. GNEITING. *Compactly Supported Correlation Functions*, in "Compactly Supported Correlation Functions", vol. 83, 2002, p. 493-508.

[41] B. HAMERS, J. A. SUYKENS, B. DE MOOR. *Compactly Supported RBF Kernels for Sparsifying the Gram Matrix in LS-SVM Regression Models*, in "Proceedings of the International Conference on Artificial Neural Networks, ICANN 2002, Madrid Spain", 2002.

[42] S. J., C. D.B., K. D.. *Practical Reliable Bayesian Recognition of 2D and 3D Objects Using Implicit Polynomials and Algebraic Invariants*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", vol. 18, n° 5, 1996, p. 505-519.

[43] F. MOKHTARIAN, S. ABBASI, J. KITTLER. *Efficient and robust retrieval by shape content through curvature scale space*, in "First International Workshop on Image Databases and Multi-Media Search", 1996, p. 35–42.

[44] A. NG, M. JORDAN, Y. WEISS. *On spectral clustering: Analysis and an algorithm*, 2001, http://citeseer.ist.psu.edu/ng01spectral.html.

[45] A. ROSENFELD. *Axial Representation of Shape*, in "CVGIP", vol. 33, n° 2, 1986, p. 156–173.

[46] P. SHILANE, P. MIN, M. KAZHDAN, T. FUNKHOUSER. *The Princeton Shape Benchmark*, in "Shape Modeling and Applications Conference, SMI'2004, Genova, Italy", IEEE, June 2004, p. 167-178.

[47] J. TANENBAUM, V. DE SILVA, J. LANGFORD. *A global geometric framework for non-linear dimensionality reduction*, in "Science", vol. 290, nº 5500, 2000, p. 2319–2323.

[48] S. TONG, E. CHANG. *Support vector machine active learning for image retrieval*, in "Proceedings of the 9th ACM international conference on Multimedia", ACM Press, 2001, p. 107–118, http://doi.acm.org/10.1145/500141.500159.

[49] S. TONG, D. KOLLER. *Support vector machine active learning with applications to text classification*, in "Proceedings of ICML-00, 17th International Conference on Machine Learning", Morgan Kaufmann, 2000, p. 999–1006, http://citeseer.nj.nec.com/tong00support.html.

[50] T. ZAHARIA, F. PRÊTEUX. *Hough transform-based 3D mesh retrieval*, in "SPIE Conference 4476 on Vision Geometry X, San Diego, CA", August 2001, p. 175-185.