INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Project-Team Imedia

Images and Multimedia: Indexing, Retrieval and Navigation

Paris - Rocquencourt

Theme : Vision, Perception and Multimedia Understanding

Activity Report

2010

Table of contents

1. Team

Research Scientists

Nozha Boujemaa [Team leader until October 31st 2010, Senior Researcher, HdR]
Alexis Joly [Junior Researcher]
Anne Verroust-Blondet [Interim team leader since November 1st 2010, Junior Researcher, HdR]

External Collaborators

Olivier Buisson [Research Scientist at INA, France]
Michel Crucianu [Professor at CNAM, HdR]
Itheri Yahiaoui [Associate Professor at Reims University]

Technical Staff

Jean-Paul Chièze [Senior Technical Staff INRIA (half-time)]
Mondher Khadraoui [Junior Technical Staff INRIA until August 31st 2010]
Vincent Ladeveze [Technical Staff INRIA]
Souheil Selmi [Junior Technical Staff INRIA]

PhD Students

Sahbi Bahroun [Joint tutorship with Sup'Com, national grant until November 2010]
Esma Elghoul [INRIA grant, Télécom ParisTech since October 15 2010]
Mehdi Ellouze [Joint tutorship with University of Sfax, national grant until December 2010]
Amel Hamzaoui [INRIA grant, Paris-Sud University since October 1st 2008]
Pierre Letessier [CIFRE with INA since October 1st 2009]
Sofiène Mouine [INRIA grant, Télécom ParisTech since October 15 2010]
Saloua Ouertani-Litayem [INRIA grant, Télécom ParisTech since October 1st 2009]
Wajih Ouertani [INRA grant, Paris-Sud University since October 1st 2008]
Ahmed Rebaï [INRIA grant, Paris-Sud University since September 1st 2007]
Asma Rejeb Sfar [INRIA grant, Télécom ParisTech since October 15 2010]
Mohamed Riadh Trad [INRIA grant, Télécom ParisTech since October 1st 2009]

Post-Doctoral Fellows

Mohamed Chaouch [Expert engineer INRIA until August 15th 2010]
Raffi Enficiaud [Expert engineer INRIA]
Joost Geurts [Postdoctoral fellow until August 31st 2010]
Hervé Goëau [Expert engineer INRIA]
Laurent Joyeux [Expert engineer INRIA since July 1st 2010]

Visiting Scientist

Donald Geman [Professor at Johns Hopkins University, USA June and July 2010, HdR]

Administrative Assistant

Laurence Bourcier [Secretary (TRS) INRIA (shared with Salsa and Micmac project-team)]

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 and may contain 3D objects. 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 and efficient 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 (text, image, video, 3D models). 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 proved 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. 2D or 3D indexing: this part mainly concerns modeling the visual aspect of images and of 3D shapes, by means of image analysis techniques. It leads to the design of signatures that can then be obtained automatically.
2. feature space structuring: to increase in efficiency the search by content in very large collections of images.
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.

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

2.2. Highlights of the year

- VITALAS project ended with success in March 2010 (see a video demo on VITALAS Multimedia search engine on http://www-rocq.inria.fr/imedia/index.php?option=com_content&view=article&id=95&Itemid=60)
- A book on the future of the multimedia search engines [15].
- Our Pl@ntScan demo on automatic tree identification (http://imedia-ftp.inria.fr:8080/demo_plantscan/)

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 and efficient search in image collections, pattern recognition and personalisation. In the following we give a short introduction to each of these themes.

3.2. Modelling, construction and structuring of the feature space

Participants: Nozha Boujemaa, Mohamed Chaouch, Jean-Paul Chièze, Raffi Enficiaud, Amel Hamzaoui, Alexis Joly, Sofiène Mouine, Saloua Ouertani-Litayem, Ahmed Rebai, Mohamed Riadh Trad, Anne Verroust-Blondet, Itheri Yahiaoui.

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. Today's scalability issues already put brake on growth of multi-media search engines. The searchable space created by the massive amounts of existing multimedia files greatly exceeds the area searched by today's major engines. Consistent breakthroughs are therefore urgent if we don't want to be lost in data space in ten years. We believe that reducing algorithm complexity remains the main key. Whatever the efficiency of the implementation or the use of powerful hardware and distributed architectures, the ability of an algorithm to scale-up is strongly related to its time and space complexities. Nowadays, efficient multimedia search engines rely on various high level tasks such as content-based search, navigation, knowledge discovery, personalization, collaborative filtering or social tagging. They involve complex algorithms such as similarity search, clustering or machine learning, on heterogeneous data, and with heterogeneous metrics. Some of them still have quadratic and even cubic complexities so that their use in the large scale is not affordable if no fundamental research is performed to reduce their complexities. In this way, efficient and generic high-dimensional similarity search structures are essential for building scalable content-based search systems. Efficient search requires a specific structuring of the feature space (multidimensional indexing, where indexing is understood as data structure) for accelerating the access to collections that are too large for the central memory. The applications we have in mind are related to biodiversity (as in Pl@ntNet), to the detection of illegal copies of images and video (with INA) and to video surveillance and monitoring (with AVT).

3.3. Pattern recognition and statistical learning

Participants: Nozha Boujemaa, Michel Crucianu, Donald Geman, Hervé Goëau, Amel Hamzaoui, Alexis Joly, Wajih Ouertani, Ahmed Rebai.

Statistical learning and classification methods are of central interest for content-based image retrieval [23], [28]. 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 adaptively identify homogeneous clusters of images, which represent a challenging problem due to feature space configuration.

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.

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

Participants: Sahbi Bahroun, Nozha Boujemaa, Michel Crucianu, Mehdi Ellouze, Donald Geman, Wajih Ouertani, Anne Verroust-Blondet, 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 interactive visualisation tools to help the user when querying and browsing the database,
- 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.

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).
- **Audio-visual applications** 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), IRT (German broadcasters) and press agencies AFP and Belga in the context of an 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 for species identifications.
- **Culture, art and design** 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. IMEDIA has been involved in TRENDS European project dedicated to provide designers (CRF Fiat, Stile Bertone) with advanced content selection and visualisation tools.

5. Software

5.1. IKONA/MAESTRO Software

Participants: Nozha Boujemaa, Marin Ferecatu, Nicolas Hervé, Jean-Paul Chièze, Mathieu Coutaud, Alexis Joly, Mehdi Bouabta, Raffi Enficiaud, Mondher Khadhraoui, Souheil Selmi, Francois Fleuret.

IKONA is a generalist software dedicated to content-based visual information indexing and retrieval. It has been designed and implemented in our team during the last years [24]. Its main functionalities are the extraction, the management and the indexing of many state-of-the-art global and local visual features. It offers a wide range of interactive search and navigation methods including query-by-example, query-by-window, matching, relevance feedback, search results clustering or automatic annotation. It can manage several types of input data including images, videos and 3D models.

Based on a client/server architecture, it is easily deployable in any multimedia search engine or service. 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. 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
- on-line processes: answer the client requests

Several clients can communicate with the server. A good starting point for exploring the possibilities offered by IKONA is our web demo, available at http://www-roc.inria.fr/cgi-bin/imedia/circario.cgi/bio_diversity?select_db=1. 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.

IKONA is a pre-industrial prototype, with exploitation as a final objective. Currently, there does not exist a licensed competitor with the same range of functionalities. It exists several commercial softwares or systems exploiting technologies similar to some functionalities of IKONA but usually not the most advanced ones. We can for example cite the SDK developed by LTU company, the service proposed by AdVestigo company, etc.. It finally exists many prototypes and demonstrators, industrial or academic, sharing some functionalities of IKONA but here again not the most advanced (e.g. Google Image Similarity Search Beta, IBM Muffin, etc.).

The main originality of IKONA is its **genericity** (in terms of visual features, metrics, input data, storage format, etc.), its **adaptivity** (to new visual features, new indexing structures or new search algorithms), its innovative **interactive** search functionalities (Local and Global Relevance Feedback, Local Search with Query Expansion, Search results clustering, etc.) and its **scalability** thanks to a generic indexing structure module than can support the integration of any new advances.

Current Users of IKONA include European and National Projects Participants through its integration in prototype multimedia systems, commercial companies through user trials (EXALEAD, INA, BELGA, AFP), General or Specific Public through Web demos (Pl@ntNet leaf identification demo).

IKONA software provides a high degree of visibility to IMEDIA scientific works through demos in commercial, scientific and general public events (notably in most INRIA national showrooms). It is also the mainstay of several Multimedia Systems developed at the European level, in conjunction with many Leader European Companies and Research Centers. Discussions for a technological transfer of IKONA are currently progressing, notably with EXALEAD company (now part of Dassault) and Belga Press Agency.

5.2. PMH Library

Participants: Alexis Joly, Olivier Buisson [INA].

PMH is a generalist software library dedicated to locality sensitive hashing in metric spaces for approximate similarity search. It allows to index and exploit efficiently large datasets of content descriptors, usually represented by high dimensional feature vectors. The construction of the index and the required memory space are linear in dataset size. The nearest neighbour search algorithm is sub-linear in dataset size.
PMH is globally related to Locality Sensitive Hashing methods (LSH) that have been proved to be the most efficient ones for approximate similarity search in large and high-dimensional datasets. Contrary to classical LSH method (such as the ones used in MIT E2LSH package), PMH includes a multi-probe search algorithm which allows to drastically reduce the memory space complexity enabling to deal with datasets of several order of magnitude larger. Our multi-probe algorithm being based on a probabilistic control of buckets success probability also offers to control accurately the quality of the approximate search. Finally, PMH library is widely more generic than concurrent libraries (such as FLANN or LSHKIT). It allows the use of different

metric types (L1, L2, Hamming, inner product, weighted distances, etc.), different data types (binary, float, sparse, non vectorial, etc.), different query types (K nearest neighbours, range queries, probabilist queries, empirical models, etc.), different hashing functions families (random projections with different distributions, kernel based projections, optimized projections such as PCA or LDA, etc.).
Notably, PMH library is the core technology for the scalability issues addressed by VITALAS European project and is fully integrated in the resulting VITALAS multimedia search engine. It has been successfully applied to multi-users real-time content-based retrieval in 20 millions Flickr images and to real-time local search of small objects in a 100K images collection (including 120 millions SIFT features).

6. New Results

6.1. Feature space modelling

6.1.1. Detection of the grape leaves meaningful parts

Participants: Sofiène Mouine, Raffi Enficiaud, Nozha Boujemaa, Ezzedine Zagrouba.

In the scope of the Pl@ntNet project, we are interested in extracting meaningful parts of the grape leaves. In a previous work, morphological tools proved to be efficient for the segmentation of the leaves. For that purpose, we used a the watershed transform with manually marked areas. In the current work [20], we automated the segmentation process. For that purpose, we extracted the venation by means of morphological filtering, and inspired from [35], and used it as a marker for the blade area. Since the detected veins are disconnected, we proposed a reconstruction method. Having the veins, we are able to detect the base point, and together with the contour, the lobes and their apexes. A result is shown in Fig.1.

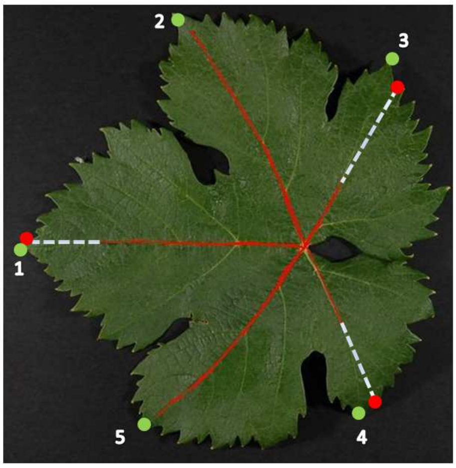

Figure 1. Detection of the main venation network, the lobes and their apexes.

We evaluated the relevance of the extracted informations with expert botanists. Theses tests showed that the venation and the boundary of the leaves carry pertinent information. However, limiting on these two morphological aspects is not sufficient for classifying properly the complete corpus, as the aspects of some varieties are very similar.

6.1.2. Automatic petiole detection

Participants: Olfa Mzoughi, Itheri Yahiaoui, Nozha Boujemaa.

Automatic leaf identification becomes the focus of many computer vision studies. Many well-known shape descriptions (such as the Hu and the Zernike moments) have been applied for this purpose. However, one of the problems that haves influenced the precision of these descriptors is the presence of the petiole, which is the small stalk located at the lower part of the leaf. This is due to the variability of its length and its bending. For that, some recent works [29], [36] have proposed to remove the petiole from the leaf before computing the descriptor. Nevertheless, their proposed solutions are not totally efficient: they are manual, semi-automated or fully-automated but specific to some particular species. For that, we introduced a new method in order to detect the petiole for a large variability of species. It is based on the local translational symmetry, which is defined as the local invariance of geometric properties along a direction [25], associated with a width description of the leaf shape that outlines the transition between the blade and the petiole. We define the components of the width descriptor as the distance between two consecutive points of the shape line, that have a distance lower than a threshold. The threshold is defined as a percentage of the width of the bounding box associated to the shape. The percentage is chosen w.r.t two criteria: the elongation of the shape and its blade organization.

- The elongation allows the separation between elongated and flattened shapes. In fact, in elongated shapes, the width ratio between blade and petiole is small so the threshold percentage should be small, unlike flattened shapes.
- The organization allows the separation between compound leaves and simple leaves that may have high elongation. In fact, for compound leaves, the percentage should be high to eliminate leaflets that may have high local translational invariance. To accomplish this task, we use the ratio between the number of lines that intercepts the shape in two points and the rest. This enables a coarse classification according the organization criteria but it meets our requirements.

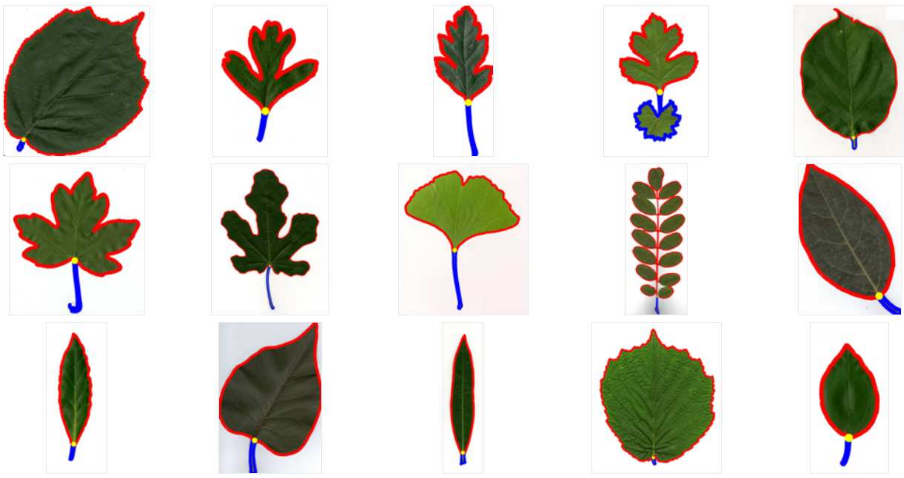

Figure 2. Automatic petiole detection results.

