INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Project-Team Imedia

Images and Multimedia: Indexing, Retrieval and Navigation

Paris - Rocquencourt

THEME COG

Activity Report

2007

Table of contents

1. Team

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

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

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

Research scientists (external)
Olivier Buisson [Research Scientist at Institut National de l'Audiovisuel]
Michel Crucianu [Professor at CNAM, HdR]
Valérie Gouet-Brunet [Assistant Professor at CNAM]
Marie-Luce Viaud [Research Scientist at Institut National de l'Audiovisuel]
Itheri Yahiaoui [Assistant Professor at Reims University]

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

Post-doctoral Fellow
Jaume Amores [Postdoctoral fellow]
Marin Ferecatu [Expert engineer Inria]
Joost Geurts [Postdoctoral fellow]
Hichem Houissa [Expert engineer Inria since June 1st 2007]
Laurent Joyeux [Expert engineer Inria since February 1st 2007]
Julien Law To [Expert engineer Inria since May 15 2007]
Hong Tang [Postdoctoral fellow (shared with ENST Paris)]

Technical staff
Mehdi Bouabta [Junior Technical Staff Inria]
Mathieu Coutaud [Junior Technical Staff Inria]

Ph. D. Student
Olfa Besbes-Abdelhak [Joint tutorship with Sup'Com, national grant since September 1st 2004]
Mohamed Chaouch [INRIA grant, Télécom Paris since October 1st 2005]
Nicolas Hervé [INRIA grant, Paris-Sud University since December 1st 2005]
Hichem Houissa [INRIA grant, Paris-Sud University until May 31st 2007]
Julien Law-To [CIFRE grant with INA until May 1st 2007]
Ahmed Rebaï [INRIA grant, Paris-Sud University since September 1st 2007]

Interns
Marwen Ben Ali [Paris V University, from April 1st until July 31st 2007]
Mehdi Ellouze [ISIET, Sfax, since November 1st 2007]
Amel Hamzaoui [ISI, Tunis, from April 1st until August 15 2007]
Saloua Litayem [ISI, Tunis, from April 1st until August 31st 2007]
Ahmed Rebaï [Sup'Com, Tunis, till August 31 2007]
Philippe Tony [ENSIEE, Evry, from June 1st till August 31 2007]

2. Overall Objectives

2.1. Introduction

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 scalability of such interactive systems in terms of indexing and response times. Of course, what value these times should have and how critical they are depend heavily on the domain (specific or generic) and on the cost of the errors.

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

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

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

2.2. Highlights of the year

- IMEDIA won all tasks in image retrieval during Technovision - ImagEVAL Benchmarking Campaign (DGA initiative supervised by CEA).

- Lunching of VITALAS integrated project which was first ranked during the EC evaluation of call6. Imedia have strong involvement and the scientific leadership of this project.

- Lunching of CHORUS coordination action dedicated to "multimedia search engines" which become a strategic objective for the European commission. The objective is to coordinate the effort of all existing European and national initiatives running under this topic. IMEDIA has the scientific leadership of this European initiative. The outcome are prospective and recommendations for the future calls objectives (see the article “Search For Tomorrow” in “The International Herald Tribune” of May 9 2007, http://www.iht.com/articles/2007/05/09/technology/ptend10.php).

3. Scientific Foundations

3.1. Introduction

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

3.2. Modelling, construction and structuring of the feature space

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

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

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

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

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

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

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

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

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

3.3. Pattern recognition and statistical learning

Statistical learning and classification methods are of central interest for content-based image retrieval [36] [39].

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

3.3.1. Statistical learning and object detection

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

Participants: Donald Geman, Nozha Boujemaa, Nicolas Hervé, Jaume Amores, Alexis Joly, Ahmed Rebai.

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 generalisation rules from training examples. The generalisation error of these techniques can be controlled, given the complexity of the models considered and the size of the training set.

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

3.3.2. Clustering methods

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

Participants: Nozha Boujemaa, Michel Crucianu, Itheri Yahiaoui, Hichem Houissa, Nicolas Hervé.

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

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

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

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

3.4. Interactive search and personalisation

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

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

We are studying here the approaches that allow for a reduction of the "semantic gap" There are several ways to deal with the semantic gap. One prior work is to optimise 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 minimise the numerical gap, we have to develop efficient images signatures. The weakness of visual retrieval results, due to the numerical gap, is often confusingly attributed to the semantic gap. We think that providing richer user-system interaction allows user expression on his preferences and focus on his semantic visual-content target.

Rich user expression comes in a variety of forms:

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

3.5. Cross-media indexing and retrieval

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

Participants: Nicolas Hervé, Hong Tang, Nozha Boujemaa.

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

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

4. Application Domains

4.1. Application Domains

- **Security applications** Examples: Identify faces or digital fingerprints (biometry). Biometry is an interesting specific application for both a theoretical and an application (recognition, supervision, ...) point of view. Two PhDs were defended on themes related to biometry. Our team also worked with a database of images of stolen objects and a database of images after a search (for fighting pedophilia). We are currently collaborating with the Ministry of the Interior.
- **Multimedia** Examples: Look for a specific shot in a movie, documentary or TV news, present a video summary. Help archivists to annotate the contents. Detect copies of a given material in a TV stream or on the web. Our team has a collaboration with INA (french TV archives) and IRT (german broadcasters) in the context of a European project. Text annotation is still very important in such applications, so that cross-media access is crucial.
- **Scientific applications** Examples: environmental images databases: fauna and flora; satellite images databases: ground typology; medical images databases: find images of a pathological character for educational or investigation purposes. We have an ongoing project on multimedia access to biodiversity collections.
- **Culture, art and education** Examples: encyclopedic research, query by example of paintings or drawings, query by a detail of an image. IMEDIA has been contacted by the French ministry of culture and by museums for their image archives.
  Finding a specific texture for the textile industry, illustrating an advertisement by an appropriate picture. IMEDIA is working with a picture library that provides images for advertising agencies.
- **Telecommunications** Examples: image representation and content-based queries stand as the basis of MPEG-4 and MPEG-7. IMEDIA does not contribute to their normative aspects but is interested in the latest results related to the MPEG-7 group. Note that the signatures developed by IMEDIA can be used with this norm.

5. Software

5.1. IKONA/MAESTRO Software

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

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

IKONA is a framework for building Content Based Image Retrieval software prototypes. It has been designed and implemented in our team during the last four years [37]. The current version is fully generic and is highly adaptable to any CBIR scenario thanks to its level of abstraction. As a research environment, IKONA offers support to the researchers in their work by providing stable and tested tools. As an application, it can easily be deployed and used by non-specialist users.

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

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

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

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

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

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

- **query by global example**: The user selects an entire image as visual query.
- **partial queries**: the user is looking for regions in images that are visually similar to a the selected region.
- **relevance feedback on global and partial query**: the user interacts with the system in a feedback loop, by giving positive and negative examples to help the system identify the category of images she/he is interested in [37];
- **mental image search**: Two different methods are investigated. The first is Target Image Search with relevance feed-back model based on mutual information, the second one consist on Logical Query Composition.

