IN PARTNERSHIP WITH:
CNRS

Institut national des sciences appliquées de Rennes

Université Rennes 1

Activity Report 2012

Project-Team TEXMEX

Multimedia content-based indexing

IN COLLABORATION WITH: Institut de recherche en informatique et systèmes aléatoires (IRISA)

RESEARCH CENTER
Rennes - Bretagne-Atlantique

THEME
Vision, Perception and Multimedia Understanding

Table of contents

Project-Team TEXMEX

Keywords: Multimedia, Video, Natural Language, Large Scale, Data Mining

Creation of the Project-Team: November 01, 2002 .

1. Members

Research Scientists

Patrick Gros [Team Leader, Senior Research Scientist, Inria, HdR]
Laurent Amsaleg [Research Scientist, CNRS]
Vincent Claveau [Research Scientist, CNRS]
Teddy Furon [Research Scientist, Inria]
Guillaume Gravier [Research Scientist, CNRS, HdR]
Hervé Jégou [Research Scientist, Inria]

Faculty Members

Ewa Kijak [Associate Professor, Univ. Rennes 1]
Annie Morin [Associate Professor, Univ. Rennes 1, HdR]
François Poulet [Associate Professor, Univ. Rennes 1, HdR]
Christian Raymond [Associate Professor, INSA Rennes]
Pascale Sébillot [Professor, INSA Rennes, HdR]
Philippe-Henri Gosselin [Professor, ENSEA, Inria delegation since 01/09, HdR]
Camille Guinaudeau [Assistant professor at Univ. Rennes until 31/08]

External Collaborators

Emmanuelle Martienne [Associate Professor, Univ. Rennes 2]
Fabienne Moreau [Associate Professor, Univ. Rennes 2]
Laurent Ughetto [Associate Professor, Univ. Rennes 2]

Engineers

Rachid BenMokhtar [Inria Technical Staff, Quaero project]
Morgan Bréhinier [Inria Technical Staff, OpenSem project, until 14/09]
Sébastien Campion [Inria Research Engineer]
Jonathan Delhumeau [Inria Technical Staff, Quaero project]
Carryn Hayward [Inria Technical Staff, Quaero project, also with SAF]
Diana Moise [Inria Technical Staff, Quaero project since 01/03]

PhD Students

Petra Bosilj [Grant from University of South Brittany since 01/10]
Mohamed-Haykel Boukadida [CIFRE grant with Orange, since 02/01]
Thanh Toan Do [MESR grant, until 30/09]
Thanh Nghi Doan [Vietnamese government grant and Brittany Council grant]
Ali Reza Ebadat [Quaero project, until 30/09]
Khaoula Elagouni [CIFRE grant with Orange]
Julien Fayolle [Quaero project and Brittany council grant]
Gylfi Gudmundsson [Quaero project]
Mihir Jain [Inria Cordis Grant]
Ludivine Kuznik [CIFRE grant with INA]
Abir Ncibi [Inria Grant]
Cédric Penet [CIFRE grant with Technicolor]
Bingjing Qu [CIFRE grant with INA, since 09/17]
Anca-Roxana Simon [MESR grant, since 10/01]

Post-Doctoral Fellows

Benjamin Mathon [Inria postdoc on ANR project Secular, since 01/10]
Josip Krapac [Inria postdoc, Quaero Project]
Bogdan Ludusan [Inria postdoc, Quaero Project, from 15/05 to 14/12]
Denis Shestakov [Inria postdoc, Quaero Project, since 18/04]
Anh Phuong Ta [Inria postdoc, Quaero project, until 30/11]
Wanlei Zhao [Inria postdoc, Quaero project, since 02/07]

Visiting Scientists

Michele Trevisiol [PhD at Yahoo Research and University of Pompeu Fabra (Barcelona), Spain]
Giorgos Tiolias [PhD at National Technical University of Athens, Greece]

Administrative Assistant

Elodie Lequoc [Inria Secretary, partial position in the team]

2. Overall Objectives

2.1. Overall Objectives

With the success of sites like Youtube or DailyMotion, with the development of the Digital Terrestrial TV, it is now obvious that the digital videos have invaded our usual information channels like the web. While such new documents are now available in huge quantities, using them remains difficult. Beyond the storage problem, they are not easy to manipulate, browse, describe, search, summarize, visualize as soon as the simple scenario "1. search the title by keywords 2. watch the complete document" does not fulfill the user's needs anymore. That is, in most cases.

Most usages are linked with the key concept of repurposing. Videos are a raw material that each user recombines in a new way, to offer new views of the content, to adapt it to new devices (ranging from HD TV sets to mobile phones), to mix it with other videos, to answer information queries... Somehow, each use of a video gives raise to a new short-lived document that exists only while it is viewed. Achieving such a repurposing process implies the ability to manipulate videos extracts as easily as words in a text.

Many applications exist in both professional and domestic areas. On the professional side, such applications include transforming a TV broadcast program into a web site, a DVD or a mobile phone service, switching from a traditional TV program to an interactive one, better exploiting TV and video archives, constructing new video services (video on demand, video edition, etc). On the domestic side, video summarizing can be of great help, as can a better management of the videos locally recorded, or simple tools to face the exponential number of TV channels available that increase the quantity of interesting documents available, overall increasing but make them really hard to find.

In order to face such new application needs, we propose a multi-field work, gathering in a single team specialists that are able to deal with the various media and aspects of large video collections: image, video, text, sound and speech, but also data analysis, indexing, machine learning... The main goal of this work is to segment, structure, describe, or de-linearize the multimedia content in order to be able to recombine or re-use that content in new conditions. The focus on the document analysis aspect of the problem is an explicit choice since it is the first mandatory step of any subsequent application, but using the descriptions obtained by the processing tools we develop is also an important goal of our activity.

To illustrate our research project in one short sentence, we would like our computers to be able to watch TV and use what has been watched and understood in new innovative services. The main challenges to address in order to reach that goal are: the size of the documents and of the document collections to be processed, the necessity to process jointly several media and to obtain a high level of semantics, the variety of contents, of contexts, of needs and usages, linked to the difficulty to manage such documents on a traditional interface.

Our own research is organized in three directions: 1- developing advanced algorithms of data analysis, description and indexing, 2- searching new techniques for linguistic information acquisition and use, 3- building new processing tools for audiovisual documents.

2.1.1. Advanced algorithms of data analysis, description and indexing

Processing multimedia documents produces most of the time lots of descriptive metadata. These metadata can take many different aspects ranging from a simple label issued from a limited list, to high dimensional vectors or matrices of any kind; they can be numeric or symbolic, exact, approximate or noisy. As examples, image descriptors are usually vectors whose dimension can vary between 2 and 900, while text descriptors are vectors of much higher dimension, up to 100,000 but that are very sparse. Real size collections of documents can produce sets of billions of such vectors.

Most of the operations to be achieved on the documents are in fact translated in terms of operations on their metadata, which appear as key objects to be manipulated. Although their nature is much simpler than the data used to compute them, these metadata require specific tools and algorithms to cope with their particular structure and volume. Our work concerns mainly three domains:

- data analysis techniques, possibly coupled to data visualization techniques, to study the structure of large sets of metadata, with applications to classical problems like data classification, clustering, sampling, or modeling,
- advanced data indexing techniques in order to speed-up the manipulation of these metadata for retrieval or query answering problems,
- description of compressed, watermarked or attacked data.

2.1.2. New techniques for linguistic information acquisition and use

Natural languages are a privileged way to carry high level semantic information. Used in speech from an audio track, in textual format or overlaid in images or videos, alone or associated with images, graphics or tables, organized linearly or with hyperlink, expressed in English, French, or Chinese, this linguistic information may take many different forms, but always exhibits a common basic structure: it is composed of sequences of words. Building techniques that preserve the subtle links existing between these words, their representations with letters or other symbols and the semantics they carry is a difficult challenge.

As an example, actual search engines work at the representation level (they search sequences of letters), and do not consider the meaning of the searched words. Therefore, they do not use the fact that “bike” and “bicycle” represent a single concept while “bank” has at least two different meanings (a river bank and a financial institution).

Extracting high level information is the goal of our work. First, acquisition techniques that allow us to associate pieces of semantics with words, to create links between words are still an active field of research. Once this linguistic information is available, its use raises new issues. For example, in search engines, new pieces of information can be stored and the representation of the data can be improved in order to increase the quality of the results.

2.1.3. New processing tools for audiovisual documents

One of the main characteristics of audiovisual documents is their temporal dimension. As a consequence, they cannot be watched or listened to globally, but only by a linear process that takes some time. On the processing side, these documents often mix several media (image track, sound track, some text) that should be all taken into account to understand the meaning and the structure of the document. They can also have an endless stream structure with no clear temporal boundaries, like on most TV or radio channels. Therefore, there is an important need to segment and structure them, at various scales, before describing the pieces that are obtained.

Our work is organized in three directions. Segmenting and structuring long TV streams (up to several weeks, 24 hours a day) is a first goal that allows to extract program and non program segments in these streams. These programs can then be structured at a finer level. Finally, once the structure is extracted, we use the linguistic information to describe and characterize the various segments. In all this work, the interaction between the various media is a constant source of difficulty, but also of inspiration.

2.2. Highlights of the Year

- The project-team has participated to three tasks in the MediaEval'2012 evaluation campaign. We have obtained the best results for the Placing Task in the run without external data.
- We have obtained top results the ETAPE named entities evaluation campaign. Our system was rank first, significantly outperforming the concurrent submitted systems.

3. Scientific Foundations

3.1. Image description

In most contexts where images are to be compared, a direct comparison is impossible. Images are compressed in different formats, most formats are error-prone, images are re-sized, cropped, etc. The solution consists in computing descriptors, which are invariant to these transformations.

The first description methods associate a unique global descriptor with each image, *e.g.*, a color histogram or correlogram, a texture descriptor. Such descriptors are easy to compute and use, but they usually fail to handle cropping and cannot be used for object recognition. The most successful approach to address a large class of transformations relies on the use of local descriptors, extracted on regions of interest detected by a detector, for instance the Harris detector [82] or the Difference of Gaussian method proposed by David Lowe [84].

The detectors select a square, circular or elliptic region that is described in turn by a patch descriptor, usually referred to as a local descriptor. The most established description method, namely the SIFT descriptor [84], was shown robust to geometric and photometric transforms. Each local SIFT descriptor captures the information provided by the gradient directions and intensities in the region of interest in each region of a 4×4 grid, thereby taking into account the spatial organization of the gradient in a region. As a matter of fact, the SIFT descriptor has became a standard for image and video description.

Local descriptors can be used in many applications: image comparison for object recognition, image copy detection, detection of repeats in television streams, etc. While they are very reliable, local descriptors are not without problems. As many descriptors can be computed for a single image, a collection of one million images generates in the order of a billion descriptors. That is why specific indexing techniques are required. The problem of taking full advantage of these strong descriptors on a large scale is still an open and active problem. Most of the recent techniques consists in computing a global descriptor from local ones, such as proposed in the so-called bag-of-visual-word approach [89]. Recently, global description computed from local descriptors has been shown successful in breaking the complexity problem. We are active in designing methods that aggregate local descriptors into a single vector representation without loosing too much of the discriminative power of the descriptors.

3.2. Corpus-based text description and machine learning

Our work on textual material (textual documents, transcriptions of speech documents, captions in images or videos, etc.) is characterized by a chiefly corpus-based approach, as opposed to an introspective one. A corpus is for us a huge collection of textual documents, gathered or used for a precise objective. We thus exploit specialized (abstracts of biomedical articles, computer science texts, etc.) or non specialized (newspapers, broadcast news, etc.) collections for our various studies. In TEXMEX, according to our applications, different kinds of knowledge can be extracted from the textual material. For example, we automatically extract terms characteristic of each successive topic in a corpus with no a priori knowledge; we produce representations for documents in an indexing perspective [88]; we acquire lexical resources from the collections (morphological families, semantic relations, translation equivalences, etc.) in order to better grasp relations between segments of texts in which a same idea is expressed with different terms or in different languages...

In the domain of the corpus-based text processing, many researches have been undergone in the last decade. While most of them are essentially based on statistical methods, symbolic approaches also present a growing interest [78]. For our various problems involving language processing, we use both approaches, making the most of existing machine learning techniques or proposing new ones. Relying on advantages of both methods, we aim at developing machine learning solutions that are automatic and generic enough to make it possible to extract, from a corpus, the kind of elements required by a given task.

3.3. Stochastic models for multimodal analysis

Describing multimedia documents, *i.e.*, documents that contain several modalities (*e.g.*, text, images, sound) requires taking into account all modalities, since they contain complementary pieces of information. The problem is that the various modalities are only weakly synchronized, they do not have the same rate and combining the information that can be extracted from them is not obvious. Of course, we would like to find generic ways to combine these pieces of information. Stochastic models appear as a well-dedicated tool for such combinations, especially for image and sound information.

Markov models are composed of a set of states, of transition probabilities between these states and of emission probabilities that provide the probability to emit a given symbol at a given state. Such models allow generating sequences. Starting from an initial state, they iteratively emit a symbol and then switch in a subsequent state according to the respective probability distributions. These models can be used in an indirect way. Given a sequence of symbols (called observations), hidden Markov models (HMMs, [87]) aim at finding the best sequence of states that can explain this sequence. The Viterbi algorithm provides an optimal solution to this problem.

For such HMMs, the structure and probability distributions need to be a priori determined. They can be fixed manually (this is the case for the structure: number of states and their topology), or estimated from example data (this is often the case for the probability distributions). Given a document, such an HMM can be used to retrieve its structure from the features that can be extracted. As a matter of fact, these models allow an audiovisual analysis of the videos, the symbols being composed of a video and an audio component.

Two of the main drawbacks of the HMMs is that they can only emit a unique symbol per state, and that they imply that the duration in a given state follows an exponential distribution. Such drawbacks can be circumvented by segment models [86]. These models are an extension of HMMs were each state can emit several symbols and contains a duration model that governs the number of symbols emitted (or observed) for this state. Such a scheme allows us to process features at different rates.

Bayesian networks are an even more general model family. Static Bayesian networks [80] are composed of a set of random variables linked by edges indicating their conditional dependency. Such models allow us to learn from example data the distributions and links between the variables. A key point is that both the network structure and the distributions of the variables can be learned. As such, these networks are difficult to use in the case of temporal phenomena. Dynamic Bayesian [85] networks are a generalization of the previous models. Such networks are composed of an elementary network that is replicated at each time stamp. Duration variable can be added in order to provide some flexibility on the time processing, like it was the case with segment models. While HMMs and segment models are well suited for dense segmentation of video streams, Bayesian networks offer better capabilities for sparse event detection. Defining a trash state that corresponds to non event segments is a well known problem is speech recognition: computing the observation probabilities in such a state is very difficult.

3.4. Multidimensional indexing techniques

Techniques for indexing multimedia data are needed to preserve the efficiency of search processes as soon as the data to search in becomes large in volume and/or in dimension. These techniques aim at reducing the number of I/Os and CPU cycles needed to perform a search. Multi-dimensional indexing methods either perform exact nearest neighbor (NN) searches or approximate NN-search schemes. Often, approximate techniques are faster as speed is traded off against accuracy.

Traditional multidimensional indexing techniques typically group high dimensional features vectors into cells. At querying time, few such cells are selected for searching, which, in turn, provides performance as each cell contains a limited number of vectors [79]. Cell construction strategies can be classified in two broad categories: *data-partitioning* indexing methods that divide the data space according to the distribution of data, and *space-partitioning* indexing methods that divide the data space along predefined lines and store each descriptor in the appropriate cell.

Unfortunately, the "curse of dimensionality" problem strongly impacts the performance of many techniques. Some approaches address this problem by simply relying on dimensionality reduction techniques. Other approaches abort the search process early, after having accessed an arbitrary and predetermined number of cells. Some other approaches improve their performance by considering approximations of cells (with respect to their true geometry for example).

Recently, several approaches make use of quantization operations. This, somehow, transforms costly nearest neighbor searches in multidimensional space into efficient uni-dimensional accesses. One seminal approach, the LSH technique [81], uses a structured scalar quantizer made of projections on segmented random lines, acting as spatial locality sensitive hash-functions. In this approach, several hash functions are used such that co-located vectors are likely to collide in buckets. Other approaches use unstructured quantization schemes, sometimes together with a vector aggregation mechanism [89] to boost performance.

3.5. Data mining methods

Data Mining (DM) is the core of knowledge discovery in databases whatever the contents of the databases are. Here, we focus on some aspects of DM we use to describe documents and to retrieve information. There are two major goals to DM: description and prediction. The descriptive part includes unsupervised and visualization aspects while prediction is often referred to as supervised mining.

The description step very often includes feature extraction and dimensional reduction. As we deal mainly with contingency tables crossing "documents and words", we intensively use factorial correspondence analysis. "Documents" in this context can be a text as well as an image.

Correspondence analysis is a descriptive/exploratory technique designed to analyze simple two-way and multi-way tables containing some measure of correspondence between the rows and columns. The results provide information, which is similar in nature to those produced by factor analysis techniques, and they allow one to explore the structure of categorical variables included in the table. The most common kind of table of this type is the two-way frequency cross-tabulation table. There are several parallels in interpretation between correspondence analysis and factor analysis: suppose one could find a lower-dimensional space, in which to position the row points in a manner that retains all, or almost all, of the information about the differences between the rows. One could then present all information about the similarities between the rows in a simple 1, 2, or 3-dimensional graph. The presentation and interpretation of very large tables could greatly benefit from the simplification that can be achieved via correspondence analysis (CA).

One of the most important concepts in CA is inertia, *i.e.*, the dispersion of either row points or column points around their gravity center. The inertia is linked to the total Pearson χ^2 for the two-way table. Some rows and/or some columns will be more important due to their quality in a reduced dimensional space and their relative inertia. The quality of a point represents the proportion of the contribution of that point to the overall inertia that can be accounted for by the chosen number of dimensions. However, it does not indicate whether or not, and to what extent, the respective point does in fact contribute to the overall inertia (χ^2 value). The relative inertia represents the proportion of the total inertia accounted for by the respective point, and it is independent of the number of dimensions chosen by the user. We use the relative inertia and quality of points to characterize clusters of documents. The outputs of CA are generally very large. At this step, we use different visualization methods to focus on the most important results of the analysis.

In the supervised classification task, a lot of algorithms can be used; the most popular ones are the decision trees and more recently the Support Vector Machines (SVM). SVMs provide very good results in supervised classification but they are used as "black boxes" (their results are difficult to explain). We use graphical

methods to help the user understanding the SVM results, based on the data distribution according to the distance to the separating boundary computed by the SVM and another visualization method (like scatter matrices or parallel coordinates) to try to explain this boundary. Other drawbacks of SVM algorithms are their computational cost and large memory requirement to deal with very large datasets. We have developed a set of incremental and parallel SVM algorithms to classify very large datasets on standard computers.

4. Application Domains

4.1. Copyright protection of images and videos

With the proliferation of high-speed Internet access, piracy of multimedia data has developed into a major problem and media distributors, such as photo agencies, are making strong efforts to protect their digital property. Today, many photo agencies expose their collections on the web with a view to selling access to the images. They typically create web pages of thumbnails, from which it is possible to purchase high-resolution images that can be used for professional publications. Enforcing intellectual property rights and fighting against copyright violations is particularly important for these agencies, as these images are a key source of revenue. The most problematic cases, and the ones that induce the largest losses, occur when "pirates" steal the images that are available on the Web and then make money by illegally reselling those images.

This applies to photo agencies, and also to producers of videos and movies. Despite the poor image quality, thousands of (low-resolution) videos are uploaded every day to video-sharing sites such as YouTube, eDonkey or BitTorrent. In 2005, a study conducted by the Motion Picture Association of America was published, which estimated that their members lost 2,3 billion US$ in sales due to video piracy over the Internet. Due to the high risk of piracy, movie producers have tried many means to restrict illegal distribution of their material, albeit with very limited success.

Photo and video pirates have found many ways to circumvent even the protection mechanisms. In order to cover up their tracks, stolen photos are typically cropped, scaled, their colors are slightly modified; videos, once ripped, are typically compressed, modified and re-encoded, making them more suitable for easy downloading. Another very popular method for stealing videos is cam-cording, where pirates smuggle digital camcorders into a movie theater and record what is projected on the screen. Once back home, that goes to the web.

Clearly, this environment calls for an automatic content-based copyright enforcement system, for images, videos, and also audio as music gets heavily pirated. Such a system needs to be effective as it must cope with often severe attacks against the contents to protect, and efficient as it must rapidly spot the original contents from a huge reference collection.

4.2. Video database management

The existing video databases are generally little digitized. The progressive migration to digital television should quickly change this point. As a matter of fact, the French TV channel TF1 switched to an entirely digitized production, the cameras remaining the only analogical spot. Treatment, assembly and diffusion are digital. In addition, domestic digital decoders can, from now on, be equipped with hard disks allowing a storage initially modest, of ten hours of video, but larger in the long term, of a thousand of hours.

One can distinguish two types of digital files: private and professional files. On one hand, the files of private individuals include recordings of broadcasted programs and films recorded using digital camcorders. It is unlikely that users will rigorously manage such collections; thus, there is a need for tools to help the user: Automatic creation of summaries and synopses to allow finding information easily or to have within few minutes a general idea of a program. Even if the service is rustic, it is initially evaluated according to the added value brought to a system (video tape recorder, decoder), must remain not very expensive, but will benefit from a large diffusion.

On the other hand, these are professional files: TV channel archives, cineclubs, producers... These files are of a much larger size, but benefit from the attentive care of professionals of documentation and archiving. In this field, the systems can be much more expensive and are judged according to the profits of productivity and the assistance which they bring to archivists, journalists and users.

A crucial problem for many professionals is the need to produce documents in many formats for various terminals from the same raw material without multiplying the editing costs. The aim of such a *repurposing* is for example to produce a DVD, a web site or an alert service by mobile phone from a TV program at the minimum cost. The basic idea is to describe the documents in such a way that they can be easily manipulated and reconfigured easily.

4.3. Textual database management

Searching in large textual corpora has already been the topic of many researches. The current stakes are the management of very large volumes of data, the possibility to answer requests relating more on concepts than on simple inclusions of words in the texts, and the characterization of sets of texts.

We work on the exploitation of scientific bibliographical bases. The explosion of the number of scientific publications makes the retrieval of relevant data for a researcher a very difficult task. The generalization of document indexing in data banks did not solve the problem. The main difficulty is to choose the keywords, which will encircle a domain of interest. The statistical method used, the factorial analysis of correspondences, makes it possible to index the documents or a whole set of documents and to provide the list of the most discriminating keywords for these documents. The index validation is carried out by searching information in a database more general than the one used to build the index and by studying the retrieved documents. That in general makes it possible to still reduce the subset of words characterizing a field.

We also explore scientific documentary corpora to solve two different problems: to index the publications with the help of meta-keys and to identify the relevant publications in a large textual database. For that, we use factorial data analysis, which allows us to find the minimal sets of relevant words that we call meta-keys and to free the bibliographical search from the problems of noise and silence. The performances of factorial correspondence analysis are sharply greater than classic search by logical equation.

5. Software

5.1. Software

When applicable, we provide the IDDN is the official number, which is obtained when registering the software at the APP (Agence de Protection des Programmes).

5.1.1. New Software

5.1.1.1. Aabot

Participant: Jonathan Delhumeau.

AABOT is a tool to facilitate annotation of large video databases. It's primary design focus has been for the annotation on commercials in two 6-month long TV databases. The software keeps a database of already annotated commercials and suggests when it finds a new probable instance. It also validates user annotations by suggesting similar existing commercials if it finds any which are similar by name or content. The user can then confirm the creation of new commercials or accept the correction if he was mistaken.

AABOT is accessed via a web-browser. It is mostly used by uploading and downloading an annotation file. An interactive HTML5 interface is also available when some user feedback is needed (during validation). It uses Peyote as an description / indexation engine.

First APP deposit: IDDN.FR.001.4200010.000.S.P.2012.000.20900.

5.1.1.2. Peyote

Participants: Sébastien Campion, Jonathan Delhumeau [correspondent], Hervé Jégou.

Peyote is a framework for Video and Image description, indexation and nearest neighbor search. It can be used as-is by a video-search or image-search front-end with the implemented descriptors and search modules. It can also be used via scripting for large-scale experimentation. Finally, thanks to its modularity, it can be used for scientific experimentation on new descriptors or indexation methods. Peyote is used in the AABOT software and was used for the Mediaeval Placing task [68] and the Trecvid Instance Search task.

First APP deposit: IDDN.FR.001.4200008.000.S.P.2012.000.20900.

5.1.1.3. Watermarking Effective Key Length Evaluation

Participant: Teddy Furon [correspondent].

This software was developed in collaboration with Patrick BAS (CNRS, Ecole Centrale de Lille)

Weckle is a software suite in Matlab and R for the numerical evaluation of the effective key length of watermarking schemes based on Spread Spectrum, a concept which was proposed in [22], [23].

5.1.2. Most active software started before 2012

5.1.2.1. Babaz

Participants: Jonathan Delhumeau, Guillaume Gravier, Hervé Jégou [correspondent].

Babaz (http://babaz.gforge.inria.fr/) is a audio database management system with an audio-based search function, which is intended for audio-based search in video archives. First APP deposit: IDDN.FR.001.010006.000.S.P.2012.000.10000. It is licensed under the terms of the GNU General Public License v3.0.

5.1.2.2. Bigimbaz

Participant: Hervé Jégou [correspondent].

Bigimbaz is a platform originally developed in the LEAR project-team, and now co-maintained by TEXMEX. It integrates several contributions on image description and large-scale indexing: detectors, descriptors, retrieval using bag-of-words and inverted files, and geometric verification.

5.1.2.3. BonzaiBoost

Participant: Christian Raymond [correspondent].

The software homepage is available at http://bonzaiboost.gforge.inria.fr/.

BonzaiBoost stands for boosting over small decisions trees. BonzaiBoost is a general purpose machine-learning program based on decision tree and boosting for building a classifier from text and/or attribute-value data. Currently one configuration of BonzaiBoost is ranked first on http://mlcomp.org a website which propose to compare several classification algorithms on many different datasets

5.1.2.4. Irisa_Ne

Participant: Christian Raymond [correspondent].

IRISA_NE is a couple of Named Entity tagger, one of them is based on CRF and the other HMM. It is dedicated to automatic transcriptions of speech. It does not take into account uppercase or punctuation and has no concept of sentences. However, they also manage texts with punctuation and capitalization.

5.1.2.5. Nero

Participant: Sébastien Campion [correspondent].

The service is available at https://nero.irisa.fr.

NERO is an online Named Entities Recognition system. It is implemented within a web service that allows other member of the community to evaluate our results online without any client side setup. An HTTP Rest API, Shell and Python client are provided. The protocol used is HTTPS to secure the transactions between the user and the server. A user account is needed, which allow a fine monitoring. Usage are also limited to 100 thousand characters per account.

5.1.2.6. SURVA

Participants: Sébastien Campion [correspondent], Jonathan Delhumeau.

Speed Up Robust Video Alignement enables to quickly and efficiently synchronize the same video with two coding and quality formats (i.e. without the same number of frame). First APP deposit: IDDN.FR.001.420009.000.S.P.2012.000.20900.

5.1.2.7. PimPy

Participant: Sébastien Campion [correspondent].

PIMPY provides a convenient and high level API to manage common multimedia indexing tasks. It includes severals features. It is used, in particular

- to retrieve video features, such as histogram, binarized DCT descriptor, SIFT, SURF, etc ;
- to detect video cuts and dissolve (GoodShotDetector) ;
- for fast video frame access (pyffas) ;
- for raw frame extraction, or video segment extraction and re-encoding ;
- to search a video segment in another video (content based retrieval) ;
- to perform scene clustering.

First APP deposit: IDDN.FR.001.260038.000.S.P.2011.000.40000

5.1.2.8. Pqcodes

Participant: Hervé Jégou [correspondent].

This software is jointly maintained by Matthijs Douze, from Inria Grenoble.

Pqcodes is a library which implements the approximate k nearest neighbor search method of [83] based on product quantization. This software has been transferred to two companies (in August 2011 and May 2012, respectively).

The current version registered at the APP is IDDN.FR.001.220012.001.S.P.2010.000.10000.

5.1.2.9. Yael

Participant: Hervé Jégou [correspondent].

This software is jointly maintained by Matthijs Douze, from Inria Grenoble.

Yael is a C/python/Matlab library providing (multi-threaded, Blas/Lapack, low level optimization) implementations of computationally demanding functions. In particular, it provides very optimized functions for k-means clustering and exact nearest neighbor search. The library has been downloaded about 1000 times in 2012.

The current version registered at APP is IDDN.FR.001.220014.001.S.P.2010.000.10000.

5.1.2.10. IRISA News Topic Segmenter (irints)

Participants: Guillaume Gravier [correspondent], Camille Guinaudeau, Pascale Sébillot, Anca-Roxana Simon.

This software is dedicated to unsupervised topic segmentation of texts and transcripts. The software implements several of our research methods and is particularly adapted for automatic transcripts. It provides topic segmentation capabilities virtually for any word-based language, with presets for French, English and German. The software has been licensed to several of our industrial partners.

5.1.3. Other softwares

- BAG-OF-COLORS, implements a technique to describe the images based on color.
- I-DESCRIPTION. IDDN.FR.001.270047.000.S.P.2003.000.21000.
- ASARES is a symbolic machine learning system that automatically infers, from descriptions of pairs of linguistic elements found in a corpus in which the components are linked by a given semantic relation, corpus-specific morpho-syntactic and semantic patterns that convey the target relation. IDDN.FR.001.0032.000.S.C.2005.000.20900.
- ANAMORPHO detects morphological relations between words in many languages IDDN.FR.001.050022.000.S.P.2008.000.20900.
- DIVATEX is a audio/video frame server. IDDN.FR.001.320006.000.S.P.2006.000.40000,
- NAVITEX is a video annotation tool. IDDN.FR.001.190034.000.S.P.2007.000.40000,
- TELEMEX is a web service that enables TV and radio stream recording.
- VIDSIG computes a small and robust video signature (64 bits per image).
- VIDSEG computes segmentation features such as cuts, dissolves, silences in audio track, changes of ratio aspect, monochrome images. IDDN.FR.001.250009.000.S.P.2009.000.40000 ,
- ISEC, web application used as graphical interface for image searching engines based on retrieval by content.
- GPU-KMEANS, implementation of k-means algorithm on graphical process unit (graphic cards)
- CORRESPONDENCE ANALYSIS computes a factorial correspondence analysis (FCA) for image retrieval.
- GPU CORRESPONDENCE ANALYSIS is an implementation of the previous software Correspondence Analysis on graphical processing unit (graphical card).
- CAVIZ is an interactive graphical tool that allows to display and to extract knowledge from the results of a Correspondence Analysis on images.
- KIWI (standing for Keywords Extractor) is mostly dedicated to indexing and keyword extraction purposes.
- TOPIC SEGMENTER, is a software dedicated to topic segmentation of texts and (automatic) transcripts.
- S2E (Structuring Events Extractor) is a module which allows the automatic discovery of audiovisual structuring events in videos.
- 2PAC builds classes of words of similar meanings (“semantic classes“) specific to the use that is made of them in that given topic. IDDN.FR.001.470028.000.S.P.2006.000.40000
- FAESTOS (Fully Automatic Extraction of Sets of keywords for TOpic characterization and Spotting) is a tool composed of a sequence of statistical treatments that extracts from a morpho-syntactically tagged corpus sets of keywords that characterize the main topics that corpus deals with. IDDN.FR.001.470029.000.S.P.2006.000.40000
- FISHNET is an automatic web pages grabber associated with a specific theme.
- MATCH MAKER, semantic relation extraction by statistical methods.
- IRISAPHON produces phonetic words.
- PYTHON-GEOHASH is an implementation of the Geometric Hashing algorithm of [90] to check if geometrical consistency between pairs of images.
- AVSST is an Automatic Video Stream Structuring Tool. First, it allows the detection of repetitions in a TV stream. Second, a machine learning method allows the classification of programs and inter-programs such as advertisements, trailers, etc. Finally, the electronic program guide is synchronized with the right timestamps based on dynamic time warping. A graphical user interface is provided to manage the complete workflow.

- TVSEARCH is a content based retrieval search engine used to search and propagate manual annotation such as advertisement in a TV corpora.
- SAMUSA detects speech and/or musical segment in multimedia content.
- KERTRACK is a visual graphical interface for tracking visual targets based on particle filter tracking or based on mean-shift.
- MOZAIC2D creates of spatio-temporal mosaic based on dominant motion compensation.

5.2. Demonstration: Texmix

Participants: Morgan Bréhinier, Sébastien Campion [correspondent], Guillaume Gravier.

The gradual migration of television from broadcast diffusion to Internet diffusion offers tremendous possibilities for the generation of rich navigable contents. However, it also raises numerous scientific issues regarding de-linearization of TV streams and content enrichment. In this Texmix demonstration, we illustrate how speech in TV news shows can be exploited for de-linearization of the TV stream. In this context, de-linearization consists in automatically converting a collection of video files extracted from the TV stream into a navigable portal on the Internet where users can directly access specific stories or follow their evolution in an intuitive manner.

Structuring a collection of news shows requires some level of semantic understanding of the content in order to segment shows into their successive stories and to create links between stories in the collection, or between stories and related resources on the Web. Spoken material embedded in videos, accessible by means of automatic speech recognition, is a key feature to semantic description of video contents. We have developed multimedia content analysis technology combining automatic speech recognition, natural language processing and information retrieval to automatically create a fully navigable news portal from a collection of video files.

The demonstration was presented in several workshops (Futur en Seine - Paris, Futur TV - Berlin, ICMR - Hong Kong, French Minister for higher education and research - Rennes, RFIA - Lyon) and a video has been made available online on the portal of the EIT ICT Labs OpenSEM project. An article about this demonstrator was also published in 'Emergences' http://emergences.inria.fr/2012/newsletter-n22/L22-TEXMIX.

See the demo at http://texmix.irisa.fr.

5.3. Experimental platform

Participants: Laurent Amsaleg, Sébastien Campion [correspondent], Patrick Gros, Pascale Sébillot.

Until 2005, we used various computers to store our data and to carry out our experiments. In 2005, we began some work to specify and set-up dedicated equipment to experiment on very large collections of data. During 2006 and 2007, we specified, bought and installed our first complete platform. It is organized around a very large storage capacity (155TB), and contains 4 acquisition devices (for Digital Terrestrial TV), 3 video servers, and 15 computing servers partially included in the local cluster architecture (IGRIDA).

In 2008, we build up a corpus of multimedia data. It consists in a continuous recording (6 months) of two TV channels and three radios. It also includes web pages related to these contents captured on broadcaster's website. This corpus is to be used for different studies like the treatment of news along the time and to provide sub-corpus like TV news within the Quaero project (see below). The manual annotation of all the TV programs is under progress. A dedicated website has been developed in 2009 to provide a user support. It contains useful information such as references of available and ready to use software on the cluster, list of corpus stored on the platform, pages for monitoring disk space consumption and cluster loading, tutorials for best practices and cookbooks for treatments of large datasets. In 2010, we have acquired a new large memory server with 144GB of RAM which is used for memory demanding tasks, in particular to improve the speed of building index or language model. The previous server dedicated to this kind of jobs (acquired in 2008) has been upgraded to 96GB of RAM.

This year, we extended our storage capacity to 215TB and expanded our computing resources with two new large memory servers with 256GB of RAM for each of them.

This platform is funded by a joint effort of Inria, INSA Rennes and University of Rennes 1.

6. New Results

6.1. Description of multimedia content

6.1.1. Face Recognition

Participants: Thanh Toan Do, Ewa Kijak.

Face recognition is an important tool for many applications like video analysis. We addressed the problem of faces representation and proposed a weighted co-occurrence Histogram of Oriented Gradient as facial representation. The approach was evaluated on two typical face recognition datasets and has shown an improvement of the recognition rate over state of the art methods [31].

6.1.2. Violent scene detection

Participants: Guillaume Gravier, Patrick Gros, Cédric Penet.

Joint work with Technicolor.

We have worked on multimodal detection of violent scenes in Hollywood movies, in collaboration with Technicolor. Two main directions were explored. On the one hand, we investigated different kinds of Bayesian network structure learning algorithms for the fusion of multimodal features [49]. On the other hand, we studied the use of audio words for the detection of violent related events—gunshots, screams and explosions—in the soundtrack, demonstrating the benefit of product quantization and multiple words representations for increased robustness to variability between movies.

6.1.3. Text detection in videos

Participants: Khaoula Elagouni, Pascale Sébillot.

Joint work with Orange Labs.

Texts embedded in videos often provide high level semantic clues that can be used in several applications and services. We thus aim at designing efficient Optical Character Recognition (OCR) systems able to recognize these texts. In 2012, we proposed a novel approach that avoids the difficult step of character segmentation. Using a multi-scale scanning scheme, texts extracted from videos are first represented by sequences of features learnt by a convolutional neural network. The obtained representations fed a connectionist recurrent model, that relies on the combination of a BLSTM and a CTC connectionist classification model, specifically designed to take into account dependencies between successive learnt features and to recognize texts. The proposed video OCR, evaluated on a database of TV news videos, achieves very high recognition rates (character recognition rate: 97%; word recognition rate: 87%). Experiments also demonstrate that, for our recognition task, learnt feature representations perform better than standard hand-crafted features ([34]). We also carried out a comparison between two of our previous text recognition methods, one relying on a character segmentation step, the other one avoiding it by using a graph model, both on natural scene texts and embedded texts, highlighting the advantages and the limits of each of them. This work is submitted to the journal IJDAR.

6.1.4. Automatic speech recognition

Participants: Guillaume Gravier, Bogdan Ludusan.

This work was partly performed in the context of the Quaero project and the ANR project Attelage de Systèmes Hétérogènes (ANR-09-BLAN-0161-03), in collaboration with the METISS project-team.

In a multimedia context, automatic speech recognition (ASR) provides semantic access to multimedia but faces robustness issues due to the diversity of media sources. To increase robustness, we explore new paradigms for speech recognition based on collaborative decoding and phonetically driven decoding. We investigated mechanisms for the interaction of multiple ASR systems, exchanging linguistic information in a collaborative setting [15]. Following the same idea, we proposed phonetically driven decoding algorithms where the ASR system makes use of phonetic landmarks (place and manner of articulation, stress) to bias and prune the search space [65], [70]. In particular, we proposed a new classification approach to broad phonetic landmark detection [69].

6.2. Large scale indexing and classification techniques

6.2.1. Image retrieval and classification

Participants: Rachid BenMokhtar, Jonathan Delhumeau, Patrick Gros, Mihir Jain, Hervé Jégou, Josip Krapac.

This work was partially done in collaboration with Matthijs Douze and Cordelia Schmid (LEAR), Florent Perronnin and Jorge Sanchez (Xerox), Patrick Pérez (Technicolor) and Ondrej Chum (CVUT Prague). It was partly done in the context of the Quaero project.

Our work on very large scale image search has addressed [14] the joint optimization of three antinomic criterions: speed, memory resources and search quality. We have considered techniques aggregating local image descriptors into a vector and show that the Fisher kernel achieves better performance than the reference bag-of-visual words approach for any given vector dimension. The joint optimization of dimensionality reduction with indexing allowed us to obtain a precise vector comparison as well as a compact representation. The evaluation shows that the image representation can be reduced to a few dozen bytes while preserving high accuracy. Searching a 100 million image dataset takes about 250 ms on one processor core.

This work has been further improved [45] by modifying the way the similarity between images is computed, in particular we have shown that whitening is an effective way to fully exploit multiple vocabularies along with bag-of-visual words and VLAD representations.

We have also considered the problem of image classification, which goal is to produce a semantic representation of the images in the form of text labels reflecting the object categories contained in the images. We have proposed a technique derived from a matching system [44] based on Hamming Embedding and a similarity space mapping. The results outperform the state-of-the-art among matching systems such as NBNN. On some datasets such as Caltech-256, our results compare favorably to the best techniques, namely the Fisher vector representation.

6.2.2. Intensive use of SVM for text mining and image mining

Participants: Thanh Nghi Doan, François Poulet.

Following our previous work on large scale image classification [58], we have developed a fast and efficient framework for large scale image classification. Most of the state of the art approaches use a linear SVM (eg LIBLINEAR) for the training task. Another solution can be to use the new Power Mean SVM (PmSVM) with power mean kernel functions that can solve a binary classification problem with millions of examples and tens of thousands of dense features in a few seconds (excluding the time to read the input files). We are working on a parallel version of this algorithm and trying to deal with unbalanced datasets: in ImageNet1000 dataset, there are 1,000 classes, this is a very unbalanced classification task so we use a balanced bagging parallel algorithm. The time needed to perform the training task on ImageNet1000 was almost 1 day with the original PmSVM algorithm and 2.5 days for LIBLINEAR, we achieve it within 10 min and with a relative precision increase of more than 20%. We are currently working to reduce the RAM needed to perform the task (today 30GB).

6.2.3. Audio indexing

Participants: Jonathan Delhumeau, Guillaume Gravier, Patrick Gros, Hervé Jégou.

This work was done in the context of the Quaero project.

Our new Babaz audio search system [46] aims at finding modified audio segments in large databases of music or video tracks. It is based on an efficient audio feature matching system which exploits the reciprocal nearest neighbors to produce a per-match similarity score. Temporal consistency is taken into account based on the audio matches, and boundary estimation allows the precise localization of the matching segments. The method is mainly intended for video retrieval based on their audio track, as typically evaluated in the copy detection task of Trecvid evaluation campaigns. The evaluation conducted on music retrieval shows that our system is comparable to a reference audio fingerprinting system for music retrieval, and significantly outperforms it on audio-based video retrieval, as shown by our experiments conducted on the dataset used in the copy detection task of the Trecvid'2010 campaign, which was used as an external evaluation in the Quaero project.

6.2.4. Approximate nearest neighbor search with compact codes

Participants: Teddy Furon, Hervé Jégou.

This work was done in collaboration with the Metiss project team (Anthony Bourrier and Rémi Gribonval). It was partly done in the context of the Quaero project.

Following recent works on Hamming Embedding techniques, we proposed [47] a binarization method that aim at addressing the problem of nearest neighbor search for the Euclidean metric by mapping the original vectors into binary vectors ones, which are compact in memory, and for which the distance computation is more efficient. Our method is based on the recent concept of anti-sparse coding, which exhibits here excellent performance for approximate nearest neighbor search. Unlike other binarization schemes, this framework allows, up to a scaling factor, the explicit reconstruction from the binary representation of the original vector. We also show that random projections which are used in Locality Sensitive Hashing algorithms, are significantly outperformed by regular frames for both synthetic and real data if the number of bits exceeds the vector dimensionality, i.e., when high precision is required.

Another aspect we have investigated in this line of research is the problem of efficient nearest neighbor search for arbitrary kernels. For this purpose, we have combined [76] the product quantization technique [4] with explicit embeddings, and showed that this solution significantly outperforms the state-of-the-art technique designed for arbitrary kernels, such as Kernelized Locality Sensitive Hashing. In addition, we have proposed a variant to perform the exact search.

6.2.5. Indexing and searching large image collections with map-reduce

Participants: Laurent Amsaleg, Gylfi Gudmundsson.

This work was done in the context of the Quaero project.

Most researchers working on high-dimensional indexing agree on the following three trends: (i) the size of the multimedia collections to index are now reaching millions if not billions of items, (ii) the computers we use every day now come with multiple cores and (iii) hardware becomes more available, thanks to easier access to Grids and/or Clouds. This work shows how the Map-Reduce paradigm can be applied to indexing algorithms and demonstrates that great scalability can be achieved using Hadoop, a popular Map-Reduce-based framework. Dramatic performance improvements are not however guaranteed a priori: Such frameworks are rigid, they severely constrain the possible access patterns to data and the RAM memory has to be shared. Furthermore, algorithms require major redesign, and may have to settle for sub-optimal behavior. The benefits, however, are numerous: Simplicity for programmers, automatic distribution, fault tolerance, failure detection and automatic re-runs and, last but not least, scalability. We report our experience of adapting a clustering-based high-dimensional indexing algorithm to the Map-Reduce model, and of testing it at large scale with Hadoop as we index 30 billion SIFT descriptors. We draw several lessons from this work that could minimize time, effort and energy invested by other researchers and practitioners working in similar directions.

6.2.6. Vectorization

Participant: Vincent Claveau.

The vectorization principle allows the description of any object in a vector space based on its similarity with pivots objects. During the last years, we have shown that such a technique can be successfully used for Information Retrieval or Topic Segmentation. This year, TexMex has demonstrated how it can be used in a pure data-mining framework by participating to the JRS2012 framework. The task proposed was a high-dimensional and multi-class machine learning problem. Our approach, based on a simple kNN using vectorization has proved its interest, since it was ranked in top-methods while requiring no training phase nor complex setting.

6.3. Security of media

6.3.1. Security of content based image retrieval

Participants: Laurent Amsaleg, Thanh Toan Do, Teddy Furon, Ewa Kijak.

The performance of Content-Based Image Retrieval Systems (CBIRS) is typically evaluated via benchmarking their capacity to match images despite various generic distortions such as cropping, rescaling or Picture in Picture (PiP) attacks, which are the most challenging. Distortions are made in a very generic manner, by applying a set of transformations that are completely independent from the systems later performing recognition tasks. Recently, studies have shown that exploiting the finest details of the various techniques used in a CBIRS offers the opportunity to create distortions that dramatically reduce the recognition performance [30]. Such a *security perspective* is taken in our work. Instead of creating generic PiP distortions, we have proposed a creation scheme able to delude the recognition capabilities of a CBIRS that is representative of state of the art techniques as it relies on SIFT, high-dimensional k-nearest neighbors searches and geometrical robustification steps. We have ran experiments using 100,000 real-world images confirming the effectiveness of these security-oriented PiP visual modifications [29]. This work goes together with the completed PhD of Thanh-Toan Do [8].

6.3.2. The concept of effective key length in watermarking

Participant: Teddy Furon.

Whereas the embedding distortion, the payload and the robustness of digital watermarking schemes are well understood, the notion of security is still not completely well defined. The approach proposed in the last five years is too theoretical and solely considers the embedding process, which is half of the watermarking scheme. In collaboration with Patrick BAS (CNRS, Ecole Centrale de Lille), we propose a new measure of watermarking security. This concept is called the *effective key length*, and it captures the difficulty for the adversary to get access to the watermarking channel: The adversary proposes a test key and the security is measured as the probability that this test key grants him the watermarking channel (he succeeds to decode hidden messages).

This new methodology is applied to the most wide spread watermarking schemes where theoretical and practical computations of the effective key length are proposed: Zero-bit 'Broken Arrows' technique [22], spread spectrum (SS) based schemes (like additive SS, improved SS, and correlation aware SS) [23], and quantization index modulation (QIM) scheme (like Distortion Compensated QIM) [38]. A journal article about this new concept has been submitted to IEEE Trans. on Information Forensics and Security. The keystone of the approach is the evaluation of a security level to the estimation of a probability. Experimental protocols using rare event probability estimator allow good evaluation of this quantity. The soundness of this latter estimator has been theoretically proven in [11] (collaboration with Inria team-project ALEA and ASPI).

6.3.3. A practical joint decoder for active fingerprinting

Participant: Teddy Furon.

This work deals with active fingerprinting, a.k.a. traitor tracing. A robust watermarking technique embeds the user's codeword into the content to be distributed. When a pirated copy of the content is scouted, the watermark decoder extracts the message, which identifies the dishonest user. However, there might exist a group of dishonest users, so called collusion, who mix their personal versions of the content to forge the pirated copy. The extracted message no longer corresponds to the codeword of one user, but is a mix of several codewords. The decoder aims at finding back some of these codewords to identify the colluders, while avoiding accusing innocent users.

This work follows our breakthrough on Tardos code joint decoding, mentioned in last year's activity report, and whose journal version has been published this year in [16]. Information theory proves that a joint decoder computing scores for pairs, triplets, or in general t-tuples of users is more powerful than single decoders working with scores for single users. However, nobody did try them for large scale setups since the number of t-tuples is in $O(n^t)$. In practical scenarios, n is at least 10,000 and t is around 10, which implies the computation of $\sim 10^{40}$ scores. Last year, we were the first team to design an approximate joint decoder. If its complexity was well under control (in $O(n)$), its iterative structure was much intricate.

Our new design of joint decoder is based on the Monte-Carlo Markov Chain method. It is a simpler iterative process allowing us to sample collusion subsets according to the A Posteriori distribution. Then, the probability that user j is guilty is empirically evaluated over this sample, and threshold to yield a reliable decision. This work has been done under a collaboration with Inria team-project ASPI, and published in [39].

6.4. Multimedia content structuring

6.4.1. Motif discovery

Participants: Guillaume Gravier, Hervé Jégou, Anh Phuong Ta, Wanlei Zhao.

This work was done in the context of the Quaero project.

We have pursued our work on unsupervised discovery of repeating motifs in multimedia data along three directions:

- Discovery of multiple recurrent audio-visually consistent sequences: We proposed two unsupervised approaches to automatically detect multiple structural events in videos using audio and visual modalities. Both approaches rely on cross-modal cluster analysis techniques to directly define events from the data without any prior assumption [51], [52].
- Large-scale unsupervised discovery of near-duplicate shots in TV streams: We developed an efficient method with little a priori knowledge which relies on a product k-means quantizer to efficiently produce hash keys adapted to the data distribution of the frame descriptors. This hashing technique combined with a temporal consistency check allows the detection of meaningful repetitions in TV streams [54].
- Audio motif discovery: This joint work with the METISS project-team extends the generic audio motif discovery method developed in the Ph. D. thesis of Armando Muscariello [17]. We developed an efficient implementation, which will be made publicly available. The software was benchmarked on near duplicate audio motif discovery in the framework of the Quaero project.

6.4.2. Stream labeling for TV structuring

Participants: Vincent Claveau, Guillaume Gravier, Patrick Gros, Emmanuelle Martienne, Abir Ncibi.

In this application, we focus on the problem of labeling the segments of a TV stream according to their types (*e.g.,*, programs, commercial breaks, sponsoring, ...). During this year, we performed an in-depth analysis of the use of Conditional Random Fields (CRF) for our task. In particular, we studied:

- how sequentiality is modeled with the CRF;
- the links with other probabilistic graphical techniques (HMM, MEMM...);
- the robustness of the approach when dealing with few training data or few features;

The use of this model for semi-supervised and unsupervised learning are under study. We also studied the use of very simple descriptors (simple shot lengths, and use of global image descriptors only to complete the results) in order to fasten the initial repetition detection stage. This allows us to process 6 months of TV in a few minutes.

6.4.3. Multimedia browsing

Participant: Laurent Amsaleg.

Traditionally, research in multimedia has focused primarily on analyzing and understanding the contents of media documents, by defining clever ways to extract relevant information from the multimedia files, thereby hoping to eventually bridge the semantic gap. We have observed that much of the research in multimedia is trying to *link* the information automatically extracted from the contents to create a meaningful user-experience. Most of the state-of-art solutions are very ad-hoc, and we believe that multimedia is lacking a powerful and flexible data model where multimedia data (ranging from entire documents to elements automatically extracted from the contents such as faces, scenes, objects, ...) can be appropriately represented as well as the relationships between data items. Instead, we propose a multi-dimensional model for media browsing, called ObjectCube, based on the multi-dimensional model commonly used in On-Line Analytical Processing (OLAP) applications. This model has been implemented in a prototype called ObjectCube, and its performance evaluated using personal photo collections of up to one million images. We also worked on exposing plug-in API for image analysis and browsing methods, facilitating the use of the prototype and its model as a demonstration platform.

6.4.4. Video summarization

Participants: Mohamed-Haykel Boukadida, Patrick Gros.

Joint work with Orange labs.

Up to now, most video summarization methods are based on concepts like saliency and often use a single modality. In order to develop a more general framework, we propose to use a constraint programming approach, where summarizing a video is seen as a constraint resolution problem, which consists in choosing certains excerpts with respect to various criteria. This first year of work on the topic was mainly devoted to discover the abilities of Choco, a constraint solver, and to study how summarization can be formulated as a constraint resolution problem.

6.4.5. Graph organization of large scale news archives

Participants: Guillaume Gravier, Ludivine Kuznik, Pascale Sébillot.

This work is done in collaboration with Jean Carrive at Institut National de l'Audiovisuel in the framework of a joint Ph. D. thesis within the Quaero project.

The idea of this work is to automatically create links and threads between reports in several years of broadcast news shows, based either on the documentary records of the shows and/or on the automatic transcripts. We studied how standard information retrieval measures of similarity can be used to build an epsilon-nearest neighbor graph from the various fields of the documentary records. Depending on the field used (title, keywords from a thesaurus, summary, speech transcript) and the metrics, different types of clusters can be obtained in the graph. We proposed metrics mimicking recall and precision on documents to analyze the graphs obtained and quantify the potential interest of various graph construction strategies for topic threading.

6.5. Language processing in multimedia

6.5.1. Lexical-phonetic automata for spoken utterance indexing and retrieval

Participants: Julien Fayolle, Guillaume Gravier, Fabienne Moreau, Christian Raymond.

This work was partly done in the context of the Quaero project.

Spoken content retrieval relies on the fields of automatic speech recognition and information retrieval (IR). However, IR tools made for text are not adapted to automatic transcripts which are particularly incomplete and uncertain. Even if in-vocabulary words are usually well-recognized, these transcripts contain many recognition errors affecting notably out-of-vocabulary words and named entities that convey important discourse information (e.g., person names, localizations, organizations) necessary for IR. This year, we have proposed a method for indexing spoken utterances which combines lexical and phonetic hypotheses in a hybrid index built from automata [35], [36]. The retrieval is performed by a lexical-phonetic and semi-imperfect matching whose aim is to improve the recall. A feature vector, containing edit distance scores and a confidence measure, weights each transition to help the filtering of the candidate utterance list for a more precise search. We have demonstrated the complementarity of the lexical and phonetic levels (extracted from the 1-best speech recognition hypothesis) and the advantage of using a hybrid index, a semi-imperfect matching and a supervised filtering (combining edit distance scores and a confidence measure).

6.5.2. Information extraction and text mining

Participants: Ali Reza Ebadat, Vincent Claveau, Pascale Sébillot.

This work was partly done in the framework of the Quaero project.

In the framework of Ali-Reza Ebadat's thesis on information extraction for multimedia analysis, we have investigated techniques for robust text-mining on texts or speech transcripts. We have developed several supervised models:

- entity detection and entity classification; the goal is to detect, into a text, pre-defined categories of entities and to label them accordingly. The techniques that we developed cascade chunk parsing with simple classification tools, resulting in a very efficient and simple to train NLP tool.
- relation detection; this model relies on k-NN approach with a language-modeling based distance. Since it relies on surface elements, it can handle noisy data such as speech transcripts.

We have also developed unsupervised models for information discovery:

- entity clustering; the goal is to detect and group, without a priori knowledge, entities. We have shown that weighting techniques used in information retrieval can be used as relevant features to describe the entity.
- relation clustering: as for entity, the goal is to group relations (that is, pairs of entities) without a priori or pre-defined categories. Our approach is pioneer is this field and relies on clustering with language-modeling based distances.

Some of these models have been evaluated in the framework of the Quaero evaluation campaign and TexMex ranked first in three of the tracks (entity detection and categorization) and close second in the last one (relation detection and categorization).

6.5.3. Morphological analysis for information retrieval

Participants: Vincent Claveau, Ewa Kijak.

In the biomedical field, the key to access information is the use of specialized terms (like *photochemotherapy*). These complex morphological structures may prevent a user querying for *gastrodynia* to retrieve texts containing *stomachalgia*. In that context, we have developed a new unsupervised technique to identify the various meaningful components of these terms and use this analysis to improve biomedical information retrieval. Our approach combines an automatic alignment using a pivot language, and an analogical learning that allows an accurate morphological analysis of terms. We ave shown that these morphological analyses can be used to greatly improve the indexing of medical documents.

6.5.4. Unsupervised hierarchical topic segmentation

Participants: Guillaume Gravier, Pascale Sébillot, Anca-Roxana Simon.

Linear topic segmentation has been widely studied for textual data and recently adapted to spoken contents. However, most documents exhibit a hierarchy of topics which cannot be recovered using linear segmentation. We investigated hierarchical topic segmentation of TV programs exploiting the spoken material. Recursively applying linear segmentation methods is one solution but fails at the lowest levels of the hierarchy when small segments are targeted, in particular when transcription errors jeopardize lexical cohesion. To skirt these issues, we investigated the use of indirect comparison between segments via vectorization techniques at the lower level of the hierarchy, using simple segmentation methods based on TextTiling. Results were similar to those obtained by the recursive use of a more elaborate probabilistic topic segmentation method. Future work will focus on using indirect comparison within the probabilistic framework.

6.6. Competitions and international evaluation campaign

6.6.1. Mediaeval's affect task: Violent scenes detection task

Participants: Guillaume Gravier, Patrick Gros, Cédric Penet.

The project-team participated in the Affect Task of the MediaEval 2012 benchmark, both as part of the organizing team and as competitor [64], [67].

6.6.2. Mediaeval's placing task: Geo-localization of videos

Participants: Jonathan Delhumeau, Guillaume Gravier, Hervé Jégou, Michele Trevisiol.

This work was partly done in the context of the Quaero project.

We developed an efficient and effective system to identify the geographic location of videos using a multimodal cascade of techniques exploiting all available sources of information, from user assigned tags to user data and image content. We also proposed a novel hierarchical strategy to exploit tags using information retrieval techniques. A coarse geographic area is first identified before refining the search to find exact geo-coordinates. Area and coordinates are obtained from a vector space representation of the tags using appropriate weighting and normalization [68].

We participated in the Placing Task of the MediaEval 2012 benchmark, where we ranked first on one of the mandatory runs (no gazeteers, no dictionary).

6.6.3. Mediaeval: Search & hyperlinking

Participants: Guillaume Gravier, Camille Guinaudeau, Pascale Sébillot.

We participated in the Search and Hyperlinking task proposed in the framework of the MediaEval benchmark initiative in 2012. We developed a solution for the hyperlinking subtask in which participants were required to return a ranked list of video segments potentially relevant to the answer provided for an initial query, thus creating links between video segments.

Our solution, based on information retrieval techniques, implements two separate module: The retrieval of relevant videos, followed by the selection of short segments specifically corresponding to the information need. First, the hyperlinking module computes the similarity between a video segment query and the collection of videos and returns a ranked list of relevant videos. We investigated several parameterization and ranking strategies. In the second step, we extract from each video the segment that is the closest, from a meaning point of view, to the video segment query, using topic segmentation methods [42].

Our system ranked either first or second depending on the evaluation conditions.

6.6.4. ETAPE named entities evaluation campaign

Participant: Christian Raymond.

Christian Raymond participated to the ETAPE Named Entities evaluation campaign. The goal was to propose a system able to tag NE following the new tree-stuctured NE definition given in the Quaero project. The evaluation has been done on manual and 5 automatic transcriptions of french TV and Radio shows produced by 5 different automatic speech recognition systems. The system was ranked first with results far better than those of the other participating systems.

6.6.5. DEFT evaluation campaign participation

Participants: Vincent Claveau, Christian Raymond.

Christian Raymond and Vincent Claveau participated to DEFT. The task proposed was to work on a corpus of scientific papers, by focusing the work on the issue of indexing the scientific papers: identifying the keywords chosen by the authors to index their paper, considering both abstract and whole article. Two tasks were proposed which led them to test two different strategies . For the first task, a list of keywords was provided. Based on that, our first strategy is to consider that as an Information Retrieval problem in which the keywords are the queries that are attributed to the best ranked documents. This approach yielded very good results. For the second task, only the articles were known. For this task, our approach is mainly based on a term extraction system whose results are reordered by a machine learning [27] technique.

6.6.6. Trecvid: Multimedia Indexing task

Participants: Jonathan Delhumeau, Philippe-Henri Gosselin, Hervé Jégou.

This work was partly done in the context of the Quaero project.

Texmex has taking part to the Quaero [50] and IRIM [21] submissions of Trecvid in the Multimedia indexing task, by providing some state-of-the-art image descriptors and collaborating with the LIG to set up the dimensionality reduction tool for high-dimensional vectors. The Quaero Rank was ranked 3rd in the full task (1st amongst European submissions).

7. Bilateral Contracts and Grants with Industry

7.1. Bilateral Contracts with Industry

7.1.1. Consulting agreement with the LTU company

Participant: Hervé Jégou.

Following the interaction in the context of the Quaero Project, LTU and Inria have signed a consulting agreement. In this context, Hervé Jégou has provided some expertise to the company in March 2012.

7.2. Bilateral Grants with Industry

7.2.1. Contract with Orange Labs

Participants: Pascale Sébillot, Khaoula Elagouni.

Duration: 36 months, since October 2009.

K. Elagouni's Ph.D. thesis is supported by a CIFRE grant in the framework of a contract between Orange Labs and TEXMEX. The aim of the work is to investigate a more semantic approach to describe multimedia documents based on textual material found inside the images.

7.2.2. Contract with INA (Institut national de l'audiovisuel)

Participants: Guillaume Gravier, Ludivine Kuznik, Pascale Sébillot.

Duration: 36 months, since April 2011.

Ludivine Kuznik's Ph.D. thesis is supported by a CIFRE grant in the framework of a contract between INA and TEXMEX within the OSEO/QUAERO project. The aim of the work is to investigate a more semantic approach to structure and navigate very large collections of TV archives.

7.2.3. Contract with Orange Labs

Participants: Patrick Gros, Mohamed-Haykel Boukadida.

Duration: 36 months, since January 2012.

M.H. Boukadida's Ph.D. thesis is supported by a CIFRE grant in the framework of a contract between Orange Labs and TEXMEX. The aim of the work is to investigate the use of constraint programming to define a general framework for video summarization and repurposing.

7.2.4. Contract with INA

Participants: Guillaume Gravier, Bingjing Qu.

Duration: 36 months, since May 2012.

Bingjing Qu's Ph.D. thesis is supported by a CIFRE grant in the framework of a contract between INA and TEXMEX. The aim of the work is to infer the structure of collections of homogeneous documents.

7.2.5. Contract with Technicolor

Participants: Cédric Penet, Guillaume Gravier, Patrick Gros.

Duration: 36 months, since October 2010.

C. Penet's Ph.D. thesis is supported by a CIFRE grant in the framework of a contract between Technicolor and TEXMEX. The aim of the work is to develop methods to detect audio events and to apply these techniques to violence detection in films.

8. Partnerships and Cooperations

8.1. National Initiatives

8.1.1. ANR Attelage de systèmes hétérogènes

Participants: Guillaume Gravier, Bogdan Ludusan.

Duration: 3 years, started in November 2009.
Partners: IRISA, LIA, LIUM

The project ASH (Automatic System Harnessing – ANR-09-BLAN-0161-03) aims at developing new collaborative paradigms for speech recognition. Many current ASR systems rely on an a posteriori combination of the output of several systems (e.g., confusion network combination). In the ASH project, we investigate new approaches in which three ASR systems work in parallel, exchanging information at every step of the recognition process rather than limiting ourselves to an a posteriori combination. What information is to be shared and how to share such information and make use of it are the key questions that the project is addressing. The collaborative paradigm is being extended to landmark-based speech recognition where detection of landmarks and speech transcription can be considered as two (or more) collaborative processes.

8.1.2. ANR FIRE-ID

Participants: Sébastien Campion, Philippe-Henri Gosselin, Patrick Gros, Hervé Jégou.

Duration: 3 years, started in May 2012.
Partner: Xerox Research Center Europe

The FIRE-ID project considers the semantic annotation of visual content, such as photos or videos shared on social networks, or images captured by video surveillance devices or scanned documents. More specifically, the project considers the fine-grained recognition problem, where the number of classes is large and where classes are visually similar, for instance animals, products, vehicles or document forms. We also assumed that the amount of annotated data available per class for the learning stage is limited.

8.1.3. ANR Secular

Participants: Laurent Amsaleg, Teddy Furon, Benjamin Mathon, Ewa Kijak.

Duration: 3 years, started in September 2012.
Partners: Morpho, Univ. Caen GREYC, Telecom ParisTech, Inria Rennes

Since their invention, content based image retrieval systems (CBRS) and biometric systems have evolved separately. This is due to the fact that they originate from different research and industrial communities. This Basic Research project, called SecuLar, groups researchers from both communities who have observed that both type of systems have indeed a lot in common in terms of goals and technological blocks. These techniques are used, however, in quite different settings possibly explaining the gap between the two. The people involved in this SecuLar project believe that what is specific to each family of approach can now benefit the other for the two following fundamental reasons.

Biometrics needs scale. The size of biometric databases quickly increases. It grows in terms of the number of records kept in the database. It also grows in terms of the size of each record as larger biometric templates maintain high quality recognition. The amount of data becomes large enough to require powerful indexing techniques. CBRS are good at this as they allow ultra fast searches of nearest neighbours in huge datasets. But porting these techniques to a biometric context is far from being easy. Biometric databases are typically protected to enforce confidentiality and privacy as security is paramount. Indexing biometric data is thus difficult because the techniques enforcing security in biometrics conflict with the technique bringing efficiency to database searches. No biometric system can today cope with both all the privacy and security constraints and the scale at which they should work in the real world for new applications.

CBRS need security and privacy. We witness a new use of CBRS these days. CBRS become the main multimedia security technology to enforce copyright laws (content monetization) or to spot illegal contents (detection of copies, paedophile images, ...) over the Internet. However, they were not designed with privacy, confidentiality and security in mind. This comes in serious conflict with their use in these new security-oriented applications. Privacy is endangered due to information leaks when correlating users, queries and the contents stored-in- the-clear in the database. It is especially the case of images containing faces which are so popular in social networks. Biometrics systems have long relied on protection techniques and anonymization processes that have never been used in the context of CBRS. Here, we plan to understand how biometrics related techniques can help increasing the security levels of CBIRS while not degrading their performance.

8.2. European Initiatives

8.2.1. Quaero

Participants: Laurent Amsaleg, Sébastien Campion, Vincent Claveau, Ali Reza Ebadat, Julien Fayolle, Patrick Gros, Gylfi Gudmundsson, Camille Guinaudeau, Carryn Hayward, Hervé Jégou, Ewa Kijak, Fabienne Moreau, Christian Raymond, Pascale Sébillot.

Duration: 5 years, starting in May 2008. Prime: Technicolor.

Quaero is a large research and applicative program in the field of multimedia description (ranging from text to speech and video) and search engines. It groups 5 application projects, a joint Core Technology Cluster developing and providing advanced technologies to the application projects, and a Corpus project in charge of providing the necessary data to develop and evaluate the technologies. The large scope of QUAERO's ambitious objectives allows it to take full advantage of Texmex's many areas of research, through its tasks on: Indexing Multimedia Objects, Term Acquisition and Recognition, Semantic Annotation, Video Segmentation, Multi-modal Video Structuring, Image and video fingerprinting.

In 2012, a key fact is our strong participation to Mediaeval to evaluate the technologies developed in Quaero.

8.3. International Initiatives

8.3.1. Participation in International Programs

Participants: Patrick Gros, Guillaume Gravier.

Duration: 2 years

Collaboration Inria-FAPEMIG with PUC Minas and UFMG – Brazil

The collaboration started this year with a visit of Patrick Gros to Belo Horizonte. The thesis of a brazilian student, Bruno Teixeira, will be co-advised, and he will spend 6 months in France next year. His work focuses on video high level description for video classification.

8.4. International Research Visitors

8.4.1. Visits of International Scientists

- **Visit of Fabio Guimaraes, 1 week in March 2012.** This visit was the opportunity to launch our collaboration with Brazil, which will take place in the framework of the Inria-FAPEMIG program. The main topic of the collaboration will be video multimodal description.
- **Visit of Michael Houle, National Institute of Informatics, Tokyo, Japan.** This visit was dedicated to share knowledge and initiate a collaboration for high-dimensional indexing.

8.4.2. Internships

- Michele Trevisiol

  Dates: May 2012–July 2013 (3 months)

  Subject: Geo-Tagging of Flickr videos, evaluated in the context of the Mediaeval's Placing task.

  Institution: Yahoo Research & Universitat Pompeu Fabra (Barcelona)
- Giorgos Tolias

  Dates: October 2012–January 2013 (5 months)

  Subject: Large scale visual search

  Institution: National Technical University of Athens (Greece)

9. Dissemination

9.1. Scientific Animation

- Laurent Amsaleg
  - was a program committee member of CBMI 2012;
  - was a program committee member of ICMR 2012;
  - was a program committee member of MMM 2012;
  - was a program committee member of BDA 2012;
  - was a program committee member of ICME 2012;
  - was a program committee member of MMM 2013;
  - was a member of the "commission de spécialistes, ISTIC, University Rennes 1";
  - was a reviewer for Advances in Multimedia Journal.
- Vincent Claveau
  - was a program committee member of the Hybrid methods in NLP Workshop, collocated with EACL2012, Avignon, France;
  - was a program committee member of TALN'12 (18^{e} conférence nationale Traitement automatique des langues naturelles), Grenoble, France;
  - was a program committee member of IEEE/WIC/ACM Web Intelligence (WI-IAT) conference, Macau, China;
  - was a program committee member of Conférence en Recherche d'Information et Applications, CORIA 2012, Bordeaux, France;
  - is a member of the editorial board of the journal TAL, Traitement Automatique des Langues;
  - was a reviewing committee member for the journal BioInformatics;

  - is a board member of Association pour le Recherche d'Information et Application (ARIA) and was implied in the organization of the CORIA 2012 conference and the EARIA 2012 summer school
  - was a selection committee member in Univ. Paris 11 and Univ. Paris 13.
- Teddy Furon
  - was an associate editor of Elsevier Digital Signal Processing, 2012;
  - was an associate editor of IEEE Trans. on Information Forensics and Security, 2012;
  - was a member of the technical committee of IEEE Information Forensics and Security sub-society, 2012.
  - was a program committee member of Information Hiding 2012, Berkeley, CA, USA;
  - was a program committee member of CMS 2012, Canterbury, UK;
  - was a program committee member of IEEE WIFS 2012, Tenerife, Spain;
- Guillaume Gravier
  - was a program committee member of the Journées d'étude sur la Parole, 2012;
  - was a technical program committee member of Workshop on Content-based Multimedia Indexing, 2012;
  - was a technical program committee member of IEEE Intl. Conf. on Multimedia and Exhibition, 2012;
  - was a program committee member of the ECCV Workshop on Information Fusion in Computer Vision for Concept Recognition, 2012;
  - founded in 2012 the Special Interest Group of the Intl. Speech Communication Association, Speech and Language in Multimedia (SLIM), which he is leading;
  - is vice-president of the Association Francophone de la Communication Parlée, the French-speaking Speech Communication Association, also a regional group of the Intl. Speech Communication Association;
  - was a program committee member of the ACM Multimedia 2012 Workshop on Multimedia Analysis For Ecological Data;
  - was a technical program committee member of the ACM Worskhop on Audio and Multimedia Methods for Large-Scale Video Analysis, 2012;
  - was a program committee member of the 2012 5th International Congress on Image and Signal Processing;
  - was appointed as an expert for the ANR CONTINT program;
  - is a member of the Interspeech 2013 Steering Committee, in charge of the conference program coordination;
  - was the scientific leader of the national evaluation initiative ETAPE;
  - was a member of the organizing committee of the MediaEval 2012 evaluation benchmark;
  - has served as a reviewer for several major journals in the speech and multimedia domains.
- Patrick Gros
  - was a program committee member of the Fourth International Conference on Creative Content Technologies, Content 2012, Nice, France;
  - was a program committee member of Conférence en Recherche d'Information et Applications, CORIA 2012, Bordeaux, France;
  - was a Technical Program Committee member of the 20th European Signal Processing Conference Eusipso 2012, Bucharest, Romania;

  - was a program committee member of the 18th Conférence en Reconnaissance des Formes et Intelligence Artificielle, RFIA 2012, Lyon, France;
  - was appointed as scientific officer of the Inria research center of Rennes – Bretagne Atlantique;
  - was member of the scientific board of Université européenne de Bretagne;
  - was member of the Evaluation Board of Inria.
- Hervé Jégou
  - has co-organized a tutorial on Large-Scale Visual Recognition at CVPR'2012;
  - was a program committee member and keynote speaker at the Workshop on Web-scale Vision and Social Media, in conjunction with ECCV'2012;
  - was area chair for the BMVC'2012 conference;
  - was a program committee member of CVPR'2012;
  - was a program committee member of ECCV'2012;
  - was a program committee member of CBMI'2012;
  - was a program committee member of MMM'2012;
  - was a program committee member of RFIA'2012;
  - was appointed as an expert for the ANR CONTINT'2012 project call and for the ANRT in 2012;
  - was a member of the selection committee for associate professors at ENSEA.
- Ewa Kijak
  - is head of the Image engineering track of ESIR (Ecole Supérieure d'Ingénieur de Rennes, associated with University of Rennes 1).
- François Poulet
  - co-organized and edited, together with B. Le Grand, the proceedings of the 10^{th} Workshop Visualisation et Extraction de Connaissances co-located with Extraction et Gestion de Connaissances, (EGC'12), Bordeaux, France, Jan. 2012.
  - was a program committee member of KDIR'12, International Conference on Knwoledge Discovery and Information Retrieval, Barcelona, Spain, Oct. 2012;
  - was a program committee member of KICSS'12, International Conference on Knowledge Information and Creativity Support Systems), Melbourne, Australia, Nov. 2012;
  - was a program committee member of VINCI'12, Visual INformation Communications International, Hangzhou, China, Sept. 2012;
  - was a program committee member of AusDM'12, Autralasian Data Mining Conference, Sydney, Australia, Dec. 2012;
  - was a program committee member of EGC'12, Extraction et Gestion de Connaissances, Bordeaux, France, Jan. 2012;
  - was a reviewer of ISCAS 2013, IEEE International Symposium on Circuits and Systems, Beijing, China, May 2013;
  - was a reviewer of DAMI, Journal on Data Mining and Knowledge Discovery, Springer;
  - was a reviewer of Advances in Knowledge Discovery and Management, Springer;
  - was co-organizer of the 10th workshop Visualisation et Extraction de Connaissances, (AVEC-EGC'12), Bordeaux, France, January 2012.
- Christian Raymond
  - is a member of the editorial board of the e-journal "Discours";

  - was a reviewing committee member of Interspeech (13th Annual Conference of the International Speech Communication Association);
  - was a reviewing committee member of ICMLA (The tenth International Conference on Machine Learning and Applications).
- Pascale Sébillot
  - was a member of the program committee of LREC 2012 (8th international conference on Language Resources and Evaluation);
  - was a member of the program committee of TALN 2012 (19e conférence francophone Traitement automatique des langues naturelles);
  - is an editorial committee member of the Journal TAL (Traitement automatique des langues; since July 2009);
  - was a member of the reading committee of several issues of the Journal TAL (Traitement automatique des langues) in 2012.

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

- Doctorate: H. Jégou, Tutorial at CVPR'12, 3h, Providence, USA
- Doctorate: H. Jégou, Tutorial at BMVC'12, 2h, Surrey, UK
- Doctorate: H. Jégou, Lecture at SSMS'12 summer school, 2h, Santorini, Greece
- Master: L. Amsaleg, High dimensional Indexing, 14h, M2R, University of Rennes 1
- Master: L. Amsaleg, Database Tuning, 8h, ENSAI
- Master: G. Gravier, Data analysis and stochastic modeling, 27h, M2, University of Rennes 1
- Master: V. Claveau, Multimedia Indexing, 12h, M2, ISTIC, University of Rennes 1, Rennes
- Master: V. Claveau, Indexing and multimedia databases, 15h, M2, ENSSAT
- Master: V. Claveau, Natural Language Processing, 36h, M2, Univ. Rennes 1
- Master: V. Claveau, Data-Based Knowledge Acquisition 2: Symbolic Methods, 27 hours, M1, INSA de Rennes
- Master: V. Claveau, Symbolic and sequential data, 10h, M2R, University of Rennes 1
- Master: P. Gros, Mathematics workshop, 8h, ISTIC, University of Rennes 1
- Master: E. Kijak, Image analysis and classification, 30h, M1, ESIR
- Master: E. Kijak, Image processing, 64h, M1, ESIR
- Master: E. Kijak, Supervised learning, 16h, M2R, University Rennes 1
- Master: E. Kijak, Computer Vision, 15h, M2, ESIR
- Master: E. Kijak, Indexing and multimedia databases, 15h, M2, ENSSAT
- Master: E. Kijak, Digital Documents Indexing and Retrieval, 26h, M2, University Rennes 1
- Master, François Poulet is in charge of the Master in computer science (2^{nd} year), MITIC, Computer Science Methods and Information and Communication Technologies, ISTIC, University of Rennes 1
- Master: F. Poulet, Managing Large Collections of Digital Data, 10h, M2R, ISTIC, University of Rennes 1
- Master: F. Poulet, Introduction to Data Mining, 15h, M2P, ISTIC, University of Rennes 1
- Master: F. Poulet, Mining Symbolic Data, 25h, M2P, ISTIC, University of Rennes 1
- Master: F. Poulet, Applications and Problem Solving, 10h, M2P, ISTIC, University of Rennes 1

- Master: F. Poulet, Learning Methods for Multimedia Data, 26h, M2P, ISTIC, University of Rennes 1
- Master: P. Sébillot, Advanced Databases and Modern Information Systems, 70 hours, M2, INSA de Rennes,
- Master: P. Sébillot, Data-Based Knowledge Acquisition 2: Symbolic Methods, 18 hours, M1, INSA de Rennes
- Licence: G. Gravier, Databases, 9h, L3, Université de Rennes 1
- