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

2008

Table of contents

1. Team

Research Scientist
Nozha Boujemaa [Research Director (DR), HdR]
Alexis Joly [Research Associate (CR) Inria]
Anne Verroust-Blondet [Research Associate (CR) Inria, HdR]

External Collaborator
Olivier Buisson [Research Scientist at Institut National de l'Audiovisuel]
Michel Crucianu [Professor at CNAM, HdR]
Itheri Yahiaoui [Assistant Professor at Reims University]

Technical Staff
Jean-Paul Chièze [Senior Technical Staff Inria (half-time)]
Laurent Joyeux [Expert engineer Inria until July 31 2008]
Mehdi Bouabta [Junior Technical Staff Inria until August 31 2008]
Mathieu Coutaud [Junior Technical Staff Inria until September 30 2008]

PhD Student
Sahbi Bahroun [Joint tutorship with Sup'Com, national grant since September 1st 2006]
Olfa Besbes-Abdelhak [Joint tutorship with Sup'Com, national grant since September 1st 2004]
Mohamed Chaouch [INRIA grant, Télécom Paris since October 1st 2005]
Amel Hamzaoui [INRIA grant, Paris-Sud University since October 1st 2008]
Nicolas Hervé [INRIA grant, Paris-Sud University since December 1st 2005]
Wajih Ouertani [INRA grant, Paris-Sud University since October 1st 2008]
Ahmed Rebaï [INRIA grant, Paris-Sud University since September 1st 2007]

Post-Doctoral Fellow
Jaume Amores [Postdoctoral fellow until February 29 2008]
Raffi Enficiaud [Expert engineer Inria since June 1st 2008]
Joost Geurts [Postdoctoral fellow]
Hichem Houissa [Expert engineer Inria]
Julien Law To [Expert engineer Inria until August 31 2008]

Visiting Scientist
Donald Geman [Professor at Johns Hopkins University, USA (May and June 2008), HdR]
Michael Houle [Research scientist at NII, Japan (July and October 2008), HdR]

Administrative Assistant
Laurence Bourcier [Secretary (SAR) 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.

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

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

2.2. Highlights of the year

- **Finalisation of prospective study on cross disciplinary search engine challenges** Coordination of European efforts to produce gap analysis technical challenges, user studies, socio-economic and legal aspects on Multimedia Search Engines in the context of CHORUS project;
- **TRECVID 08** Organisation of the first TRECVID (international evaluation campaign) copy detection track;
- **Alignment of 3D models.** It is an important challenge for many applications in computer graphics, including object matching, retrieval, recognition, and clustering. A new alignment method, based on symmetries, has been selected to appear in the International Journal of Graphical Models (GMOD). Our approach produces alignments similar to what a human would select and improves the performances of 3D shape retrieval systems.

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

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 [24] [29].

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 detection*, *object retrieval*.

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: Donald Geman, Nozha Boujemaa, Hichem Houissa, Raffi Enficiaud, 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.

3.5. Cross-media indexing and retrieval

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

Participants: Nicolas Hervé, 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).

- **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 is 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

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, Raffi Enficiaud, 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 [25]. 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 [25];
- **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/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.

The architecture of this client/server software and several visual signatures were a subject of a deposit to APP. It is distributed to INA, AFP, INRA, Ministry of Interior, JRC and Alinari.

6. New Results

6.1. Construction and organisation of the visual feature space

6.1.1. Multisource RSC clustering for multiple search results clustering

Keywords: *clustering*, *multisource*, *search result*, *shared neighbours*.

Participants: Amel Hamzaoui, Alexis Joly, Nozha Boujemaa, Michael Houle.

Todays Multimedia search engines are often based on multiple decentralised search services, multiple informations sources (text search, audio search, visual search, semantic search engine, etc.) and multiple data representations and similarity measures.Heterogeneous multiple search results need to be combined and structured efficiently and generically. we propose to consider all information sources as simple oracles returning ranked lists of relevant objects and to use an efficient shared neighbours based clustering algorithm called RSC "Relevant Set Correlation" [37]. We propose a multisource version of the RSC algorithm and we compare it with two others fusion strategies (Early fusion, Late fusion). Experimentally, we have shown multi source RSC clustering performs better than early fusion without any knowledge on data. Our approach has Linear complexity in number of oracles. we have compared also an approximate similarity search structure to a sequential method to compute ranked lists of relevant objects and we have concluded that approximate similarity search does not degrade too much clustering performances. We plan to investigate more heterogeneous oracles (text/image,matching scores, etc.) , to study the effects of outliers and to compare our approach to others clustering methods. For more details see [31].

6.1.2. A posteriori Multi-Probe Locality Sensitive Hashing

Keywords: *LSH*, *Multi-Probe*, *complexity*, *indexing*, *similarity search*, *structure*.

Participants: Alexis Joly, Olivier Buisson.

Today's scalability issues already put brake on growth of multi-media search engines. The searchable space created by the massive amounts of existing video and multimedia files greatly exceeds the area searched by today's major engines. To process the massive scale of new data created every day with more and more complex technologies, breaking algorithms complexity remains a key solution. In this joint work with INA, we were interested in reducing both space and time complexity of similarity search algorithms, by proposing a new high dimensional similarity search structure, which improves upon recent theoretical work on the popular Locality Sensitive Hashing (LSH) technique. To overcome the over-linear space cost drawback of common LSH, we define a new multi-probe scheme allowing to probe multiple buckets in each hash table and thus to reduce drastically the number of required hash tables. Our method is based on an a posteriori model taking account some prior about the queries and the searched objects. Compared to other multi-probe schemes, this prior knowledge allows a better quality control of the search and a more accurate selection of the most probable buckets. We show that our a posteriori scheme outperforms other multi-probe LSH while offering a better quality control. Comparisons to the basic LSH technique show that our method allows consistent improvements both in space and time efficiency. Figure 1 plots the search time of our technique compared to LSH when varying the size of a visual local features dataset. It shows that at constant available memory, the increase of LSH search speed over dataset size is supra-linear whereas the one of our method is sub-linear, leading to strong search time improvements. Another advantage of our method is its genericity, notably compared to LSH restrictions on query types: Our probabilistic filtering algorithm is indeed fully independent of the query type. It just requires query samples and corresponding relevant objects sets, not necessarily nearest neighbours. Examples of other relevant objects are distorted features obtained after transformation of a multimedia content or nearest neighbours of the query in another dataset (e.g. a category specific dataset or a training dataset). The search can also be easily adapted to different objectives by pre-computing different prior models and corresponding probabilities Look-up tables for the same index structure. A typical application is to achieve class dependent queries. This work was published in the proceedings of ACM multimedia conference 2008 [17] (oral presentation).

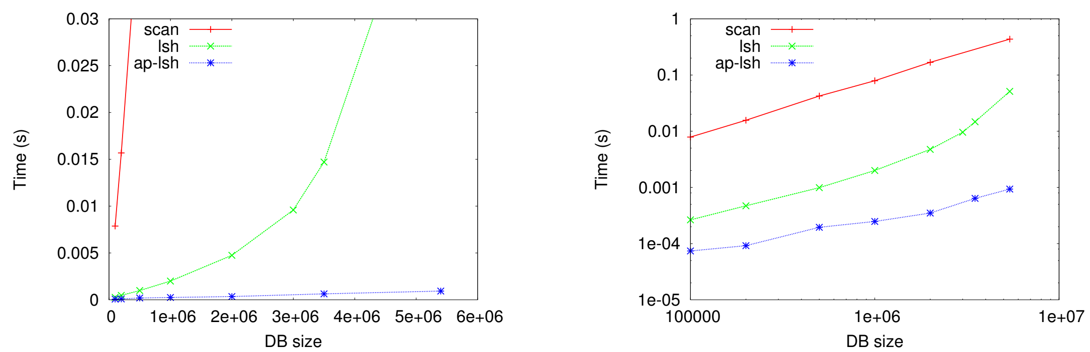

Figure 1. Search time efficiency comparison when varying the size of the dataset - the bottom graph represents the same curves in logarithmic coordinates

6.1.3. 3D indexing: alignment of 3D models

Keywords: *3D alignment*, *3D model retrieval*, *Principal Component Analysis*, *symmetry detection*, *viewpoint selection*.

Participants: Mohamed Chaouch, Anne Verroust-Blondet, Skander El Fekih.

The 2D/3D retrieval approaches developed the previous years ([26], [27]) need a normalisation step as most of the 3D shape retrieval methods. In fact, 3D models are generally given in arbitrary scale, position and orientation in 3D-space. The normalisation consists in two steps: the alignment to determine the pose invariant and the scaling to make the scale-invariant. Generally, an approach based on the Principle Component Analysis method (PCA) is used to compute the reference axes. We have introduced and developed a new approach, based on the symmetry properties of the model: the reflective symmetry and the local translational symmetries along a direction. For that purpose, a local translational invariance cost (LTIC) have been defined. It measures the local translational symmetries of a shape along a given direction. The good properties of the PCA methods w.r.t. the mirror symmetries of a 3D model transforms the detection of the reflective symmetries into a selection of some PCA-axes. This lead to an efficient alignment method which consistently aligns most of the 3D objects (see Figure 2).

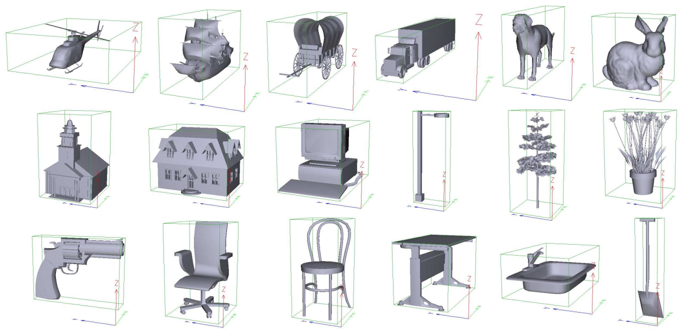

Figure 2. Alignments results of models of different classes.

Moreover, our alignment improves the retrieval performances, when used inside a normalisation step in a 3D retrieval process. This work is described in [16].

To help the user when consulting the query results, a new interface has been built. A 3D viewer, where the user can rotate the model, fix the associated reference frame and select the best viewpoint, is associated to each resulting model (cf. Figure 3 and [28] for more details).

Figure 3. Left: Our new interface with aligned models and viewpoints selected by the user. Right: a 3D viewer associated to a model

6.1.4. Stochastic image segmentation by using multi-cue combination and Conditional Random Field framework

Keywords: *contextual interactions*, *homogeneous/non-homogeneous CRF model*, *image segmentation*, *multi-cue combination*.

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

Image segmentation is a long standing problem in computer vision and is the basis of many visual tasks. For instance, the extraction of coherent regions in images is very useful for object recognition, region-based retrieval and semantic scene interpretation of medical and satellite images. Despite many thoughtful attempts, it still remains a challenging task due to the complexity of natural images. This complexity may arise because those images are heterogeneous and contain both non-textured and textured regions such that within them pixels are spatially dependent and follow various perceptual organisation rules. Thus, it is useful to combine cues such as edges, colour and texture to discriminate in a given image its constituent regions. Therefore, we define in [15] the objective function in terms of region and boundary properties of segments and according to some Gestalt principles such as similarity, proximity and good continuation. We formulate the problem as partitioning an hybrid adjacency graph of atomic region and atomic contour primitives. We define directly the posterior probability $P(R, B|F)$, of region and boundary partitions (R, B) given observations F, by using a Conditional Random Field (CRF) framework [32]. 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. In order to capture contextual interactions of the labels as well as the labels, we propose in [14] a non-homogeneous CRF model with spatially dependent association and pairwise interaction potentials. Both atomic regions and segments are described with contextual statistics. In our energy formulation [15], [14], similarity is measured by learned likelihood ratios so that we reduce the number of model parameters to few weighting parameters. The inference is performed using a cluster sampling method, the Swendsen-Wang Cut algorithm [23] to combine the representational advantages of CRF and graph cut approaches. Although we don't use region description parameters since we measure similarity by the learned likelihood ratios, we obtain promising results as shown in figures 4 and 5.

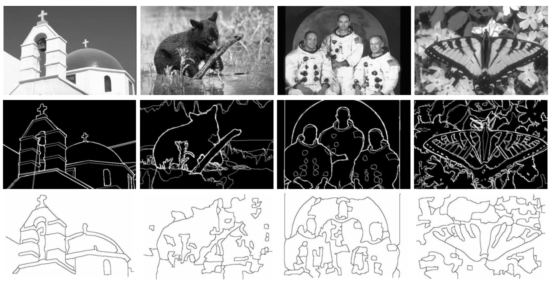

Figure 4. Example results: The original images, their ground truth segmentation and the obtained results by combining region and edge cues in our homogeneous CRF model.

However as illustrated in figure 5, the non-homogeneous CRF model provides better results than the homogeneous CRF model which demonstrate the importance of contextual information integration. In future work, we plan to add more informative features, learn the weighting parameter and extend our model to a multi-scale framework.

6.1.5. Vein Networks extraction and description

Keywords: *Vein-Networks*, *local features*, *points of interest*.

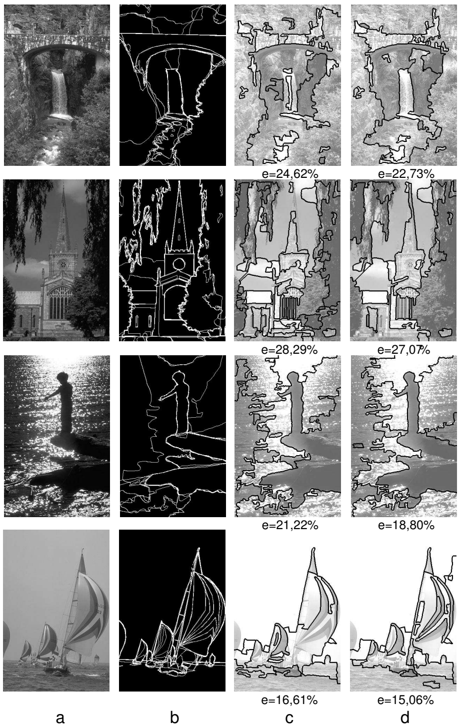

Figure 5. Example results: (a, b) The original images and their ground truth segmentation. The obtained results by (c) homogeneous and (d) non-homogeneous versions of our CRF model. The error rates are also shown.

Participant: Itheri Yahiaoui.

The identification and reconstruction of relationships between plants have been based largely on the reproductive organs. There are, however, situations in which these organs are not available and leaves are the only means to identify a plant. Few researchers have considered the automatic indexation of plant images. Those who do use different types of images, notably, reproductive organs, flowers, leaves etc. Researchers who use images of leaves focus on the shape and the size of the leaf. However, these features (shape and size) are insufficient to recognise a plant species. Although leaves may have very similar shape and size, they may represent several plant species. Botanist use the venation of the leave as one of the most informative feature to identify a plant.

Therefore, we are developing technology thats use both shape and venation to automatically index leaf images. To extract the vein network we use different methods based on first and second derivative computation. The quality of the extracted network depends largely on the resolution of the original image and lighting conditions. To extract the network from poor quality images we are currently developing a method that uses statistical learning to discriminate between the vein and background texture.

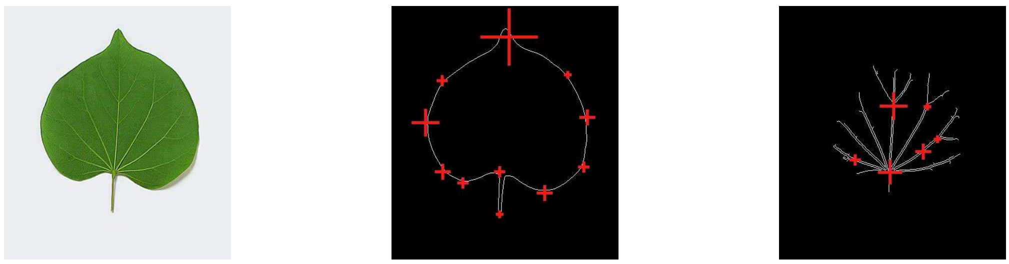

Figure 6. Vein Networks Extraction

Figure 6 shows three images, the most left portrays the original leaf, the subsequent image represents its silhouette and the most right shows its extracted vein networks. We compute points-of-interest to describe both contour and vein network.

6.1.6. Coarse-to-fine content-based video copy detection fusion

Keywords: *Vicopt*, *coarse-to-fine*, *copy detection*, *dissociated dipoles*, *video*.

Participants: Julien Law-to, Nozha Boujemaa, Alexis Joly.

We designed and developed a coarse-to-fine Video Copy Detection fusion strategy for heterogeneous degradation levels. Trading quality for time is a usual process in content-based copy detection systems in order to be fast enough in very large datasets. However, some specific contents are sometimes strongly transformed and require more accurate matching strategies. In this study, we proposed a 3 steps framework that allows to adapt to different levels of degradation of the original contents. The first step is based on a very light spatio-temporal local features representation and allows to perform very fast copy detection while being robust to most frequently encountered transformations, such as re-encoding, noise, patterns insertion, re-editing or ratio changes. The straight of this method is to track the local features over time during offline indexing step in order to obtain a strong compression of the indexed features. An assymetric and adaptative temporal matching strategy is then used during the online retrieval step and allows to trade temporal granularity of the detected segments for high speed improvements. The second step, applied when the first one fails, relies an a more robust method that does not use any temporal compression. All the local features extracted in sampled key frames are indexed and a robust spatio-temporal coherence measure is used during matching. This step allows to be invariant to most photometric and geometric transformations. Finally, the last step is applied when the two first ones fail by increasing drastically the number of extracted features in resized queries. It allows to be robust to very strong geometric distortions such as very small picture in picture. The proposed strategy was evaluated within TRECVID 2008 copy detection task and obtained the second best result among 26 participants. A publication was accepted to TRECVID proceedings and a presentation was given during the workshop.

6.1.7. Object recognition

Keywords: *features selection*, *interest points*, *interpretability*, *local descriptors*.

Participants: Ahmed Rebai, Alexis Joly, Nozha Boujemaa.

Current techniques used for object recognition are still by far poor compared to the human's ability in accomplishing such tasks. In object recognition, we are confronted to many problems about invariance as viewpoint and illumination. There are also other problems like dealing with non-rigid objects, complex background and occlusion by unknown objects just to name a few. During the beginning of 2008, we studied the most powerful existing techniques related to local features detection, stability selection and machine learning. We believe that each object category does possess a specific description that defines it perfectly. That's why using one kind of descriptor or feature extractor would not lead to a global cognitive system. This year, we have developed a generic framework that aims to use multiple interest points detectors combined with different local signatures in order to reach a complementary description. The idea is to create a specific descriptor that best fits a category of interest points. That is, using interpretability—in terms of human vision—as a criterion to reach an efficient description. Indeed, we carried out some experiments with different shape masks and obtained promising results. To study in depth the used descriptors, we developed a visual tool integrated to the IKONA interface. Currently, we are experimenting a feature selection technique to improve the learning steps.

6.2. Interactive retrieval

6.2.1. Clustering for pallet harmony generation and pattern trends browsing

Keywords: *clustering*, *harmony rules*.

Participants: Hichem Houissa, Nozha Boujemaa.

In the context of TRENDS we have planed to provide end-users (car designers) new functionnalities of assessing and browsing their image collections. The information retrieval process combines textual and visual search engines besides of novel functionality of browsing images such as pallets generation and clustering modules. We have contributed to the formalization and the development of the harmony generation part. This part derives pairs of harmonious colors from low level photometric features according to harmony rules applied on the chromatic circle. According to complementary harmony rule, colors should be antagonist around the chromatic circle. Secondary harmony rules stipulates that harmonious colors should compose a triangle on the chromatic circle. Pallets of textures reflecting the most relevant textures composing a set of selected images are also generated (cf. Figure 7 and [13] for more details). This process goes through several steps: first images are roughly segmented using competitive agglomeration algorithm, local descriptors were extracted to separate uniform regions from textured ones by means of coherence criterion assessing the spatial distribution of these local descriptors. Finally, clustering is applied to aggregate most similar textures and display most relevant ones to the users.

Another novelty concerns the fast browsing into a database by means of clustering images according to a combination of their photometric features (color, texture and shape signatures) but reduced to smaller dimension by applying a Principal Component Analysis.

In this module, several difficulties were raised due to the TRENDS database size that handicaps the image retrieval in terms of computational time and relevance of results in case relevance feedback is concerned. Furthermore, the external query has faced the issue of the database size since it can be applied to a subpart of it, hence excluding the richness provided by the TRENDS database that contains around 1.8 million images.

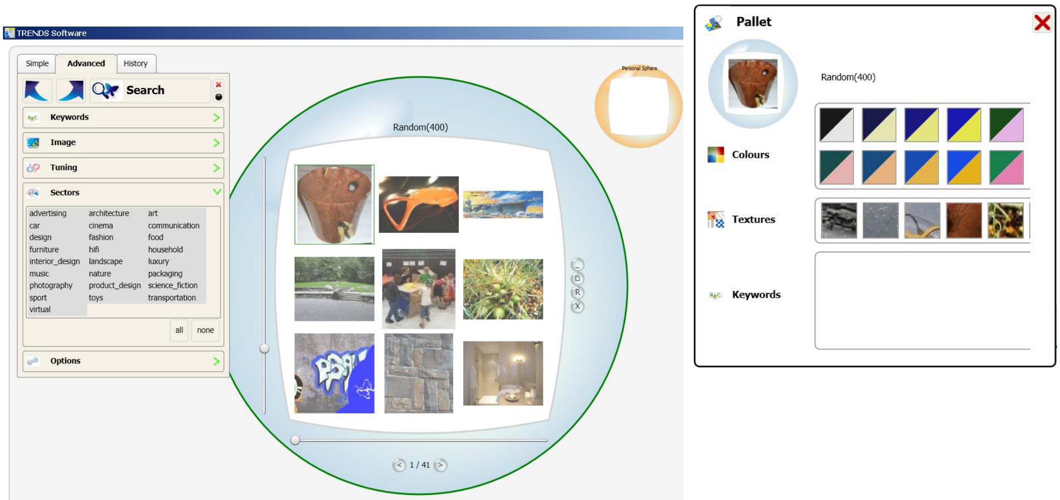

Figure 7. TRENDS pallet generation. On the left, the set of images from which to extract pallets of colours and textures. On the right, the generated pallet of harmonious pairs of colours (top) and most relevant textures (bottom).

6.2.2. Enhanced result visualisation functionnalities

Keywords: *Isomap*, *SNE*, *graph representation*, *optimisation methods*.

Participants: Raffi Enficiaud, Anne Verroust-Blondet.

Visualising the result of a query is a major issue. Most of the systems today display the results as a long list of images. This is the way the IKONA client software was articulated. This work aims at providing Maestro with a new visualising module, in order to enhance the end user navigation among the set of results. It then should be adapted to video visualising.

Visualising may be formulated as a dimensionality reduction task. The problem is to project a high dimensional space - the features space - into a 2D or 3D space, while preserving at best the similarity relations between the elements in the database. For this purpose, an usual approach is to first consider the similarity matrix between the database entries, which provides an information about the relative distance between the entries. Several strategies can then be adopted for the projection, both providing results as maps.

Two methods have been tested with interesting results. The first one is the ISOMAP [38] method, which aims at preserving the geodesic distances by using a multi-dimensional scaling after the elements being projected into a graph representation. An example of result is provided in figure 8.

The second method called *Stochastic Neighbour Embedding* [36] is formulated as a cost function between two probability distributions. The *true* distribution is computed within the high dimensional space, while the *target* is computed in the projection plane. Each distribution describes the relation between an element of the database and its potential neighbours. The better the distributions match, the better the neighbouring relations are preserved. A Kullback-Leibler divergence between these distributions is the cost function to be minimised. The minimum is computed with a classical gradient-descent method. An example of result is given in figure 8.

While presenting good results, some issues should be addressed. For the ISOMAP method, a neighbourhood size should be chosen by hand and when used on complex sets this step often creates several disconnected components. For the SNE method, the optimisation formulation is of complexity $O(N^2)$ which does not match with the scalability issue of the VITALAS project. Moreover, an optimisation method better adapted to the size of the problem should be found.

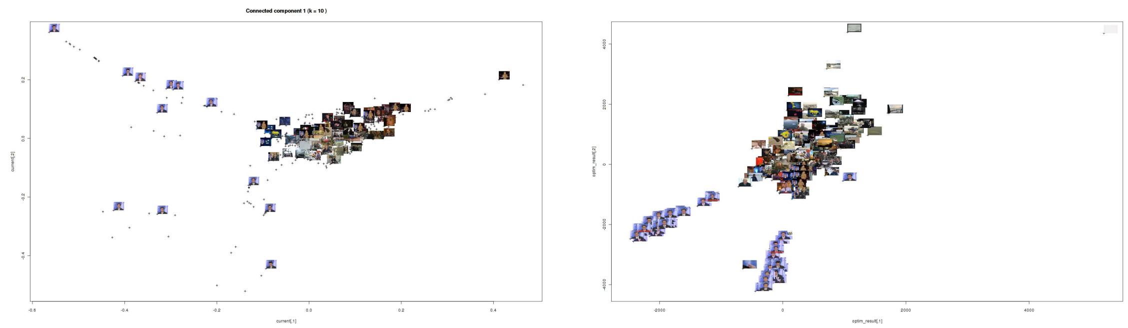

Figure 8. Examples of projection. Left: ISOMAP projection; right: SNE projection.

6.2.3. Relevance feedback on local image features, with application to the identification of plant species

Keywords: *active learning*, *local features*, *relevance feedback*, *semi-supervised learning*.

Participants: Wajih Ouertani, Nozha Boujemaa, Michel Crucianu.

The characterization, evaluation and use of plant biodiversity is based on the precise and efficient identification of its components and especially of the species. The identification keys issued from systematic botany mainly rely on characteristics that are ineffective in many real-world situations. The developmenf the inventory of species, of community ecology and of the monitoring of self-propagating plants is limited because it requires an active and continuing involve ment of the very few highly specialized botanists. The collaboration between the UMR AMAP and the IMEDIA team aims to address this challenge by exploiting image analysis and recognition in a generic interactive species identification system. Since the identification process should be interactive, we decided to further explore relevance feedback on sets of local image features that describe regions of interest of an image. In the case under focus here, such regions would correspond to plant organs whose attributes are potentially relevant for identification. The PhD of Wajih Ouertani addresses active and semi-supervised learning methods for relevance feedback with local features describing regions of interest, with the aim to effectively support the identification of plant species by exploiting the information acquired during the interaction with the user. The work started in October 2008 with a state of the art and the extension of the relevance feedback mechanism implemented in the search engine of IMEDIA to take into account local features with the help of an existing kernel [30] for SVM.

6.2.4. Large scale relevance feedback on local visual features

Keywords: *Adaboost*, *local descriptors*, *multi-probe locality sensitive hashing*, *relevance feedback*, *scalability*.

Participants: Saloua Litayem, Alexis Joly, Nozha Boujemaa.

Relevance feedback on freely selected regions of the image is a very promising query paradigm having many prospects such as building semi-supervised recognition models for objects categories. Unfortunately, one major challenge is the scalability of such methods. In fact, they require a very large amount of extracted visual local features that make most usual machine learning techniques unusable in the online context of relevance feedback. The scalability issues are mainly related to the complexity of the online prediction time that is usually linear in the number of images in the dataset and in the average number of features per image. Currently, we are investigating new algorithms to render the prediction time sublinear. The idea is to design learning models with bounded supports and to use efficient similarity search structures during the prediction step. Indeed, simple bounded queries in multi-dimensional spaces can be processed in sublinear time with most recent similarity search structures. More precisely, we generate the bounded learning models with the well known Adaboost algorithm but using range queries in the local features space as weak classifiers. The resulting model is then a linear combination of simple range query classifiers, whose prediction can be easily computed in sublinear time over the dataset when using a similarity search structure. For our first experiments, we plan to use the new similarity search structure developped by IMEDIA and which is based on multi-probe locality sensitive hashing. This new search method is currently being integrated in IKONA/MAESTRO software.

6.3. Automatic annotation and learning

6.3.1. Local descriptors sampling strategies for bag-of-words images representations

Keywords: *automatic image annotation*, *bag-of-word*, *cross-modal indexing*, *object detection*, *visual vocabulary*.

Participants: Nicolas Hervé, Nozha Boujemaa, Michael Houle.

The success of the bag-of-word approach for text has inspired the recent use of analogous strategies for global representation of images with local visual features. Combined with a learning strategy, these representations are often used for automatic image annotation. For a general presentation of automatic image annotation, see [34]. Applying the bag-of-word strategy is inherently easier for text than for images. Indeed, the words are clearly identified and the vocabulary, generic across all documents, is known in advance. For images on the other hand, such words are not defined and one need to analyse and characterise the data in the vicinity of all pixels, over all potential scales. The generation of such local descriptors that fully cover all images is computationally too expensive on current hardware.

In their desire to reduce image processing costs, the computer vision community has put much effort into the development of point-of-interest detectors, beginning with their original applications for image registration. Detectors are generally attracted to specific areas of images that have high variation in the visual signal, such as the vicinity of edges and corners of regions.

High visual variance is often considered to be associated with high semantic content. However, this assumption is not always justified, as low variance may also carry important semantic information in generic image datasets. We therefore believe that all areas of an image must be considered, regardless of whether they exhibit high variation or low variation. It is known that contextual information has a positive effect when performing object detection [33].

In order to verify this assumption, we introduced a generic framework for documents representation [35]. As the bag-of-word approach comes from the text community, we choose to experiment both dense sampling and point-of-interest sampling strategies for local patches selection on a text corpus and an image dataset. We have shown that on both corpuses, the behaviour of dense sampling and point-of-interest sampling is similar and, in both cases, the results are generally better with dense sampling. We believe that dense pattern sampling should be used in image representation as it leads to less information loss in comparison with point-of-interest detection.

6.3.2. Semantic region labelling using a point pattern analysis

Keywords: *homogeneous Poisson process*, *interest points*, *nearest neighbour method*, *quadrant count method*, *semantic labelling*, *spatial dispersion*.

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

Several recent studies and researches focused on the combination of global and fine local region description using points of interest. The main benefit of global approach is that homogeneous parts of the image can easily be described by means of global attributes whereas small details are ignored. In the other hand, as the region gets smaller and with high photometric variability, points of interest are more appropriate to carry local description. We proposed a new semantic labelling of regions using their interest point spatial dispersion. We introduced a point-based criterion to label regions into homogeneous and textured classes. Our point based criterion is based on a point pattern analysis study. If a region has few extracted points of interest, or concentrates some points on the contours due to the coarse segmentation and rough contours, it is more likely to be labelled as homogeneous. If a region is covered with high number of interest points more or less regularly dispersed in the region, this means that it is characterised by an important photometric variability and it is more likely to be labelled as textured.
The point pattern analysis is based on comparing the distribution of interest points detected in a given region with randomly distributed points in the same region. The Complete Spatial Randomness (CSR) model assumes that points are distributed at random, which often display features that look like clumping. In addition, this model assumes that the mean density of points per unit area is known. The mathematical construct that we used to simulate a CSR model is the homogeneous Poisson process.
The Quadrant Count Method can be described simply as partitioning the data set into n equal sized quadrants; construct the frequency distribution of the number of points per quadrant, compute the variance and the mean. The variance to mean ratio is a good measure of the degree of randomness of a given distribution of points. We compare the VMR of our distribution with the VMR of the random distribution to know whether our distribution is more clustered or more dispersed than the random distribution.

Figure 9. homogeneous labelled regions resulting from points accumulation over borders and grouping on small details. Textured labelled regions: entirely covered by interest points denoting an important photometric variation.

The Nearest Neighbour method is based on computing the ratio of expected and observed mean value of the nearest neighbour distances. This ratio is used to determine if a set of interest points is clustered or not. Figure 9 shows examples of textured and homogeneous labelled regions.
This work has been presented in [12].

6.4. Software

6.4.1. IKONA/MAESTRO software

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

Participants: Mehdi Bouabta, Mathieu Coutaud, Nicolas Hervé, Laurent Joyeux, Raffi Enficiaud, Jean-Paul Chieze.

As each year, the integration of latest research results has been achieved. One can mention the handling of the similarity matrix across the content of the database, on which the visual retrieval is based. The parsing of video files has also been successfully tested.

A disconnected protocol communication has been implemented, in order to cope wit h fully TCP based clients such as the Web Services framework. The Web Service interface to the Maestro API has been finalised and successfully tested in an industrial environment (VITALAS project). The support for new functionalities is planned within the VITALAS project.

Some major modifications of the core functionalities have also been undertaken. The main goal is to properly handle video databases which is a major issue for 2009. Also the OS specific dependencies have been abstracted in order to run MAESTRO on several platforms (Unices 32 & 64 bits, Mac OSX, Win32). This part will be fully finalised and tested in the early 2009.

A better packaging of the software is on the final stage of development. Its aim is to ease the delivery of binaries to external partners. A complete automated framework for demos (packaging and demon supervision) was finalised.

Finally, a particular care concerning the quality of the software was taken. The several projects are built on different compilers and a complete test bed with unit and regression tests is played every nights. This ensures, as in industrial environments, an early breakdown of functionality detection. This point will be further improved during 2009.

The use of Web Services API stressed the fact that a better support for concurrent access is needed. There is also the need for dynamic indexing and a totally disconnected communication protocol (session management). These will be the major issues for the next year. For the early 2009, the finalising of the video handling and the multi-OS support is also planned. The interface with database systems (Oracle, MySQL and PotgreSql), developed during 2007 and 2008, will be finalised and used in the ongoing projects. The development of an administrative interface is considered.
PMH software is a C++ package implementing the new Probabilistic Multidimensional Hashing technique developed jointly with INA [17]. It allows indexing and search of very large datasets of high dimensional feature vectors (C++ library, 15,000 lines). It has been integrated as a MAESTRO indexing structure module and will be the basis of most scalability aspects of VITALAS IP system.

7. Other Grants and Activities

7.1. National Initiatives

7.1.1. Collaboration with AMAP

Since October 2007, a collaboration between IMEDIA and AMAP (botAnique et bioInforMatique de l'Architecture des Plantes) has started. AMAP is a joint research unit between INRA, CIRAD, IRD and CNRS located in Montpellier (http://amap.cirad.fr/).

- In 2008, the RTRA project on weeds in Camargue ricefields named "Reconnaissance automatisée des espèces adventices des rizières de Camargue par les techniques de reconnaissance de contenus visuels" has been achieved. AMAP is providing financial support for research engineer contract to work on biodiversity content search problems together with IMEDIA. Demo of preliminary results are avalaible here: http://www-rocq.inria.fr/cgi-bin/imedia/circario.cgi/bio_diversity?select_db=1. These preliminary investigations have proven the relevance of content search engines for species identification and prepare the floor for the next two collaborations below.
- 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.

- The project Pl@ntNet [2009-2012] 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 (www.agropolis.fr/)and is titled “Plant Computational Identification and Collaborative Information System”.

7.1.2. 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.

7.1.3. InfoMagic[2005-2008]

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

7.2. European Initiatives

7.2.1. 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".

7.2.2. 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.

The demo of the first version of VITALAS system has been presented at NEM summit (October 13-15, 2008, Saint-Malo, France) and at ICT Lyon (November 25-27, 2008, Lyon, France).

7.2.3. 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. MUSCLE-VCD-2007 corpus dissemination [2008]

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). In 2008, the corpus created for this event was maintained on the web and the community using it has been extensively growing. More than 60 research teams, private and academic, did download the corpus and several publications referencing it are already published or will be soon. The corpus was also requested by Google lab , currently developing a product with YouTube and by the MPAA (Motion Picture Association of America, http://www.mpaa.org/) who did organise in 2007/2008 a benchmarking initiative of commercial products. Finally, the corpus was used as development data for TRECVID 2008 copy detection task (see next section).

7.3.2. Organisation of TRECVID video copy detection task [2008]

IMEDIA co-organised a new content-based video copy detection task within the international benchmarking campaign TREDVID (http://www-nlpir.nist.gov/projects/trecvid/), in collaboration with NIST (US /National Institute of Standards and Technology)/. IMEDIA was a member of the steering committee of this task and actively participated to its definition and organisation (data, queries, metrics, participants). IMEDIA defined an evaluation framework and developed several tools to create automatically the final queries in a blind way. For its first running year, the task was a real success with 26 participants all over the world including major companies and labs in the field. A special session about this task was hold during TRECVID workshop 2008 (Gaithersburg, November 17-18) with consistent IMEDIA contributions.

7.3.3. Cooperation with NII, Japan

Joint collaboration with Michael Houle has been established since 2006. Several visits and mobilities have been achieved between IMEDIA and NII. The two main topics consist on scalable clustering and object recognition.

7.3.4. Cooperation with John Hopkins University, USA

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

8. Dissemination

8.1. Seminars, presentations and other dissemination activities

8.1.1. Demos

Demos of IKONA/MAESTRO software have been presented at:

- **VITALAS awareness workshop** February 21, 2008 at IRT premises (Munich, Germany);
- **NEM summit** October 13-15, 2008 in Saint-Malo, France;
- **ICT Lyon** November 25-27, 2008 in Lyon, France.

8.1.2. Scientific Movie

Participants: Nozha Boujemaa, Alexis Joly, Julien Law-to.

Jointly with INRIA Multimedia team, a 12 minutes movie was realised about IMEDIA work on content-based copy detection, in the scope of MUSCLE Network of Excellence. The movie introduces visual based search engines and related problematics and then focus on two copy detection demos presented by the members of the team. It is available in different formats on several websites: MUSCLE website (http://www.muscle-noe.org/content/view/148/43/), INRIA website, Dailymotion (42 views for now, http://www.dailymotion.com/relevance/search/imedia%2Bcopy/video/x4e0gp_the-visual-copy-retrieval_tech).

8.1.3. Awarded student visit

IMEDIA team did welcome during one day Cécile Bérillon, a young student that did gain a French mathematical award (Olympiades de mathématiques). INRIA multimedia team realised a movie about this visit.

8.1.4. Dissemination to large public and European community

- An editorial of "eStrategies|Europe" on Vitalas, titled "The search goes on for more efficient retrieval methods" by British Publishers (http://www.britishpublishers.com)
- ICT Results: "Results that lead the way" (Muscle) http://cordis.europa.eu/ictresults/index.cfm/section/news/tpl/article/id/90028
- Sciences et Avenir Journal: announcement of Pl@ntNet lunching
- IMEDIA presented its works and activities to the head and people from the CUBE, a French creation center dedicated to numeric experiences (http://www.lesiteducube.com/homepage/01.html)

8.2. Leadership with scientific community

8.2.1. Nozha Boujemaa

- Scientific coordinator of MUSCLE NoE (Network of Excellence FP6);
- Scientific coordinator of VITALAS IP FP6;
- Scientific coordinator of CHORUS CA FP6;
- Expert for ESF (European Science Foundation : www.esf.org), appointed for 2008-2009
- 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
- Invited presentation: "Disruptive Search Applications?" in "Techno-economic trends" session, workshop "Socio-Economic Challenges of Search", Sevilla, Spain (organised by Ramón Compañó - Institute for Prospective Technological Studies (IPTS)) European Commission, September 2008
- Pannelist pour "CBMI 08 Panel Session: The Future of Multimedia Retrieval Research", London, UK; organised by Aggelos Katsaggelos (Northwestern University, USA), June 2008
- Invited speaker in the session: "Next Generation Information Retrieval and Analysis Technologies" during de EU-Japan Cooperation Forum on ICT Research, Tokyo, Japan; Organizers: European Commission, MEXT, MIC, MITI, MOFA, March 2008
- Member of steering committee of the first international evaluation campaign on Video copy detection in the context of TrecVid'08: http://www-nlpir.nist.gov/projects/tv2008/tv2008.html
- 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 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...
- 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 several recruitment committees of INRIA (CR2)
- Several PhD Jury committee members-president: CNAM, ENST...

8.2.2. Michel Crucianu

- Scientific expert for the French national Research Agency (ANR), call “Contenu et Interaction”.
- Journal reviewer: Information Science, International Journal of Image and Graphics, Journal of Intelligent & Fuzzy Systems, IEEE Trans. on Neural Networks, Data & Knowledge Engineering.

8.2.3. Joost Geurts

- Programme committee member for the International MultiMedia Modeling Conference (MMM2008) and for SAMT 2008, the international conference on Semantic and Digital Media Technologies.

8.2.4. Alexis Joly

- Member of the steering committee of VITALAS IP FP6 (leader of WP2 "Enabling technologies: Media Content Description and Summarisation").
- Member of the steering committee of TRECVID 2008 copy detection task.
- Scientific expert for the French National Research Agency (ANR), call "Contenu et interaction".
- invited presentation at INRIA LEAR seminars in October 2008
- Journal Reviewer: IEEE Transactions on Pattern Analysis and Machine Intelligence, VLDB journal, CVIU journal, IEEE Transactions on multimedia, Signal processing journal.

8.2.5. Anne Verroust-Blondet

- Member of the Humanities and Social Sciences committee for the “Blanc” and “Young researcher” programmes of 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 steering committee of the CNRS GDR IG (Informatique Graphique) ;
- Member of the editorial board of the “Revue Electronique Francophone d’Informatique Graphique” ;

8.3. Teaching

8.3.1. Nozha Boujemaa

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

8.3.2. Michel Crucianu

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

8.3.3. 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.

8.3.4. Nicolas Hervé

- 24h TP on Java Database Connectivity (JDBC), CNAM 3rd year (NFA011)

8.3.5. Alexis Joly

- 5h Tutorial on "Similarity Search Algorithms and Structures" in the scope of a CIRM meeting, May 2008 (Centre International de Rencontres Mathématiques, http://www.cirm.univ-mrs.fr/). Slides available online (http://www-rocq.inria.fr/~ajoly/index.php?n=Main.Education)

8.3.6. 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 knowledge " 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", vol. 29, n^{o} 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

Articles in International Peer-Reviewed Journal

[6] M. CRUCIANU, J.-P. TAREL, M. FERECATU. *An Exploration of Diversified User Strategies for Image Retrieval with Relevance Feedback*, in "Journal of Visual Languages and Computing", 2008.

[7] M. FERECATU, N. BOUJEMAA, M. CRUCIANU. *Semantic interactive image retrieval combining visual and conceptual content description*, in "ACM Multimedia Systems", vol. 13, n^{o} 5-6, 2008, p. 309-322.

[8] M. FERECATU, D. GEMAN. *A statistical framework for image category search from a mental picture*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", 2008.

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

[10] H. TANG, T. FANG, P. DU, P. SHI. *Intra-dimensional Feature Diagnosticity in the Fuzzy Feature Contrast Model*, in "Image and Vision Computing", vol. 26, n^{o} 6, 2008, p. 751–760.

[11] J. YU, J. AMORES, N. SEBE, Q. TIAN. *Distance Learning for Similarity Estimation*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", vol. 30, n^{o} 3, June 2008, p. 451-462.

International Peer-Reviewed Conference/Proceedings

[12] S. BAHROUN, Z. BELHADJ, N. BOUJEMAA. *Semantic region labelling using a point pattern analysis*, in "16th European Signal Processing Conference (EUSIPCO 2008), Lausanne", August 2008.

[13] A. BERECIARTUA, C. BOUCHARD, J.-F. OMHOVER, M. FERECATU, H. HOUISSA, F. GANDON, G. LOGEROT. *A new semantic text-image search engine for car designers*, in "IEEE International Workshop on Content-Based Multimedia Indexing (CBMI'08), London", June 2008.

[14] O. BESBES, N. BOUJEMAA, Z. BELHADJ. *Non-Homogeneous Conditional Random Fields for Contextual Image Segmentation*, in "IEEE International Symposium on Multimedia (ISM2008), Berkeley", December 2008.

[15] O. BESBES, N. BOUJEMAA, Z. BELHADJ. *Stochastic Image Segmentation by Combining Regions and Edge Cues*, in "IEEE International Conference on Image Processing 2008 (ICIP'08), San Diego", October 2008.

[16] M. CHAOUCH, A. VERROUST-BLONDET. *A Novel Method for Alignment of 3D Models*, in "IEEE International Conference on Shape Modeling and Applications (SMI'08), Stony Brook University", June 2008.

[17] A. JOLY, O. BUISSON. *A Posteriori Multi-Probe Locality Sensitive Hashing*, in "ACM International Conference on Multimedia (MM'08), Vancouver, British Columbia, Canada", October 2008, p. 209–218.

[18] A. JOLY, J. LAW-TO, N. BOUJEMAA. *INRIA-IMEDIA TRECVID 2008: Video Copy Detection*, in "NIST TRECVID Workshop, Gaithersburg, MD", November 2008.

Scientific Books (or Scientific Book chapters)

[19] S. P. WILSON, J. FAUQUEUR, N. BOUJEMAA. *Mental Search in Image Databases: Implicit Versus Explicit Content Query*, in "Machine Learning Techniques for Multimedia", Book Series Cognitive Technologies, Springer, 2008.

Books or Proceedings Editing

[20] N. BOUJEMAA, M. DETYNIECKI, A. NÜRNBERGER (editors). *Adaptive Multimedia Retrieval: Retrieval, User and Semantics*, Book Series - Lecture Notes in Computer Science (LNCS), vol. 4918/2008, 2008.

Research Reports

[21] M. CHAOUCH, A. VERROUST-BLONDET. *A Novel Method for Alignment of 3D Models*, RR-6408, Rapport de recherche, 2008, http://hal.inria.fr/inria-00203336/en/.

[22] M. FERECATU, D. GEMAN. *A Statistical Framework for Image Category Search from a Mental Picture*, RR-6584, Rapport de recherche, 2008, http://hal.inria.fr/inria-00303572/en/.

References in notes

[23] 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.

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

[25] 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.

[26] M. CHAOUCH, A. VERROUST-BLONDET. *Enhanced 2D/3D Approaches Based on Relevance Index for 3D-Shape Retrieval*, in "International Conference on Shape Modeling and Applications (SMI 2006), Matsushima, Japan", June 2006.

[27] 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.

[28] S. EL FEKIH. *Multiviewer pour la navigation dans les bases de données 3D*, Mémoire de stage ingénieur, ISI, Tunisia, 2008.

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

[30] K. GRAUMAN, T. DARRELL. *The Pyramid Match Kernel: Efficient Learning with Sets of Features*, in "Journal of Machine Learning Research", vol. 8, 2007, p. 725–760.

[31] A. HAMZAOUI. *Catégorisation de résultats de recherches multicritères d'images*, Masters thesis, Université Pierre et Marie Curie, UPMC, September 2008.

[32] 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.

[33] 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.

[34] N. HERVÉ, N. BOUJEMAA. *Automatic image annotation*, in "Encyclopedia of database systems", Springer, 2009.

[35] N. HERVÉ, N. BOUJEMAA, M. HOULE. *Document description: what works for images should also work for text?*, in "21st Annual IS&T/SPIE Symposium on Electronic Imaging, San Jose", January 2009.

[36] G. HINTON, S. T. ROWEIS. *Stochastic Neighbor Embedding*, in "Advances in Neural Information Processing Systems", vol. 15, 2002, p. 857–864.

[37] M. HOULE. *Clustering without data: the Relevant Set Correlation model*, in "Intl. Workshop on DataMining and Statistical Science (DMSS)", 2006, p. 54–61.

[38] J. TENENBAUM, V. DE SILVA, J. LANGFORD. *A global geometric framework for nonlinear dimensionality reduction*, in "Science", vol. 290, n^{o} 5500, 2000, p. 2319–2323.