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

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

6. New Results

6.1. Construction and organisation of the visual feature space

6.1.1. Video Copy detection

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

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

Motivation and challenge. Growing broadcasting of digital video content on different media brings the search of copies in large video databases to a new critical issue. Digital videos can be found on TV Channels, Web-TV, Video Blogs and the public Video Web servers. The massive capacity of these sources makes the tracing of video content into a very hard problem for video professionals. At the same time, controlling the copyright of the huge number of videos uploaded everyday is a critical challenge for the owner of the popular video web servers. Content Based Copy Detection (CBCD) presents an alternative to the watermarking approach to identify video sequences and to solve this challenge.

Proposed Concept.

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

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

Figure 1. Copy / similarity .

We propose a solution called **ViCopT** for Video CopyTracking that involves estimating and characterising trajectories of complementary points of interest throughout the video sequence. These complementary points of interest are characterised by a local description that is robust to changes of luminance. The trajectory properties allows the local description to be enriched by adding a spatial, dynamic and temporal behaviour of this point. Analysing the trajectories obtained makes it possible to highlight trends of behaviours and then to assign a label of behaviour to a local descriptor. The aim is to provide a *rich*, *compact* and *generic* video content description, which is used in a robust voting function dedicated to copy detection in large video databases. All the details are described in the PhD thesis [8]. An interactive demo of this technology had been presented in industrial exhibit and in conference demo session ([30]

(a) *Special Bardot* 1968 (c)ORTF.

(b) Left: *Les duos de l'impossible* 2005. Right: *Samedi et Compagnie.* 1970 (c)ORTF

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

Comparative Evaluation.

Different state-of-the-art techniques, using various kinds of descriptors and voting functions have been compared within the same framework. Global video descriptors, based on spatial and temporal features; local descriptors based on spatial, temporal as well as spatio-temporal information have been studied an e-team in the European Network of Excellence MUSCLE. Robust voting functions have been adapted to these techniques to enhance their performances and to fairly compare them ([29]). The high distinctiveness of some trajectories and the robust description of points of interest which have a strong visual content allow **ViCopT** to achieve good performances facing the other techniques as shown in figure 3 and table 1.

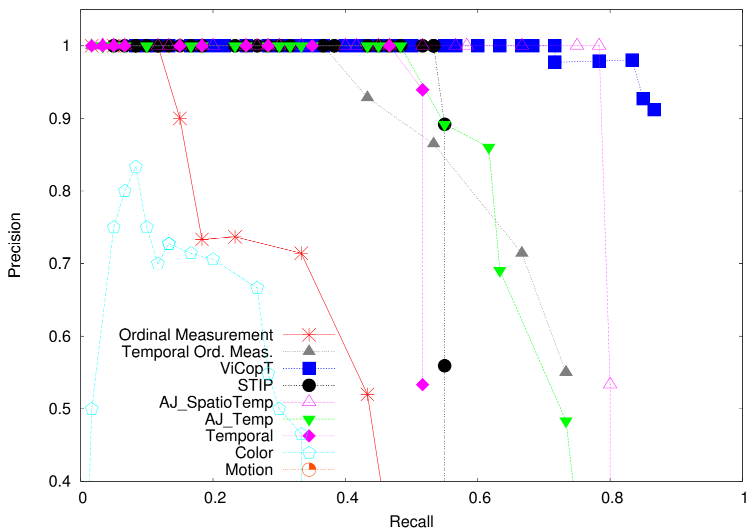

Figure 3. Precision and Recall of the different techniques

Video Linking.

Other works use **ViCopT** to create links between videos in large database ([28]). More than eliminating redundancy, linking videos in a large database is a very motivating challenge to enrich large collections of videos. Links have been created between 5,600 videos downloaded on the Internet and different kind of links have been considered: copy, common segment or similar background.

Table 1. Average precision

Technique	AveP	Technique	AveP
ViCopT	0.86	STIP	0.55
AJ_SpatioTemp	0.79	Temporal	0.51
AJ_Temp	0.68	Ord. Meas.	0.36
Temp. Ord. Meas.	0.65		

6.1.2. Stochastic satellite image segmentation by combining region and edge cues

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

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

Image segmentation and perceptual grouping in terms of identifying coherent segments in images are long standing problems in computer vision. In spite of diverse existing attempts, finding a method that can match human performance in terms of quality and speed has remained challenging because of the complexity of generic images such as satellite images. As salient segments may be separable by any of a variety of cues such as their colour, texture [20] or boundary, we make an attempt to design an efficient algorithm for image segmentation by combining all above cues. Therefore, the objective function is defined in terms of region and boundary properties of segments and according to some Gestalt principles such as similarity, proximity and good continuation. Coherent segments, which have both large similarities within them and large dissimilarities across their boundaries, are extracted. We formulate the problem as partitioning an adjacency graph with the vertices being a mixing of atomic region and atomic contour primitives and edges being their spatial relationships. Our goal is to group atomic regions into regions and atomic contours into boundaries, so that the posterior probability $P(R, B|F)$ of region and boundary partitions (R, B) given observations F is optimised. The posterior probability is directly defined by using a Conditional Random Field (CRF) framework [40]. Therefore, we take advantages of not only capturing local contexts in images by incorporating data-dependent interactions among the labels but also integrating multiple grouping cues by a simple form of a log-linear combination of features. We use a data-dependent discriminative model based on log-likelihood ratios in order to define the CRF potentials. Given our CRF-based model $P(R, B|F, \Theta)$ with respect to an adjacency graph $G = (V, E)$, we carry out inference using a cluster sampling method, the Swendsen-Wang Cut (SWC) algorithm [34], which allows large sampling moves and thus provides fast simulation and optimisation. During the stochastic simulation process, probabilistic grouping operations are influenced by both region and edge cues.

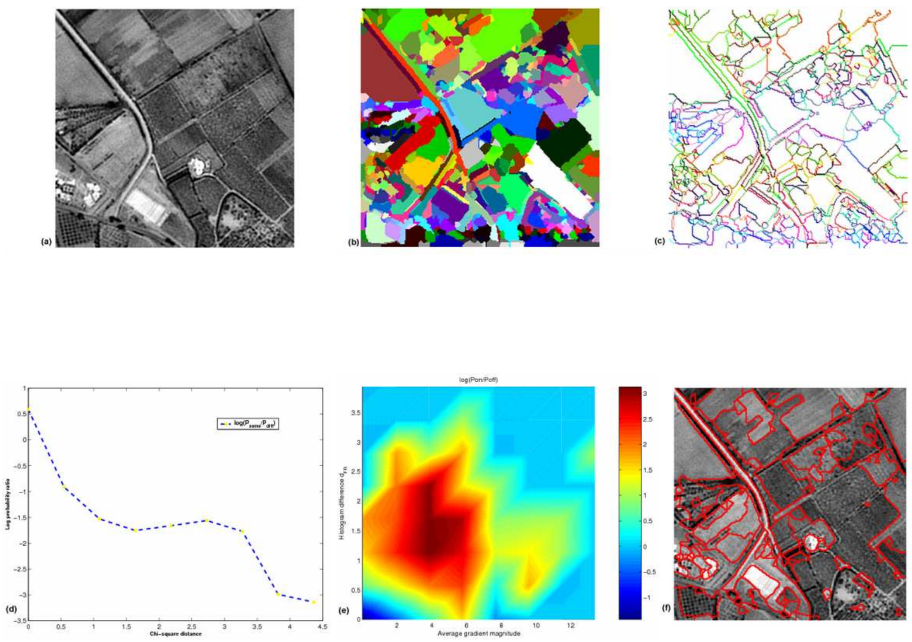

Figure 4. A panchromatic satellite image segmentation result.

In figure 4, we present respectively a panchromatic satellite image, its two maps of atomic regions and contours, the learned region and edge based log-likelihood ratios and the segmentation result.
In future work, we intend to extend the binary classification of atomic contours (B) to the multi-class case so that salient structural segments will be extracted.

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

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

Participants: Mohamed Chaouch, Anne Verroust-Blondet.

This year, a new 2D/3D approach has been developed, based on depth line descriptor (DLA approaches). Each model is represented by a set of depth lines, which are extracted from the depth-buffer images associated to the 3D model. These depth lines are transformed into sequences. The depth sequence information provides a more accurate description of 3D shape boundaries than using other 2D shape descriptors. Retrieval is performed when dynamic programming distance (DPD) or Hamming distance (HD) is used to compare the depth line descriptors. The DPD leads to an accurate matching of sequences even in the presence of local shifting on the shape. Experimentally, we have shown absolute improvement in retrieval performance on the Princeton 3D Shape Benchmark database (cf. Figure 5).

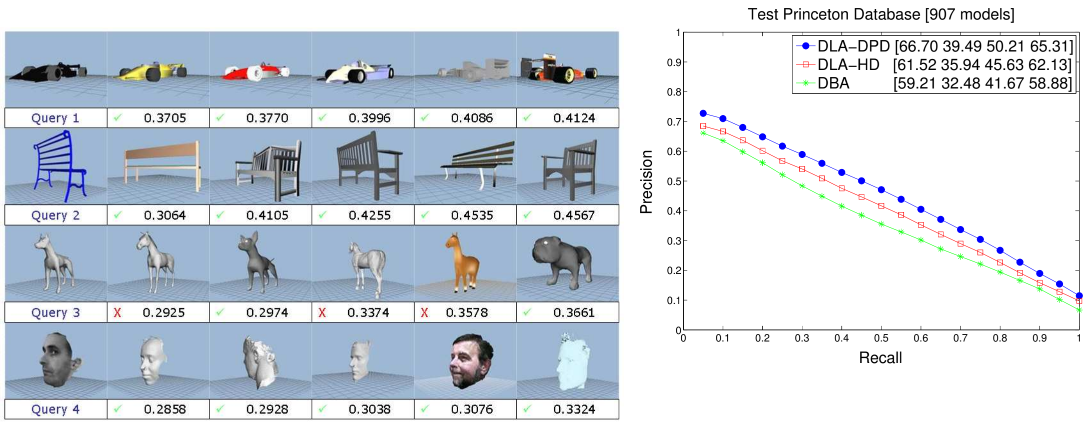

Figure 5. Left: example of queries with the DLA-DPD approach. Right: Precision-recall curves for the global average using the Depth Buffer approach (DBA) and DLA approach using Hamming distance (DLA-HD) or dynamic programming distance (DLA-DPD) to compare the depth line descriptors. The mean NN, FT, ST and DCG values are given in the legends.

This work is described in [21] and it has been improved in [22], where twenty depth images rendered from the vertices of a regular dodecahedron represent a 3D model. It has been supported in part by the DELOS NoE on Digital Libraries (Task 3.8 “Description, Matching and Retrieval by Content of 3D Objects” of WP3 “Audio/Visual and Non-traditional Objects”).
The algorithm described in [21] has been used in the SHREC - 3D shape retrieval contest (watertight models track) organised by the European Network of Excellence AIM@SHAPE in 2007. (cf. [42]).

6.1.4. Local descriptors and visual saliency

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

Participants: Ahmed Rebai, Alexis Joly, Nozha Boujemaa.

We continue working on the definition of new salient local features with the objective to increase the interpretability of the characterised patterns [31]. We propose indeed a new interest points detector. Unlike most standard detectors which concentrate on the local shape of the signal, the main objective of this new operator is to extract interpretable points from the image context. The basic principle of this operator was the detection of radial symmetries, but we have generalised it to cover other kind of interest points. Indeed, detected points constitute centers of circles or logarithmic spirals, intersections of curves and vanishing points. Detection of such points is performed using a three dimensional space called θ-space. Experiments reveal that these points are more likely to be related to visual attention. We have also applied these points to the object recognition problem and to the state of the art performances using non specialised histogram descriptors (see Fig. 6).

Figure 6. Detection of interpretable interest points in object classes

The definition of new local descriptors adapted to the extracted features has been a challenging task. To this end, we have developed a generic framework allowing the construction of different kinds of shapes and geometric patterns. Global descriptors using colon, shape and texture could be computed on these geometric supports visualised as masks. Preliminary experiments were carried out and reveal some properties of local features.

We also developed new local photometric descriptors based on dissociated dipoles for transformed images or rigid objects retrieval [27]. Dissociated dipoles are non local differential operators which are more stable than purely local standard differential operators, particularly for small localisation errors (see Fig. 8 and 7).

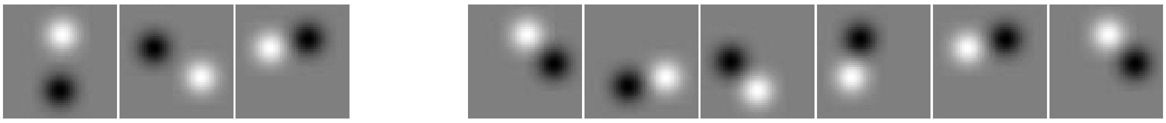

Figure 7. First order filters bank F_1

In this work, we defined and computed specific oriented dissociated dipoles around multi-resolution colon Harris points forming 20-dimensional normalised features, invariant to rotation, affine luminance transformations, negative or flip. In a comparison with extensively used SIFT descriptors, we show that such descriptors are more efficient while containing 6 times less information. This allows the complete retrieval to be both more efficient and faster. This strategy ranked first in **ImagEval** http://www.imageval.org/ benchmark, which, as far as we know, is the only competition including a transformed image recognition task (or content-based copy retrieval task).

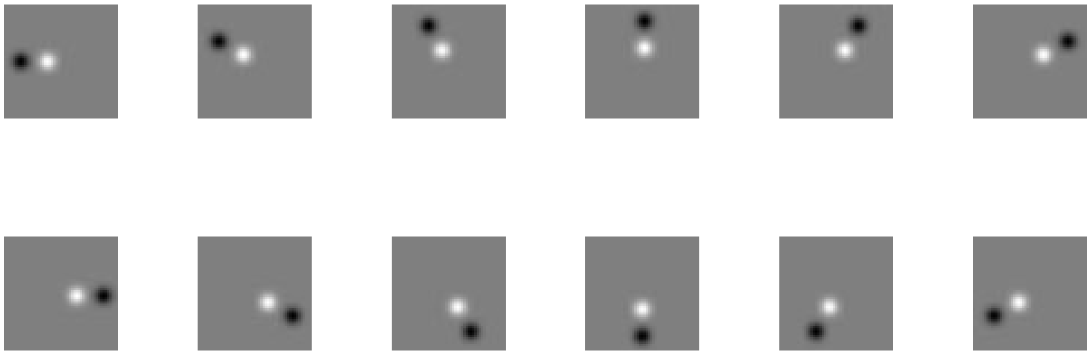

Figure 8. Second order filters bank F_2

6.1.5. Biodiversity image content description

Participants: Mathieu Coutaud, Nozha Boujemaa, Daniel Barthélémy [AMAP], Pierre Bonnet [AMAP].

Since October, a collaboration between IMEDIA and AMAP has started. The main goal is to investigate and evaluate the identification capabilities of botanical species using IMEDIA search engine. We have first started to work on orchid pictures from Lao, some preliminary experiments has been done with global descriptors to measure the difficulty to distinguish species in such a realistic database. Indeed, there are photographies of inflorescences, flowers but also leaves scans and pictures can be taken in the nature or in a greenhouse.

For the next year we will investigate the benefits of local description on this content, and we will also study the impact of merging different queries to enhance identification. For example, the user can provide multiple pictures of the same plant as a query and expects plant identification. For now, different image databases have been collected and a first RTRA project on weeds in Camargue ricefields has been accepted.

6.2. Interactive retrieval

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

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

Participants: Hichem Houissa, Nozha Boujemaa.

The paradigm of query by visual thesaurus was first initiated by Fauqueur and developed in [7] during his PhD work. The Visual thesaurus is built based on a database of regions obtained through image coarse segmentation. Depending on the complexity of regions, the use of specific descriptors is highly appropriate. On the one hand, colour distributions were used in case of homogeneous visual patches. On the other hand, we apply local descriptors depicted by Harris colour points of interest associated to Hilbert invariants to find out the visual content of textured regions [41]. The hierarchical visual thesaurus is structured into textured visual patches and homogeneous visual patches summary (see Figure 9).

Beside regions categorisation, a mental query involves spatial constraints applied by user when composing his starting example. To do so, directional relationships were introduced and a novel spatial descriptor (Weighted Angle Spatial Histogram) was developed to encapsulate the topology of regions and their compacity. The validation tests were conducted on generalistic databases. We figure out that as the formalisation of relations between regions are more and more fine and detailed, the retrieved images corresponding to a mental query example are more close to the end-user visual representation of a given scene. Figure 10 shows an example

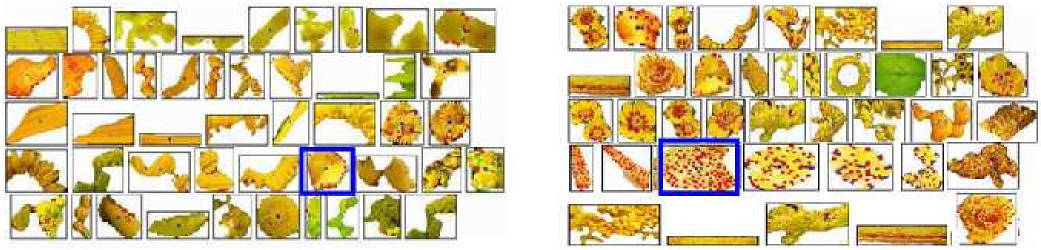

Figure 9. Example of textured yellow patches cluster together with homogeneous yellow ones. These clusters are represented in the thesaurus by their prototype highlighted by the blue box in the figure.

of mental query and its corresponding retrieved results. The query composition is a suitable search paradigm in case no starting example is available or for retrieving **semantic concepts** that cannot be reached through query by visual example [26].

Figure 10.

Example of visual composition through visual patches and spatial consolidation through WASH attributes that allows ***semantic concept*** *retrieval.*
Left: Query composition ("beach scene" concept). Right: retrieved images

6.2.2. Bayesian framework for semantic categories matching

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

Participants: Marin Ferecatu, Donald Geman, Nozha Boujemaa.

Traditional image retrieval methods, for example query by example and retrieval by relevance feedback, require a “query image” to initiate a search for members of an image category. However, when the image database is unstructured, and when the category is semantic and resides only in the mind of the user, there is no obvious way to begin (the “page zero” problem).

Our contribution is a method for discovering an instance from a semantic image category residing only in the mind of the user. The search is terminated upon displaying one of these images and performance is measured by the expected number of iterations necessary to achieve this. No semantic annotation is assumed. Also, unlike previous approaches to mental matching [38], ours extends to category search and large, unstructured

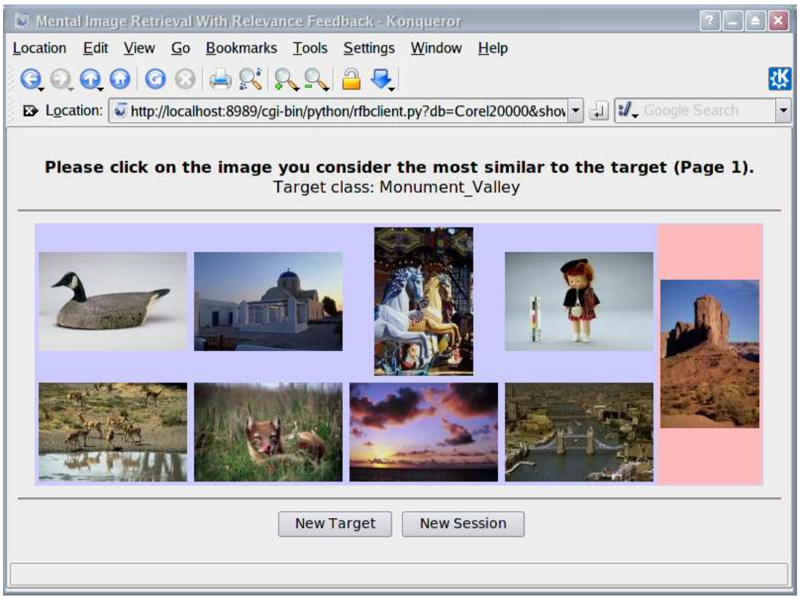

Figure 11. Retrieval by mental matching: at each feedback iteration the user chooses the image closest in his opinion to the semantic target class.

databases. It could serve either as a standalone function in a retrieval system or as a method for initialising another session, such as query-by-example, to obtain additional examples.

The theoretical contribution is a mathematical formulation of the problem adapted to large categories and large databases. We overcome the combinatorial intractability of constructing a probability distribution over categories by maintaining parallel Bayesian update systems, one per image. We also present a new response model which accounts for the nature of human decision-making, capturing the gap between the user's "metric" and the one used by the system. These two new features render our work distinct from previous efforts in category search.

The efficiency of the search is illustrated by experiments in which real users, working from an interface (see Fig. 11), attempt to locate an instance from a semantic target category of order 100 in a database of size 20,000. In fifty percent of the cases, four or fewer iterations are necessary and ten or fewer are usually sufficient. We also provide an in-depth analysis of the issue of coherence between the "metric" of the user and the one employed by the system based on low-level features [24].

6.2.3. Visualisation methods for multimedia search engines

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

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

This year, inside VITALAS Integrated project, we studied visualisation methods adapted to multimedia search engines (see deliverable D7.2 "report on state of the art on advanced visualisation methods" of VITALAS project with Hervé Goèau , Agnès Saulnier, Jérôme Thièvre and Marie-Luce Viaud of INA) to develop new approaches helping the user to browse and navigate inside multimedia data in a multimedia search engine. We have chosen to investigate graph-based approaches to build cartographic views of the data. We will have a post-doctoral student to work with us on this subject in 2008.

6.2.4. Web image replicas mining

Participants: Amel Hamzaoui, Alexis Joly, Nozha Boujemaa.

In whatsoever manner they are stored, a considerable number of the diffused images are duplicated across the Web and multi-media databases. A first essential step of post processing that we have done is to gather all the copies of each document by using hierarchical clustering. Organising the results into different clusters facilitates users browsing. One will be able then to extract from the clusters the contextual characteristics related to the various contexts of use of each document (each cluster) and to exploit these new descriptors to seek, categorise or annotate the images of the catalogue. A geographical descriptor can reveal geographical dispersion of the images. A temporal descriptor can reveal persistence or not of an event according to the frequency of use. Moreover, considering the inefficiency of the current techniques to dissociate completely the copies of images of similar images, a descriptor like the scale factor and the orientation angle can describe the diversity of the transformations which has undergoes the copies of images. We have also extract keywords related to Web images. The histogram of the frequency of these words recovered for each image of the cluster can generate an automatic annotation of the whole cluster.

6.3. Automatic annotation and learning

6.3.1. Incorporating geometrical properties into the Bag-Of-Features framework. Application to Image Classification and Object-Class Recognition

Keywords: *Bag-Of-Features*, *Boosting*, *Correlograms*, *Image Classification*, *Object recognition*.

Participants: Jaume Amores, Marwen Ben-Ali, Petia Radeva, Nicu Sebe [UVA].

We studied the incorporation of geometrical properties into the Bag-Of-Features paradigm. For this purpose, we developed two complementary approaches: i) introduction of new spatial descriptors, and ii) introduction of geometrical constraints in the Bag-Of-Features framework.

The first approach is based on our Generalised Correlograms, recently published in [9]. The Generalised Correlogram is a three-dimensional histogram that measures the spatial distribution of local properties around a given point of the image. It is based on a log-polar quantisation that represents a tradeoff between robustness against clutter (by weighting more the local context) and distinctiveness (by gradually incorporating global context).

The Generalised Correlogram is built upon generic local properties that do not depend on the specific class of images we deal with, so that the image representation can be pre-calculated off-line [9]. Lately, we have explored the alternative case where correlograms are built upon class-specific local features. This is done by extracting a vocabulary of features that are specific of our class of images, and building the correlograms onto this type of features. In [19] we show that: i) a compact representation can be obtained by binarising the correlograms based on thresholds obtained by a Boosting classifier, and ii) we can efficiently learn these correlograms if we integrate a Conditional Mutual Information Maximisation algorithm adapted to binary descriptors.

Finally, following a complementary direction, we studied the incorporation of geometrical constraints in the Bag-Of-Features paradigm. The standard approach is based on traditional distance functions, such as the Euclidean, for finding what features from the image are similar to those features stored in a given vocabulary. We studied the benefits of incorporating spatial coherence into this distance function, where now the distance function not only reflects the similarity of local appearance, but also the similarity of the matching in the spatial neighbourhood. This method is described in [35].

6.3.2. Learning distance functions for Image Retrieval

Keywords: *Boosting*, *Content-Based Image Retrieval*, *learning distance*.

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

We proposed a novel approach for learning similarity functions with application to exemplar-based image retrieval. First, several base metrics are introduced that fit the statistical distribution of noise in the data. Then, we propose to use a new Boosted Distance that, given a set of equivalence constraints introduced as training data, obtains a weighted combination of these base metrics in order to obtain an accurate distance function. The resulting distance function was seen to outperform a large battery of traditional metrics, including distance estimation techniques such as Mahalanobis, RCA-based metrics, and other recently proposed distance estimation techniques. This method appears in [18].

6.3.3. Automatic image annotation

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

Participants: Nicolas Hervé, Nozha Boujemaa.

Following the effort made during last year within ImagEVAL benchmark, we continued the evaluation of our automatic image annotation framework on standard databases used by the community. One of the main conclusion is that such databases are not appropriate any more for research on automatic image annotation. The issue here is that we could not have the same corpora for global image annotation task (such as indoor, outdoor,...) and image annotation that corresponds to local image patches including objects. We need to move to realistic images collections, with hundreds of thousand of pictures and thousands of concepts to manage. The approach used for the ImagEVAL benchmark, as well as the extended tests are published in this paper [25].

Moving to such databases introduces new challenges that we studied deeply this year. We focused more specifically on two main issues :

- the quality of a visual vocabulary for the bag-of-word learning approaches to concept detection
- the quality of available annotations

For both of these topics, we investigated new approaches in order to improve the already available results.

In order to conduct our experiments on automatic images annotation on huge databases we developed two toolkits. The Semantic toolkit provides all the necessary functionalities to manage and process the available annotations. The Visual vocabulary optimisation toolkit provides methods to create and optimise visual vocabularies, as well as bag-of-words learning strategies. They are based on our global MAESTRO framework and have been carefully developed to allow massive parallelisation on a cluster of CPUs. By using them together, we are able to manage the full concept learning/prediction chain.

6.3.4. Similarity measures and satellite image annotation

Keywords: *feature contrast model*, *latent semantic analysis*, *multiple instance learning*, *similarity measure*.

Participants: Hong Tang, Nozha Boujemaa, Henri Maitre [ENST].

Similarity measure by integrating both common and distinctive labels
We introduced new similarity measures between satellite images by integrating both common and distinctive labels. First of all, all images are segmented and a low-level feature is extracted from each region in an images. Then, a hierarchical clustering algorithm is used to cluster regions of all images, and each region is assigned a label. Moreover, we define a variable to reflect the degree of content variation in each image according to the hierarchical structure of regions. At last, we measure the similarity between two images as a linear combination between common and re-weighted distinctive labels, i.e., how many labels are common to the two images, and how many labels are different between two images. The weight is defined to reflect the “distinctiveness” of the different labels. Experimental results show the defined weight is good to discriminate those similar images in terms of part of visual features (cf. [32]).

Automatic image annotation by learning from loosely annotated images
Our annotation approach consists in assigning keywords to an image by learning the relationship between regions (or keywords) in a same image, and the correspondence between image regions and keywords.

Latent semantic analysis (LSA) has been used to annotate image for indexing through learning the relationship between keywords. However, un-observed keywords in a training image will not be modelled during learning topic models. We believe that, as co-occurrence between keywords, some un-observed keywords in training images are also an important cue to learn the topic models. The reason is that some keywords will not occur together, and absence of certain keywords might be related to a different topic. Therefore, we propose a method to explicitly "imagine" an observation for un-observed keyword through the similarity between keywords, then topic model are learned from both real observation and "imagined" observation. At last, the learned topic model is used to annotate new images. Experimental results show that our method exhibits some very attractive characters.

6.3.5. ImagEVAL benchmark

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

IMEDIA has participated to ImagEVAL benchmark initiative and won all the following tasks:

- Task 1 : Transformed images retrieval

- Task 4 : Objects and objects class retrieval

- Task 5 : Semantic attributes extraction

Task 1 was processed thanks to a new content-based copy detection technique based on new local descriptors coupled with global geometric constraints and an efficient indexing structure. Task 4 was processed with IMEDIA standards descriptors used in a local and multi-resolution context and processed by several new multiple instances learning strategies. Task 5 was processed with a semantical hierarchy of classifiers based on IMEDIA standards global descriptors and support vector machines. For further details, see Section 7.1.5.

6.4. Software

6.4.1. Clustering toolbox

Keywords: *non-exclusive clustering algorithms*.

Participants: Philippe Tony, Itheri Yahiaoui, Nozha Boujemaa.

We are investigating new approaches of non-exclusive clustering. Our goal is to propose new objective functions in order to improve clustering performance allowing the reduction of the semantic gap. In this regard, we aim to test different variants of proposed algorithms and compare them with existing methods.

Building a visual summary of an image database and image segmentation are both usages for clustering methods. Many clustering methods have been studied by team members, and many methods have been separately implemented. Regrouping the most common methods in a single versatile library, offers easier implementation and new approaches like method comparisons, and automatic parameters discovery.

The different clustering methods designed and developed by different team members have been regrouped in a single common library.

A framework for data clustering software prototype has been designed and implemented during this year as part of an internship.

The current version includes the following clustering methods:

K-Means, Fuzzy C-Means, Competitive Agglomeration (with multiple objective functions), Quality Threshold, Piecewise-Constrained CA, Relational Fuzzy C-Means, CA for Relational Data (CARD).

The different methods have been internally regrouped by types (in two levels):

- non-exclusive algorithms

- exclusive algorithms

- relational algorithms (where the only input is the distance matrix between the data points, and no barycenter can be computed)

- prototypical algorithms (where barycenter can be computed)

Each type of algorithm has its own data structure and methods, which tends to make the code less redundant, and implementing new algorithms easier. The two-level-structure may be extended with further levels to accommodate with new types and variations of algorithms.

6.4.2. IKONA/MAESTRO software

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

Participants: Mehdi Bouabta, Mathieu Coutaud, Nicolas Hervé, Laurent Joyeux.

This year, the main development relates to the integration of signatures storage in a database management system. So, signatures can still be stored in files (besides different formats are enabled: XML, full text...) but now also in a real database system (Oracle, MySQL, PostgreSql). A wide range of new functionalities and improvements have been integrated, such as a memory cache to cope with huge databases, a similarity search structures library (to speed up queries), a visual thesaurus library (construction and navigation), protocoled services (similarity matrices, signatures extraction), new image analysis features (new interest points, image masks, grids...). We can also mention some improvements concerning the mass annotation mechanism, the graphical user interface and the MAC OS portability.
At the same time, two important developments have been made outside the main platform. First, an API converting maestro services into web services was developed as well as a new web client.

For the next year, we foresee to develop an administration interface to facilitate access to services: server configuration, database indexing...

6.4.3. Development of components for a content based image Web search engine

Participants: Saloua Litayem, Alexis Joly, Nicolas Hervé.

Components were developed in order to integrate IKONA system into a complete Web search engine by making it compatible with standard web services.

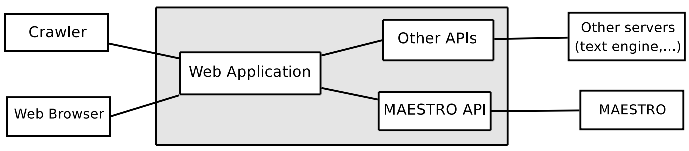

Figure 12. Search engine architecture

The figure 12 presents the search engine architecture. The three following tasks were achieved:

- Development of an API for textual search of images in the Web.
- Development of an API which will allow the MAESTRO server to communicate with other applications (deployed on Yahoo! search engine).
- Development of a Web application using the previous APIs.

On the side of the client (web browser), the server is constituted by MAESTRO + APIs + Web application. And on the side of the MAESTRO server, the developed components (APIs+web application) will constitute a new client. This client will has the following new functionalities:

- Many request ways (textual search, search by example image, random navigation ...)
- Additional informations (link to the original content, search's score, size of the image ...).

Figure 13. Web GUI of the application

The figure 13 illustrates a web search result using the associated developed API.

7. Other Grants and Activities

7.1. National Initiatives

7.1.1. Collaboration with AMAP

Since October, a collaboration between IMEDIA and AMAP has started. AMAP is a joint research unit between INRA, CIRAD, IRD and CNRS located in Montpellier. AMAP is providing financial support for research engineer contract to work on biodiversity content search problems together with IMEDIA. During 2007, a first RTRA project on weeds in Camargue ricefields has been accepted to reinforce this collaboration.

7.1.2. Industrial contract with INA [2004-2007]

A co-supervision of a Phd within CIFRE context. Julien Law-to defended his PhD on December 14th. The thesis deals with the description of digital videos, from genericity to distinctiveness. One particular targeted application was the detection of video copies on the web and in TV streams from very large video archives.

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

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

7.1.4. InfoMagic[2005-2008]

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

7.1.5. Benchmark Initiative : ImagEVAL

The ImagEVAL project has a national financial support but benefit of European research lab. participation (Techno-Vision, Ministère de la recherche http://www.imageval.org/presentation.html. It is animated by CEA-LIST and supervised by DGA and Nicephore Cité. IMEDIA did win all image retrieval tasks (3) in both the blank and the official tests.

Official results were published mid-December, 2006. ImagEVAL relates to the evaluation of technologies of content-based image retrieval (CBIR) and automatic image annotation in large-scale image databases. It is the first benchmark comparing image search engines using visual content and in a realistic context (images coming from professional content providers and annotated by real archivists: Hachette agency). French academic teams (ENSEA, CEA, Ecole des Mines de Paris,...), European academic teams (Vienna University of Technology, University of Geneva, CERTH in Grece) as well as industrial teams (Canon, LTU technologies, AdVestigo) have participated.

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). IMEDIA is in charge of the development of the "Intelligent Search and Retrieval" application module. This module brings together the software of four research teams that work on different multimedia information retrieval paradigms. This year IMEDIA activities in aceMedia stopped in March, these three months enable to fit last requirements for the PCS User Interface and to finish the video support.

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 the scientific coordinator of the network.

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

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

7.2.4. Strep "TRENDS" [2006-2008]

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

7.2.5. Integrated project "VITALAS" [2007-2009]

"Video & image indexing and retrieval in the large scale" (http://vitalas.ercim.org/) in the call6 of 6th Framework Programme. VITALAS is an innovative project designed to provide advanced solution for indexing, searching and accessing large scale digital audio-visual content through cross-modal content enrichment and personalised. The strength of this initiative relies on the capacity of the project to confront its technology to real use-cases, reflecting the joint concerns of two major European content providers. The project will develop new technological functionalities and services to access to large scale multimedia databases. The project is composed of 12 industrial and academic European partners (ERCIM, EADS, CWI, Fraunhofer, Robotiker, INA, Univ. of Sunderland, CERTH-ITI, Codeworks, Belga, IRT). AFP (Agence France Press) has recently joined the project. Nozha Boujemaa is the scientific coordinator of the project, Alexis Joly and Anne Verroust-Blondet are Workpackage leaders.

7.2.6. Coordination Action "CHORUS" [2007-2009]

CHORUS is a coordination Action in the field of Audio-Visual Search Engines accepted in the call6 of the 6th Framework Programme (http://www.ist-chorus.org/). An important objective of the project consists of supporting the preparation of an analysis and a roadmap for the realisation of Audio-visual search engines in EU. Hence, CHORUS coordinates all the ongoing European and national efforts/projects on the topic of "multimedia search engine". The consortium represents established and well-reputed research institutions and consultancies with a broad range of intellectual and technological expertise in the area, both as regards concrete actions and policy development and track records of national, Union-wide, and international cooperation and activity (Thomson, Philips, JCP consult, IRT, France Telecom, Exalead, etc.). Nozha Boujemaa is the scientific co-ordinator of the project.

7.3. International Initiatives

7.3.1. Organisation of MUSCLE/CIVR video copy detection benchmark [2007]

IMEDIA organised the first international benchmarking initiative for content-based video copy detection technologies, funded by MUSCLE Network of Excellence. A corpus of 100 hours of video was collected and a ground truth was build from real world experiences. The corpus has been nowadays distributed to more than 20 institutes and companies all around the world. A live evaluation showcase was organised during ACM CIVR 2007 conference (with the participation of IBM, AdVestigo, Chinese Accademy of sciences, Bilkent university, etc.).
For more information, see http://www-rocq.inria.fr/imedia/civr-bench/index.html.

7.3.2. Showcases and demos

IMEDIA developed a content based copy detection showcase in the context of the MUSCLE FP6 NoE. The showcase is composed of two live demos, one dedicated to video copy detection from a large dataset of INA archives and one for the live retrieval of still images captured from a digital camera. Two graphical user interfaces were developed as well as a communication layer. The demos have been played in several international conferences and trade shows:

- **IBC exhibition** (http://www.ibc.org/)
- **CeBit trade show** (http://www.cebit.de/homepage_e)
- **IEEE CIVR 2007** conference, demo section (http://www.civr2007.com/)

The showcase and other demos of the team concerning content based visual search engines were also presented during several European and national events (Salon européen de la recherche, MUSCLE meeting, Les rendez-vous d'OpticsValley, 40 ans de l'INRIA).

8. Dissemination

8.1. Leadership with scientific community

8.1.1. Nozha Boujemaa

- Scientific coordinator of MUSCLE NoE (Network of Excellence FP6);
- Scientific coordinator of VITALAS IP FP6;
- Scientific coordinator of CHORUS CA FP6;
- Co-chairman, with J.Wang, of the "9th ACM SIGMM International Workshop on Multimedia Information Retrieval" in conjunction with ACM Multimedia 2007 (MIR'07)
- Co-chairman, with A. Delbimbo, R. Cucciara of the International Workshop on Visual and Multimedia Digital Libraries (VMDL07)
- Chair of "National Initiatives on Multimedia Content Description and Retrieval" CHORUS event: http://www.ist-chorus.org/geneva—october-10th-07.php
- Chair of "Use Cases and New Services for Multimedia Content Search" http://www.ist-chorus.org/rocquencourt–mar-13–14-07.php
- Session chair "Science to Business" ACM CIVR07 "Practitioner Day"
- Session chair of EC cluster meeting on "Evaluation and Benchmarking"
- Panel Chair at CBMI 07
- Scientific evaluator (panellist) for NSF projects
- French expert for COST ICT Domain (intergovernmental network for European Cooperation in the field of Scientific and Technical Research)
- Member of "European Commission Task Force on Networked Media: long term research"
- Member of ACM - SIGMM committee
- Member of the Editorial board of scientific journals: I3, PRA
- Member of several Technical program committees (TPC) of major international conferences: ACM MM, ACM, CIVR, ACM, MIR, IEEE ICME, ICPR, CBMI, SAMT...
- Member of the Steering Board of NEM ETP (Networked and Electronic Media European Technology Platform) and represent INRIA at this ETP;

- Member of the "National Evaluation Commission" of INRIA, member of recruitment committee of INRIA (IR, IE)
- Several PhD Jury committee members-president: Tampere University (Helsinki), ENST...

8.1.2. Michel Crucianu

- Scientific expert for the French national Research Agency (ANR), call “Audiovisuel et Multimédia”.
- Coordinator of the BIOTIM project of the French national initiative “Masses of data”.
- Leader of the e-team “Active Semi-Supervised Learning” in the MUSCLE FP6 NoE.
- Journal reviewer: IEEE Trans. On Pattern Analysis and Machine Intelligence, Multimedia Tools and Applications, Pattern Recognition.
- HDR jury member:
  - University of Nantes.
  - University of Tours.

8.1.3. Alexis Joly

- Steering comittee of VITALAS IP FP6 (leader of WP2 "Enabling technologies: Media Content Description and Summarisation").
- Co-leader of the e-team "Visual saliency" in the MUSCLE FP6 NoE.
- Co-leader of the showcase "Copy detection" in the MUSCLE FP6 NoE.
- Scientific expert for the French National Research Agency (ANR), call "Audiovisuel et Multimédia" and "Masse de données et connaissances".
- Journal Reviewer: IEEE Transactions on Pattern Analysis and Machine Intelligence.

8.1.4. Anne Verroust-Blondet

- Committee member of the call “Outils de la recherche en sciences humaines et sociales, CORPUS 2007” for the French national Research Agency (ANR).
- Member of the steering committee of VITALAS IP FP6 (leader of WP7:“user interface and visualisation”),
- Member of the programme committee of Semantic And digital Media Technology (SAMT’07) conference,
- AFIG Vice-President (Association Française d’informatique Graphique) ;
- Member of the steering committee of the CNRS GDR IG (Informatique Graphique) ;
- Member of the editorial board of the “Revue Electronique Francophone d’Informatique Graphique” ;
- Co-editor of the special issue “Informatique graphique” of Technique et Sciences Informatiques (TSI) journal (issue 9 in 2007)
- Journal Reviewer: IEEE Transactions on Visualization and Computer Graphics ;

8.2. Teaching

8.2.1. Mohamed Chaouch

- 50 hours of TD on Probability and Statistics, Master 1 “Economie et gestion”, University of Paris-Sud.

8.2.2. Michel Crucianu

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

8.2.3. Valérie Gouet-Brunet

- 192 hours in the Computer Science Department of CNAM;
- National responsible for the course "Computer Vision" of the Master research STIC - Computer Science of CNAM (6 ECTS - 60 hours);
- Course “Multimedia Databases” of the Master in Computer Science of the University Paris Dauphine.

8.2.4. Nicolas Hervé

- 16h TP on Java Database Connectivity (JDBC), CNAM 3rd year (NFA011), January 2007

9. Bibliography

Major publications by the team in recent years

[1] J. AMORES, N. SEBE, P. RADEVA. *Context-Based Object-Class Recognition and Retrieval by Generalized Correlograms*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", vol. 29, n° 10, 2007, p. 1818 – 1833.

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

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

[4] M. FERECATU, N. BOUJEMAA, M. CRUCIANU. *Semantic interactive image retrieval combining visual and conceptual content description*, in "ACM Multimedia Systems", 2007.

[5] N. GRIRA, M. CRUCIANU, N. BOUJEMAA. *Active semi-supervised fuzzy clustering*, in "Pattern Recognition", In Press, corrected proof, 2007.

Year Publications

Books and Monographs

[6] D. BECHMANN, A. VERROUST-BLONDET (editors). *"Informatique Graphique" special issue of TSI*, vol. 9, 2007.

Doctoral dissertations and Habilitation theses

[7] H. HOUISSA. *Recherche par thésaurus visuel et composition spatiale dans les bases d’images*, Ph. D. Thesis, Paris-Sud University, June 2007.

[8] J. LAW-TO. *From genericity to distinctiveness of video content description. Application to Video Copy Detection*, Ph. D. Thesis, Université Versailles Saint-Quentin / INRIA IMEDIA / INA, December 2007.

Articles in refereed journals and book chapters

[9] J. AMORES, N. SEBE, P. RADEVA. *Context-Based Object-Class Recognition and Retrieval by Generalized Correlograms*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", vol. 29, n° 10, 2007, p. 1818 – 1833.

[10] D. BECHMANN, A. VERROUST-BLONDET. *Animation par déformations et métamorphoses*, in "Informatique Graphique, modélisation géométrique et animation", Traité IC2, Hermès, March 2007, p. 253-289.

[11] N. BOUJEMAA, M. FERECATU. *Description des images fixes*, in "Indexation multimédia: Description, indexation et recherche des documents multimédias par leur contenu", Traité IC2, Hermès, June 2007.

[12] M. FERECATU, N. BOUJEMAA, M. CRUCIANU. *Semantic interactive image retrieval combining visual and conceptual content description*, in "ACM Multimedia Systems", 2007.

[13] M. FERECATU, N. BOUJEMAA. *Interactive Remote-Sensing Image Retrieval Using Active Relevance Feedback*, in "IEEE Transactions on Geoscience and Remote Sensing", vol. 45, n° 4, April 2007, p. 818-826.

[14] N. GRIRA, M. CRUCIANU, N. BOUJEMAA. *Active semi-supervised fuzzy clustering*, in "Pattern Recognition", In Press, corrected proof, 2007.

[15] A. JOLY, C. FRÉLICOT, O. BUISSON. *Content-based copy retrieval using distortion-based probabilistic similarity search*, in "IEEE Trans. on Multimedia", vol. 9, n° 2, 2007, p. 293-306.

[16] H. TANG, T. FANG, P. DU, P. SHI. *Intra-dimensional Feature Diagnosticity in the Fuzzy Feature Contrast Model*, in "Image and Vision Computing", in press, 2007.

[17] J. Z. WANG, N. BOUJEMAA, Y. CHEN. *High diversity transforms multimedia information retrieval into a cross-cutting field: report on the 8th Workshop on Multimedia Information Retrieval*, in "SIGMOD Rec.", vol. 36, n° 1, 2007, p. 57-59.

[18] J. YU, J. AMORES, N. SEBE, Q. TIAN. *Distance Learning for Similarity Estimation*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", to appear, 2007.

Publications in Conferences and Workshops

[19] J. AMORES, N. SEBE, P. RADEVA. *Class-Specific Binary Correlograms for Object Recognition*, in "British Machine Vision Conference (BMVC)", 2007.

[20] O. BESBES, Z. BELHADJ, N. BOUJEMAA. *A Variational Framework for Adaptive Satellite Images Segmentation.*, in "SSVM", LNCS Springer 4485, 2007, p. 675–686.

[21] M. CHAOUCH, A. VERROUST-BLONDET. *3D Model Retrieval Based on Depth Line Descriptor*, in "IEEE International Conference on Multimedia and Expo 2007 (ICME 2007), Beijing, China", July 2007.

[22] M. CHAOUCH, A. VERROUST-BLONDET. *A New Descriptor for 2D Depth Image Indexing and 3D Model Retrieval*, in "IEEE International Conference on Image Processing 2007 (ICIP 2007), San Antonio, Texas, USA", September 2007.

[23] M. CRUCIANU, D. ESTEVEZ, V. ORIA, J.-P. TAREL. *Hyperplane Queries in a Feature-Space M-tree for Speeding up Active Learning*, in "Bases de Données Avancées, Marseille, France", October 2007.

[24] M. FERECATU, D. GEMAN. *Interactive search for image categories by mental matching*, in "Proceedigs of the Inter. Conf. on Computer Vision (ICCV '07)", October 2007.

[25] N. HERVÉ, N. BOUJEMAA. *Image annotation : which approach for realistic databases ?*, in "Proc. of the ACM International Conference on Image and Video Retrieval (CIVR'07), Amsterdam", July 2007.

[26] H. HOUISSA, N. BOUJEMAA. *A new angle-based spatial modeling for query by Visual Thesaurus composition*, in "Proc. of IEEE International Conference on Image Processing", 2007.

[27] A. JOLY. *New local descriptors based on dissociated dipoles*, in "CIVR '07: Proceedings of the 6th ACM international conference on Image and video retrieval, New York, NY, USA", ACM, 2007, p. 573–580.

[28] J. LAW-TO, O. BUISSON, V. GOUET-BRUNET, N. BOUJEMAA. *Video Copy Detection on the Internet: the Challenge of Copyright and Multiplicity*, in "IEEE International Conference on Multimedia & Expo (ICME 2007), Beijing, China", July 2007.

[29] J. LAW-TO, L. CHEN, A. JOLY, I. LAPTEV, O. BUISSON, V. GOUET-BRUNET, N. BOUJEMAA, F. STENTIFORD. *Video Copy Detection: a Comparative study*, in "ACM International Conference on Image and Video Retrieval, Amsterdam, The Netherlands", July 2007.

[30] J. LAW-TO, A. JOLY, L. JOYEUX, N. BOUJEMAA, O. BUISSON, V. GOUET-BRUNET. *Video and image copy detection demo*, in "CIVR '07: Proceedings of the 6th ACM international conference on Image and video retrieval, New York, NY, USA", ACM, 2007, p. 97–100.

[31] A. REBAI, A. JOLY, N. BOUJEMAA. *Interpretability based interest points detection*, in "CIVR '07: Proceedings of the 6th ACM international conference on Image and video retrieval, New York, NY, USA", ACM, 2007, p. 33–40.

[32] H. TANG, H. MAITRE, N. BOUJEMAA. *Similarity measure for satellite images with heterogeneous contents*, in "URBAN - 2007 - IEEE/ISPRS Joint Workshop on Remote Sensing and data fusion over urban areas, Paris, France", April 2007.

Miscellaneous

[33] M. BEN-ALI. *Méthodes de sacs de primitives dans la reconnaissance d'objets dans les images*, Technical report, UFR de Mathématiques et Informatique, Université Paris Descartes, 2007.

References in notes

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

[35] M. BEN-ALI. *Méthodes de sacs de primitives dans la reconnaissance d'objets dans les images*, Master thesis, UFR de Mathématiques et Informatique, Université Paris Descartes, 2007.

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

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

[38] I. J. COX, M. L. MILLER, T. P. MINKA, T. PAPATHOMAS, P. N. YIANILOS. *The Bayesian image retrieval system, PicHunter: theory, implementation and psychophysical experiments*, in "IEEE Transactions on Image Processing", vol. 9, n^{o} 1, January 2000, p. 20–37.

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

[40] X. HE, R. ZEMEL, M. CARREIRA-PERPINAN. *Multiscale Conditional Random Fields for Image Labeling*, in "CVPR", vol. 02, IEEE Computer Society, 2004, p. 695–702.

[41] H. HOUISSA, N. BOUJEMAA. *On the use of metrics for multi-dimensional descriptors clustering*, in "Proc. of IEEE International Conference on Image Processing", 2006, p. 153-156.

[42] R. C. VELTKAMP, F. B. TER HAAR. *SHREC2007, 3D in Shape Retrieval Contest 2007 , (M. Chaouch and A. Verroust-Blondet, 2D/3D Descriptor based on Depth Line Encoding)*, Technical report, UU-CS-2007-015, Department of Information and Computing Sciences, Utrecht University, 2007.