This procedure is proceeded by an alignment step, called alignment according the direction of the petiole. This alignment is defined by the rotation maximizing the number of lines that intercept the shape in two points separated by a small distance, among all possible rotations. The main goal of this alignment is align vertically a long parts of the petiole and then to find a high local translation invariance in this region.

Experiments carried out in the leaf image database of Tela Botanica, our partner in Pl@ntNet project, prove that this approach yields petiole detection results that are accurate for most leaf images and with various petiole structures (simple leaves, compound leaves, leaves with stipules). Figure 2 shows a number of leaves from different species of the Tela databases, where the petiole is well detected by our method. More details can be found in [21].

6.1.3. 3D segmentation

Participants: Wejdi Ben Saad Elweili, Mohamed Chaouch, Anne Verroust-Blondet.

In many applications, it may be helpful to be able to segment a set of 3D objects consistently. It can be a way to consistently decompose the 3D models into parts inside a partial shape retrieval process. For this purpose, we choose to tackle the following problem: we start with a segmentation of a 3D model and propose a method to transfer the segmentation to a set of unsegmented meshes. We obtained promising results on the mesh segmentation benchmark models of Princeton [26] that can be downloaded at http://segeval.cs.princeton.edu/.

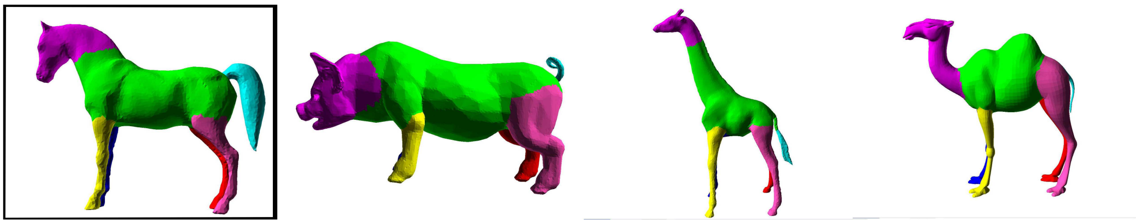

Figure 3. Several "fourlegs" models segmented w.r.t. the segmented horse

Figure 3 shows three models segmented by our method given the initial segmentation of a horse (see [18] for more details).

6.1.4. A new approach for image description based on optical context

Participants: Asma Rejeb Sfar, Hervé Goëau, Nozha Boujemaa.

Nowadays, digital cameras can bring a host of information around a photo-shooting event, and a digital photo is no longer just a collection of pixels. A digital photo brings some relevant source of contextual information stored as metadata. Indeed, besides the date informations, the metadata can give numerous technical optical informations related to the global scene capture conditions like the focal, the aperture, the exposure-time, ISO value, how the flash was used, the white-balance, ..., but also information more related to the subject in the photo like the focus-distance or the auto-focus points localization. All these informations reveal, directly or not, consciously or not, a part of the intention of the photographer during a photo shoot.

In [22], we examine the way of combining such metadata with low-level features in order to improve image retrieval performance and image object categorization. We are especially interested in the autofocus points and in the subject distance information, which were very rarely exploited in previous works, because these informations are more related to a subject in the photo than to the global scene. We first propose an "optical attentive" region localization method based on the active autofocus points as shown in Figure 4 in order to describe an hypothetical main subject in the scene. Secondly, we use the focus-distance information to estimate an order of magnitude of the real dimensions of the local visual contents in the focus plane. These estimations are useful to disambiguate two images containing similar visual contents, but which are in fact associated with dissimilar objects in terms of dimensions in the real world. Finally, we implement a new visual signature, which combines the description of the "optical" region of interest with the "real" dimensions estimated informations.

Results show the potential of this approach in the image retrieval context. We propose to enhance this description method by using other metadata. We also would like to test the relevance of this new approach on image classification.

Figure 4. Attentive region extraction based on automatic focus

6.2. Feature space structuring

6.2.1. Random Maximum Margin Hashing

Participants: Alexis Joly, Olivier Buisson.

Following the success of hashing methods for multidimensional indexing, more and more works are interested in embedding visual feature space in compact hash codes. Such approaches are not an alternative to using index structures but a complementary way to reduce both the memory usage and the distance computation cost. Several data dependent hash functions have notably been proposed to closely fit data distribution and provide better selectivity than usual random projections such as LSH. However, improvements occur only for relatively small hash code sizes up to 64 or 128 bits. We did show in several experiments that this issue is mainly due to the lack of independence between the produced hash functions. In this work, we introduced a new hash function family that attempts to solve this issue in any kernel space. Rather than boosting the collision probability of close points, our method focus on data scattering. By training purely random splits of the data, regardless the closeness of the training samples, it is indeed possible to generate consistently more independent hash functions. On the other side, the use of large margin classifiers allows to maintain good generalization performances. Experiments show that our new Random Maximum Margin Hashing scheme (RMMH) outperforms 4 state-of-the-art hashing methods, notably in kernel spaces.

6.2.2. Multi-source shared nearest neighbours for multi-modal image clustering

Participants: Amel Hamzaoui, Alexis Joly, Nozha Boujemaa.

Unsupervised data clustering remains a crucial step of many recent multimedia retrieval approaches, e.g. web objects and events mining, search results clustering or visual query suggestion.

Shared Nearest Neighbours (SNN) techniques are well known to overcome several shortcomings of traditional clustering approaches, notably high dimensionality and metric limitations. However, previous methods were limited to a single information source whereas such methods appear to be very well suited for heterogeneous data, typically in multi-modal contexts.

We propose a new technique to accelerate the calculation of shared neighbours and we introduce a new **multi-source** shared neighbours scheme applied to multi-modal image clustering. We first extend existing SNN-based similarity measures to the case of multiple sources and we introduce an original automatic source selection step when building candidate clusters. The key point is that each resulting cluster is built with its own optimal subset of modalities which improves the robustness to noisy or outlier information sources.

We experiment our method in the scope of multi-modal search result clustering, visual search mining and subspace clustering. Experimental results on both synthetic and real data involving different information sources and several datasets show the effectiveness of our method. More details can be found in [16], [9].

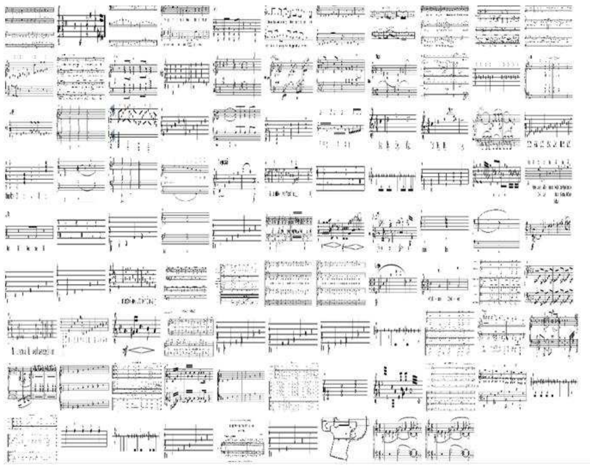

Figure 5. First three Clusters of the Wikipedia's subset clustering using visual and textual sources.

An example of a multi-modal search result clustering is shown in Figure 5. The visual and textual sources are selected for the first cluster while just the textual source is selected for the second cluster and the visual source is selected for the third cluster. This work has been done inside the R2I ANR project.

6.2.3. Visual similarity sensitive hashing methods for semantic image search in very large collections of images

Participants: Saloua Ouertani-Litayem, Alexis Joly, Nozha Boujemaa.

Over the last decade we have witnessed an explosive growth in the scale of shared data collections. Then, it is now possible to use very large training sets to address challenging tasks in machine learning for developing large scale approaches for multimedia retrieval and mining. Therefore, a key challenge is to build efficient methods for training and matching efficiently very large collections of images.

In this context, we first made experiments comparing the effectiveness of the state of the art unsupervised hashing methods used as KNN classifier. We notably integrated the Kernelized Locality Sensitive Hashing method [30] and studied the difference between it as a supervised hashing method and unsupervised methods such as a posteriori multiprobe Locality sensitive Hashing. For the experiments we used the dataset being with 256 categories (Caltech 256) from the Caltech vision group.

We then investigated with an approach consisting in benefiting from both semantic hashing like techniques and embedding approaches in order to build compact category aware codes indexed with efficient similarity search structures. Therefore, we have extended a kernelized hashing method with multi-class SVM to solve a K-class classification problem by choosing the maximum applied to the outputs of K SVMs. An important task during this process was to experimentally evaluate the quality losts induced by such representations with respect to the efficiency gains.

6.2.4. Scalable information retrieval in distributed architectures

Participants: Mohamed Riadh Trad, Alexis Joly, Nozha Boujemaa.

Whereas text search engines are now mature enough to deal with huge datasets, there are still some challenging issues querying large-scale visual datasets. Low-level visual metadata are indeed not simple textual or scalar values, their management requires efficient similarity search in high dimensional spaces.

For the last decade, the problem has been intensively researched on centralized settings. Space partitioning methods perform well on low dimensions but degrade to linear search on high dimensions. On the other hand, hash-based approaches trade accuracy for efficiency by returning approximate closest neighbors of a query point.

However, the amount of data they can handle is still several order of magnitude lower to what would be required in real world networked media applications. Distributing efficiently these methods is thus a challenge for handling complex applications.

We designed and implemented a scalable prototype for distributed similarity search based on state of the art distributed architecture, we have made several experiments querying real world large datasets. The prototype proved to be efficient while processing large queries providing a suitable framework for mining applications.

To achieve experimentations, we implemented an event-oriented retrieval service and achieved large scale experiments on real world dataset. Ongoing experiments process a 1.6 million images dataset. Results will be submitted for publication.

6.3. Interactive retrieval and navigation

6.3.1. Interactive Search with Local Visual Features for Computer Assisted Plant Identification

Participants: Wajih Ouertani, Michel Crucianu, Nozha Boujemaa.

As biological image databases are increasing rapidly, automated species identification based on digital data is of great interest for accelerating biodiversity assessment, research and monitoring. Content-based image search can provide a significant contribution to plant species identification. However, to make it successfully applicable to realistic contexts, we argue that it is necessary to let the user interact with the system on the basis of local image descriptions that allow to focus on the relevant part of an image.

In the scope of the Pl@ntNet project, we put forward an interactive identification approach in which a botanist having a partially annotated large image database is assisted by a Relevance Feedback search mechanism to identify a plant specie. We proposed a relevance feedback method relying on local images features (LF). It jointly uses search by example with local queries and supervised classification (with Support Vector Machines, SVM) [14], [13]. Every Relevance Feedback round consists of two stages: (1) QBVE using as query the LF that were previously found relevant; (2) result re-ranking by the SVM decision function, applied to the potentially relevant set of features in every returned image. This joint use of QBVE and SVM classification serves two purposes. First, it allows to locate, in the returned images, the potential regions of interest that have to be evaluated by the SVM.

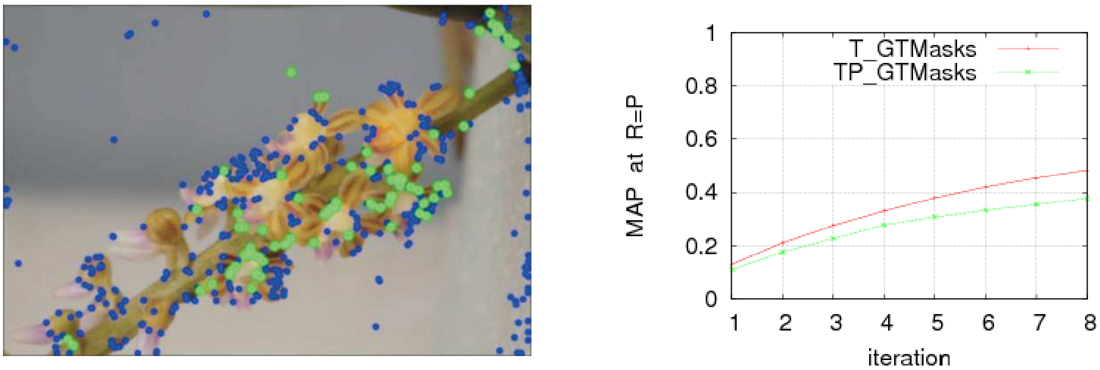

Figure 6. Implicit object localization

In this context, the task of the SVM is to solve the ambiguity and distinguish sets of LF that belong to the target from others judged relevant according to individual features. Second, QBVE can be very fast with an appropriate index structure and only images containing hit points (i.e. points that are individually similar to relevant LF) have to be evaluated by RF rather than all the images in the database, which significantly improves scalability. We have shown that this approach can be successful and that it makes prior segmentation unnecessary (see in Figure 6 right: an exp. performed on Oxford flowers dataset). The results also show how important it is to devise local features that are robust to most of the variations that can be expected when pictures are taken in more general and uncontrolled conditions.
Our tests are performed on :

- AMAP Orchids database : 1937 images , 181 species, with bounding boxes annotated regions of interest.
- Oxford Flowers 17 database : 1360 images, 17 classes , 80 images per class.

6.3.2. Thesaurus construction for interactive search in a satellite images database

Participants: Sahbi Bahroun, Nozha Boujemaa, Ziad Belhadj [SUP'COM, Tunisia].

Observing the Earth from space boomed in recent years with the launch of several satellites at very high spatial resolution (Ikonos Quikbird, Spot 5 ...). The analysis and characterization of images acquired by these sensors require the development of new modeling methods for different classes of the image (urban, forest, vegetation, sea ...) taken in different spectral and spatial resolutions ranging from 1m to 20m. Indeed, an urban class, for example, is viewed differently depending on the spatial resolution, spectral mode, the camera angle ... etc. A Content Based Image Retrieval system should allow the user to explicitly designate the visual query related to the different image components. Our goal was how to define visual search keys to allow the user to express the visual target using efficient descriptors that best represent the content of the image by taking into account all the transformations that an image can suffer.

In [10], [11], we focused on the case where the query image example does not exist or is not appropriate to the mental image that the user wants to submit. The paradigm of the Visual thesaurus is used to generate categories of regions whose represent the visual abstract of the image database as a zero page. The user can compose his mental image by selecting visual thesauruses from this page zero. Indeed, the content-based image retrieval from visual thesaurus is an effective and realistic alternative when the user is in a situation where the query image is not available (especially in the case of satellite images). We introduced in this work a new unsupervised method that allows distinguishing between two categories of regions: homogeneous and textured. These two regions are very different in nature. Using the same descriptor to describe images will not give us the best description of the data.

This category based description allows a more faithful description of the image content and will enhance the system performance. We have thus constructed two visual thesauruses, one visual thesaurus for homogeneous regions and another visual thesaurus for textured regions. The user can compose his query by selecting regions from these two visual thesauruses.

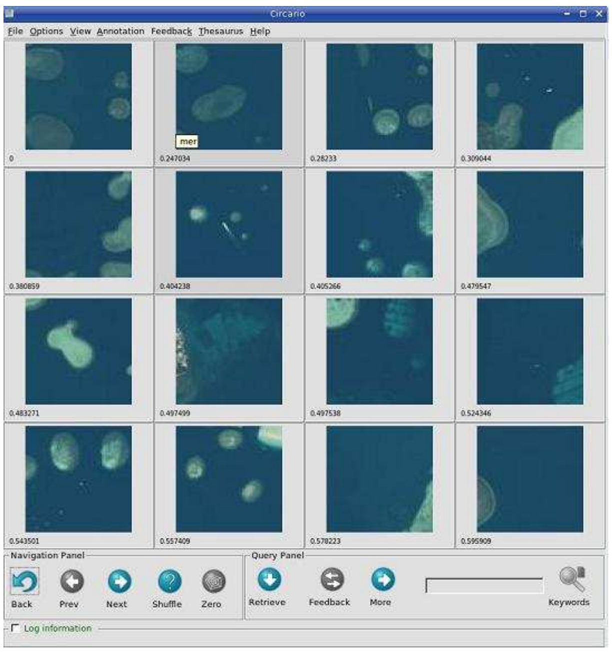

Figure 7. Example of a multi-modal query.

Today, satellite image retrieval systems are not efficient with respect to the semantic search. One reason for this comes from how the images were semantically described. The content of an image can be described at two levels: the numeric level and semantic level. During this work, we tried to find connections between these two levels of description. We believe that these two pieces of information are complementary. Describing images by exploiting these two informations increases the efficiency of the system and improve the quality of results returned at the earliest (see Figure 7 where the query is based on visual feature and the keyword sea).

6.3.3. Scene Pathfinder: Unsupervised Clustering Techniques for Movie Scenes Extraction

Participants: Mehdi Ellouze, Nozha Boujemaa, Adel Alimi [ENIS, Tunisia].

The need of summarization methods and systems becomes more and more crucial as the audio-visual material continue the exponential growth. Our work [8], [7] proposes a novel vision and a novel system for movies summarization. A video summary is an audio-visual document displaying the essential parts of the original document. However, the definition of the term "essential" is user-dependent. The advantage of this approach relatively to others [33], [31], [34], [27] is the involve of users in the summarization process. By means of $IM(S)^2$, people generate on the fly customized video summaries responding to their preferences. $IM(S)^2$, is made of an offline part and an online part. In the offline part we segment the movies into shots and we compute features describing them. In the online part users inform about their preferences by selecting interesting shots. After that, the system will analyze the selected shots to bring out the user's preferences. Finally the system will generate a summary from the whole movie which will provide more focus on the user's preferences. To show the efficiency of $IM(S)^2$, it was tested on the database of the European project MUSCLE made up of 5 movies. We invited 10 users to evaluate the usability of our system by generating for every movie of the database a semi-supervised summary and to judge at the end its quality. Obtained results are encouraging (compared to [32]) and show the merits of our approach.

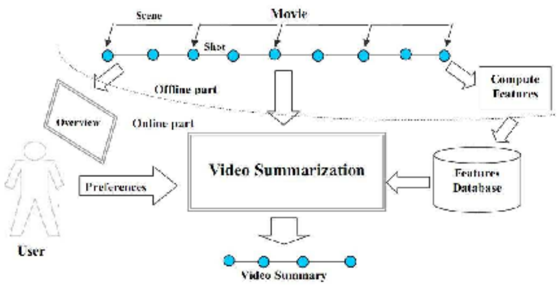

Figure 8. Overview of $IM(S)^2$ system

Figure 8, illustrates our framework in a synthetic way. It may be summarized as follows: the user will take a look on an overview, interact with it to express his preferences and the system replies by generating a summary which at the same time displays the essential parts of the original movie and includes the user's preferences

6.3.4. Visualisation of a collection of images

Participants: Boujemaa Dahoui, Anne Verroust-Blondet.

In the context of Pl@ntNet project, it is important to propose a visualisation tool to provide to the user visual representations of the whole database. This year, we developed and compared several techniques computing a 2D representation of the content of a database of botanic images such that the similarity distances between the images are preserved. The similarity matrices may concern either the visual features or the textual annotation of the of the images. Four methods of mapping : Multidimensional Scaling (MDS), isometric mapping (ISOMAP), stochastic neighbour embedding (SNE) and isometric stochastic neighbor embedding (ISOSNE). Experimental results on different databases of botanic images helped us to mesure the efficiency and the effectiveness of the visual descriptors and the mapping techniques (cf. [19]).

6.3.5. Building interpretable visual models for interactive object retrieval

Participants: Ahmed Rebai, Alexis Joly, Nozha Boujemaa.

Thanks to the advances in technology, content-based image retrieval gained further maturity during the last few years. There were indeed several improvements in terms of image description and large scale real-time efficient search. However, current search engines haven't achieved yet the point to which they correctly formulate and answer a mental user query. This is the case for example when we are tempted to retrieve images that contain an object with specific characteristics. In fact, in our natural language, we use the same word to define similar objects having the same function even if they differ visually. Moreover, what a user needs may differ from one person to another. It is then of interest to build systems that retrieve these concepts according to our perception. One solution lies in constructing an interactive system that allows users to define their own visual concept. The main problem, however, consists in how to provide concise models, and at the same time, visually interpretable. Our idea relies on two points: first, we propose to use local features that are easily understandable by humans, and second, we suggest to regularize the loss function by adding extra constraints. We believe that these two clues represent major keys to favor user interactivity.

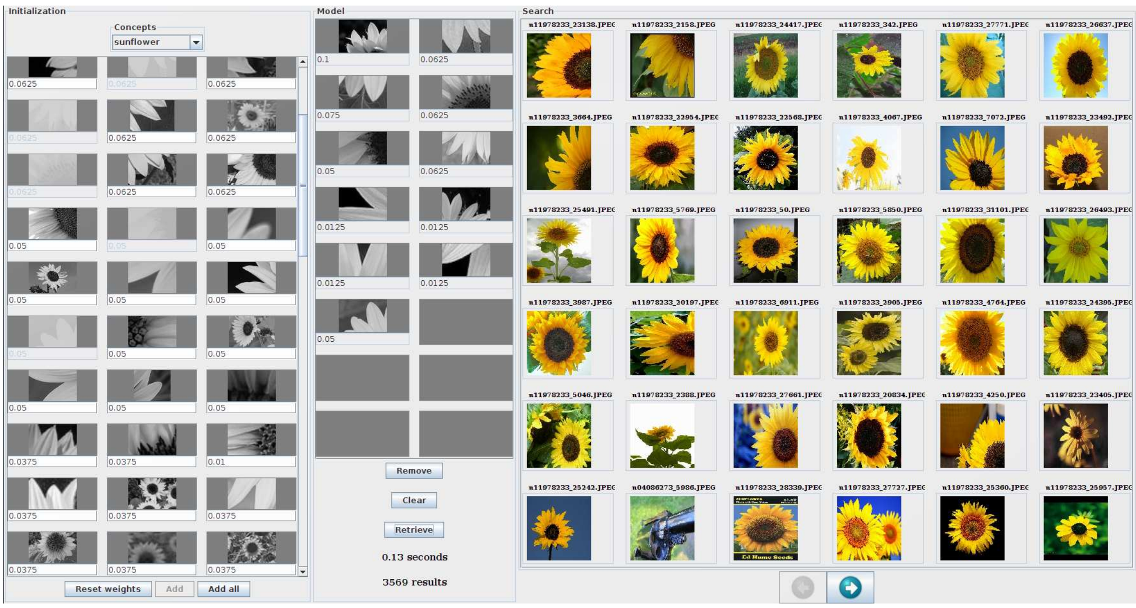

Figure 9. Example of an interactive search

Relying on interpretable local features is considered to be one first step to better perceive and understand what the model is composed of. Added to that, using a sparse representation should normally increase the knowledge of the most statistically discriminant features that summarize a given category. Sparsity is also preferable because it helps to decrease prediction time. In order to obtain sparse solutions, we need to constraint the model. We propose to build our models based on solutions of type lasso (Least Absolute Shrinkage and Selection Operator). It introduces a L_1-regularization constraint to allow for shrinking the models coefficients and producing more interpretable visual models. For a general convex loss function L, the lasso loss Γ can be written as:

$$\Gamma(\beta,\lambda) = \sum_{n=1}^{N} L(S_n,\beta) + \lambda \cdot ||\beta||_1$$

where $\lambda \geq 0$ is the parameter controlling the amount of regularization applied to the estimate. During the last year, we carried out experiments using various databases and we noticed an improvement over both precision and prediction time. Our results were submitted to a journal for publication. Furthermore, we designed a GUI for interactive retrieval (cf. Figure 9). After choosing a visual concept, users are allowed to build a visual model by collecting the patches they find relevant and attributing a confidence measure to each of them. The search engine can be interrogated afterward.

6.4. Software

6.4.1. IKONA/MAESTRO software

Participants: Jean-Paul Chieze, Mondher Khadhraoui, Souheil Selmi, Alexis Joly, Nozha Boujemaa.

This year, IKONA has been extended in the context of Pl@ntNet, R2I projects and Exealead collaboration. We choose the web services approach to interface IKONA/maestro in order to make its fonctionnalities easily accessible from the outside. Web services based on SOAP protocol (visual similarity search, external search, Relevance feedback, and Clustering) in the contect of Exalead demonstrator have been developed. In addition, the indexation of an Exalead dataset which contains 100 millions images and web services based on REST (visual similarity search and KNN classfier) in the Pl@ntNet project are currently processed.

6.4.2. PMH library

Participants: Alexis Joly, Olivier Buisson [INA].

PMH library being a joint ownership between INRIA and INA, it is not a part of IKONA, owned by INRIA only. The library is usable either as a stand alone software or as a pluggin in IKONA for large scale visual information retrieval. PMH is the core technology of most scalability issues addressed by VITALAS European project and is integrated as a service of VITALAS multimedia search engine (pre-industrial prototype). It is therefore used by all VITALAS users including the academic partners of the project (CWI, Fraunhofer, CERTH-ITI, etc.), the end-users of the project (INA, Belga) and the external users of the system (academics, companies trials, general public, etc.). Discussions with INA concerning the valorization of the library as a stand alone software are currently progressing. INRIA suggested an open source diffusion of the software to emphasize visibility. INA internal discussions are still pending and we hope converging to a final decision before the end of 2011.

7. Other Grants and Activities

7.1. National Initiatives

7.1.1. Pl@ntNet project [2009-2012]

It is a joint project with AMAP (CIRAD, Montpellier) and Tela Botanica, an international botanical network with 8,500 members and an active collaborative web platform (10,000 visits /day). The project has its financial support from Agropolis International Foundation (http://www.agropolis.fr/) and is titled “Plant Computational Identification and Collaborative Information System”.

Dissemination:

- This year, we developed jointly with AMAP and Telabotanica an automatic tree identification demonstrator called pl@ntScan(http://imedia-ftp.inria.fr:8080/demo_plantscan_new/), which is freely available as an online service.

7.1.2. Other collaborations with INRA

The PhD thesis of Wajih Ouertani, financed by INRA, in the context of a strategic collaboration between INRIA and INRA, addresses interactive species identification through advanced relevance feedback mechanisms based on local image information.

7.1.3. ANR project R2I [2008-2010]

The project "R2I - Recherche Interactive d'Images" is a joint project which aims at designing new methods for interactive image search. The final goal of this project is a system which can index about one billion of images and provide users with advanced interaction capabilities. The partners are the company Exalead, a leader in the area of corporate network indexing and a specialist for user-centered approaches, the INRIA project-team Imedia, a research group with a strong background in interactive search of multi-media documents, as well as LEAR and the University of Caen, both specialists in object recognition.

7.1.4. ANR project SCAR-FACE [2008-2010]

SCAR-FACE (*Semantic Characterization And Retrieval of FACEs*) objective is to develop new interactive technologies for recognizing people in public places provided with videosurveillance networks.
Other partners: Univ Caen - INRIA LEAR, EADS, SPIKENET, IREENAT
IMEDIA activities within the project are just starting.

7.2. European Initiatives

7.2.1. FP7 - European Coordination Action "CHORUS+" http://www.ist-chorus.org/ [2010-2012]

CHORUS + has been funded in the continuity of the former CHORUS initiative thanks to its success. Beyond CHORUS coordination objectives, CHORUS+ includes new key issues such as extended cooperation and coordination to Asian countries and US, support to integration and implementation, support to coordinated research evaluations or support to results dissemination of EU projects. Nozha Boujemaa and Alexis Joly are part of the management board of the project.

7.2.2. FP7 - European Integrated Project "GLOCAL" [2010-2012]

Glocal (*Event-Based Retrieval of Networked Media*) is a European Integrated Project involving 10 accademic and industrial partners (UNITN, ISOCO, ALINARI,CERTH, YahooSpain, AFP, DFKI, Exalead, LUH). The key idea underlying the project is to use events as the primary means for organizing and indexing media. Within networked communities, common (global) descriptions of the world can be built and continuously enriched by a continuous flow of individual (local) descriptions. With two leading search companies and four content providers, the consortium attempts to realize and evaluate this approach in several application domains, which will involve professional and amateur users dealing with professional and generic contents. IMEDIA is responsible of three research tasks related to visual-based event indexing, retrieval and mining, notably in distributed contexts.

7.2.3. FP7 - STREP "I-SEARCH" [2010-2012]

The I-SEARCH project (*A unified framework for multimodal content SEARCH*) aims to provide a novel unified framework for multimodal content indexing, sharing, search and retrieval. It involves academic and industrial partners (CERTH, JCPC, ATTC, ENG, Google, UNIGE, Exalead, FHE, ANSC, EGR). The I-SEARCH framework will be able to handle specific types of multimedia and multimodal content (text, 2D image, sketch, video, 3D objects and audio) alongside with real world information, which can be used as queries and retrieve any available relevant content of any of the aforementioned types. IMEDIA is workpackage leader of "RUCOD COMPLIANT Descriptor Extraction".

7.3. International Initiatives

7.3.1. Cooperation with John Hopkins University, USA

Don Geman is a regular visiting professor since several years; The scientific topics addressed are related to relevance feedback and mental category image search.

7.3.2. Cooperation with Tunisia: CIVE project

The CIVE (Classification d'Images d'espèces VEgétales) project is a collaborative project between AMAP, INRIA, ISI (Institut Supérieur d'Informatique de Tunisie) and Sup'Com (Ecole Supérieure de Communication de Tunis). It is financed by both the Tunisian Universities and INRIA.

7.3.3. MUSCLE-VCD corpora for benchmarking

In 2007, IMEDIA did organise the first international benchmark event about video copy detection technologies, as a "live" event during ACM CIVR 2007 conference (http://www-rocq.inria.fr/imedia/civr-bench/benchMuscle.html). This corpus is still maintained and distributed with success.

7.3.4. BelgaLogo dataset

BelgaLogo dataset is an evaluation corpus and framework dedicated to logos and trademarks retrieval in large pictures collections (the only one to the best of our knowledge). It was created in 2009 within VITALAS European IP in collaboration with Belga Press Agency. It is freely distributed on the web [2] (about 20 official download requests for now).

8. Dissemination

8.1. Seminars, presentations and other dissemination activities

8.1.1. Demos

Demos of IKONA/MAESTRO software on botanic databases and of pl@ntScan have been presented at "Journée Moteur de recherche : de l'accès à la maîtrise des contenus" organised by INA Sup, INRIA and Cap Digital at L'atelier BNP Paribas, Paris, in November 2010.

A Demo of our 3D shape retrieval system has been presented at the "journée nationale des Rencontres INRIA Industrie" on "Modélisation et systèmes sûrs" at Toulouse in May 2010.

8.1.2. Dissemination to large public and European community

The program of the CIVR 2010 Practitioner's Chairs has been set-up in collaboration with the EU Coordinated Action CHORUS+

8.2. Leadership with scientific community

8.2.1. Nozha Boujemaa

- General co-Chair of ACM Multimedia Information Retrieval (ACM MIR 2010 - March 29-31 - Philadelphia, Pennsylvania): http://riemann.ist.psu.edu/mir2010/
- Co-chair of "Track V: Multimedia and Document Analysis, Processing and Retrieval" in ICPR 2010: International Conference on Pattern Recognition 2010 (23-26 August Istanbul), http://www.icpr2010.org/ and http://www.icpr2010.org/tracks.php#Multimedia_and_Document_Analysis_Processing_and_Retrieval
- Chair of "Brave New Ideas" in ACM Multimedia 2010 25-29 October 2010, Florence, Italy (http://www.acmmm10.org/)
- Founding member of the ACM ICMR "ACM International Conference on Multimedia Retrieval" born from the fusion of: ACM MIR (International Conference on Multimedia Information Retrieval) and ACM CIVR (International Conference on Image and Video Retrieval)
- Member of the steering committee of ACM ICMR (4 years)

[2] http://www-rocq.inria.fr/imedia/belga-logo.html

- Program Chair of final Chorus conference http://www.ist-chorus.org/conference.asp
- Member of the Scientific Advisory board of the japanese project "Multimedia Web Social Analysis and Mining" supported by the MEXT (Japonese ministry of Research)
- Member of an academia Think-Tank for the PPP european initiative.
- Scientific coordinator of CHORUS+ CA FP7;
- Expert for ESF (European Science Foundation : http://www.esf.org), appointed since 2008
- Expert for the EC for FP7 preparation, participation to several expert meetings.
- Expert for NWO (Netherland)
- Elected member in the Steering Board of NEM ETP (Networked and Electronic Media European Technology Platform) and acting as INRIA representative
- French expert for COST ICT Domain (intergovernmental network for European Cooperation in the field of Scientific and Technical Research)
- Member of ACM - SIGMM committee and of ACM Multimedia Information Retrieval International Conference steering 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, IEEE ICPR, CBMI, SAMT, WIAMIS...
- Responsabilities within INRIA: Director of INRIA Research Center at Saclay (officially started in September 2010); member of COPIL SDRH of INRIA (comité de pilotage national sur les priorités et la prospective de la politique des ressources humaines), member of the “Comité d’animation scientifique” of the research topic "Perception, cognition, interaction" of INRIA (2010 till her new appointment as Director of INRIA Research Center at Saclay).

8.2.2. Michel Crucianu

- Co-organizer of the ISIS-I3 Workshop “Scalability in multimedia retrieval and mining” (http://gdr-isis.fr/index.php?page=reunion&idreunion=111), November 3, 2010, CNAM, Paris.
- Invited talk “Scalability in multimedia content retrieval and mining”, during the Assises of the GDR Information - Interaction - Intelligence (I3), July 2, 2010, Strasbourg.
- Scientific expert for the French National Research Agency (ANR), call “Programme Blanc”, and for ANCS (Romania).
- Journal reviewer: Decision Support Systems, IEEE Transactions on Neural Networks, Information Science, Pattern Recognition Letters.

8.2.3. Alexis Joly

- Member of the steering committee of Pl@ntNet: WP2 co-leader.
- Program committee member of 2010 NEM Summit, Panel chair at CBMI 2010.

8.2.4. Anne Verroust-Blondet

- Member of the steering committee of Pl@ntNet (WP5 co-leader: “Dissemination”).
- Member of the steering committee of I-SEARCH (WP4 leader).
- Member of the Humanities and Social Sciences committee for the “Blanc” and “Young researcher” 2010 programmes of the French National Research Agency (ANR) and expert for the “Equipex 2010” call for ANR.
- Member of the steering committee of the CNRS GDR IG (Informatique Graphique) ;

- Member of the technical programme committee of the Eurographics Workshop on 3D Object Retrieval in 2010, of the ACM Workshop - 3D Object Retrieval 2010 and of the Third International Conference on Advances in Multimedia (MMEDIA 2011),
- Member of the editorial board of the “Revue Electronique Francophone d’Informatique Graphique”.

8.3. Teaching

8.3.1. Nozha Boujemaa

- 20 hours course on multimedia indexing at ISI and SupCom Tunis.

8.3.2. Michel Crucianu

In charge of the courses

- “Machine learning”, Master in computer science and School of engineering of CNAM Paris.
- “Advanced databases 2 : image databases”, Master in computer science and School of engineering of CNAM Paris.
- “Multimedia Databases”, Master in computer science of the University Paris Dauphine.

8.3.3. Amel Hamzaoui

- 20 hours of supervised work on C++ for first degree students at Orsay Paris-Sud University.

8.3.4. Mohamed Riadh Trad

- 32 hours course on Algorithmic courses first degree students at Paris Dauphine University

8.3.5. Itheri Yahiaoui

- 192 hours in the Mathematic and Computer Science Departement of Reims Champagne Ardenne University;
- In charge of the course "Images Acquisition and Analyses" of the Master "engineering, images and knowlegle" of Reims Champagne Ardenne University.

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", 2007, vol. 29, n^{o} 10, p. 1818 – 1833.

[2] M. CHAOUCH, A. VERROUST-BLONDET. *Alignment of 3D models*, in "Graphical Models", March 2009, vol. 71, n^{o} 2, p. 63–76.

[3] 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.

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

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

[6] N. GRIRA, M. CRUCIANU, N. BOUJEMAA. *Active semi-supervised fuzzy clustering*, in "Pattern Recognition", 2008, vol. 41, n° 5, p. 1834-1844.

Publications of the year

Articles in International Peer-Reviewed Journal

[7] M. ELLOUZE, N. BOUJEMAA, A. M. ALIMI. *IM(S)^2: Interactive movie summarization system*, in "Journal of Visual Communication and Image Representation", 2010, vol. 21, n° 4, p. 283-294.

[8] M. ELLOUZE, N. BOUJEMAA, A. M. ALIMI. *Scene Pathfinder: Unsupervised Clustering Techniques for Movie Scenes Extraction*, in "Multimedia Tools and Applications", 2010, vol. 47, n° 2, p. 325-346.

[9] A. HAMZAOUI, A. JOLY, N. BOUJEMAA. *Multi-source shared nearest neighbours for multi-modal image clustering*, in "accepted for publication in in Multimedia Tools and Applications", 2010.

International Peer-Reviewed Conference/Proceedings

[10] S. BAHROUN, N. BOUJEMAA, Z. BELHADJ. *Combining textual and visual thesaurus for a multi-modal search in a satellite image database*, in "IEEE Geoscience and Remote Sensing Symposium (IGARSS 2010)", Honolulu, Hawai, July 2010.

[11] S. BAHROUN, N. BOUJEMAA, Z. BELHADJ. *Hierarchical visual thesaurus building for satellite image retrieval based on semantic region labelling*, in "IEEE International Conference on Image Processing (ICIP 2010)", Hong-Kong, September 2010.

[12] R. ENFICIAUD. *Queue and priority queue based algorithms for computing the quasi-distance transform*, in "International Conference on Image Analysis and Recognition (ICIAR 2010),", Povoa de Varzim, Portugal, June 2010.

[13] W. OUERTANI, P. BONNET, M. CRUCIANU, N. BOUJEMAA, D. BARTHÉLÉMY. *Iterative Search with Local Visual Features for Computer Assisted Plant Identification*, in "BioIdentify 2010", Paris, September 2010.

[14] W. OUERTANI, M. CRUCIANU, N. BOUJEMAA. *Interactive Learning of Heterogeneous Visual Concepts with Local Features*, in "ACM Multimedia (MM 2010)", Florence, Italy, October 2010.

Scientific Books (or Scientific Book chapters)

[15] N. BOUJEMAA, H. GOURAUD, R. COMPANO, J. KARLGREN, P. VAN DER LINDEN, P. KING, N. SEBE, J. KOHLER, A. JOLY, J. GEURTS, C. DOSCH, R. ORTGIES, A. RUDSTROM, M. KAUBER, J.-C. POINT, J.-Y. LE MOINE. *Cross-disciplinary Challenges and Recommendations regarding the Future of Multimedia Search Engines*, Publications Office of the European Union, 2010, http://cordis.europa.eu/fp7/ict/netmedia/docs/publications/multimedia-search.pdf.

Research Reports

[16] A. HAMZAOUI, A. JOLY, N. BOUJEMAA. *Multi-source shared nearest neighbours for multi-modal image clustering*, INRIA, 2010, n^o RR-7351, http://hal.archives-ouvertes.fr/inria-00496170/en/..

[17] A. JOLY. *Technical report on large scale indexing and simlarity search*, R2I european project, August 2010.

Other Publications

[18] W. BEN SAAD ELWEILI. *Segmentation d'objets 3D interactive et par l'exemple*, Ecole Polytechnique de Tunisie (Tunisia), June 2010.

[19] B. DAHOUI. *Méthodes de visualisation d'une collection d'images*, Ecole Polytechnique de Tunisie (Tunisia), June 2010.

[20] S. MOUINE. *Application de la morphologie mathématique pour la caractérisation de l'architecture des feuilles de vigne*, Institut Supérieur d'Informatique de Tunis, September 2010.

[21] O. MZOUGHI. *Morphological analysis of leaf shape: Application to tree leaves classification*, Higher School of Communications of Tunis (SUPCOM), 2010.

[22] A. REJEB SFAR. *Vers une nouvelle description du contenu visuel basée sur le contexte optique des images*, ENSEEIHT (Toulouse), September 2010.

References in notes

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

[24] 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.

[25] M. CHAOUCH, A. VERROUST-BLONDET. *Alignment of 3D models*, in "Graphical Models", 2009, vol. 71, p. 63-76.

[26] X. CHEN, A. GOLOVINSKIY, T. FUNKHOUSER. *A Benchmark for 3D Mesh Segmentation*, in "ACM Transactions on Graphics (Proc. SIGGRAPH)", August 2009, vol. 28, n^o 3.

[27] M. ELLOUZE, H. KARRAY, A. M. ALIMI. *Genetic algorithm for summarizing news stories*, in "International Conference on Computer Vision Theory and Applications, VISAPP", 2007, p. 303-308.

[28] F. FLEURET. *Détection hiérarchique de visages par apprentissage statistique*, Université Paris-VI, Paris, 2000.

[29] D. J. HEARN. *Shape analysis for the automated identification of plants from images of leaves*, in "Taxon", 2009, p. 934-954.

[30] B. KULIS, K. GRAUMAN. *Kernelized locality-sensitive hashing for scalable image search*, in "IEEE International Conference on Computer Vision (ICCV", 2009.

[31] Y. MA, L. LU, H. ZHANG, M. LI. *A user attention model for video summarization*, in "ACM International Conference on Multimedia, ACM MM", 2002, p. 533-542.

[32] V. PARSHIN, L. CHEN. *Video Summarization Based on User-Defined Constraints and Preferences*, in "International Conference on Adaptivity, Personalization and Fusion of Heterogeneous Information, RIAO", 2004, p. 18-24.

[33] M. SMITH. *Video Skimming and Characterization through the Combination of Image and Language Understanding Techniques*, in "IEEE International Conference on Computer Vision and Pattern Recognition, CVPR", 1997, p. 775-781.

[34] S. UCHIHASHI, J. FOOTE, A. GIRGENSOHN, J. BORECZKY. *Video manga: Generating semantically meaningful video summaries*, in "ACM International Conference on Multimedia, ACM MM", ACM Press, 1999, p. p p.383-392.

[35] T. WALTER. *Application de la Morphologie Mathématique au diagnostic de la Rétinopathie Diabétique à partir d'images couleur*, Centre de Morphologie Mathématique, Ecole Nationale Supérieure des Mines de Paris,, 2003.

[36] X. -F. WANGA, D. -S. HUANGA, J. -X. DUA, H. XUA, L. HEUTTED. *Classification of plant leaf images with complicated background*, in "Applied Mathematics and Computation", 2008, vol. 205, p. 916-926.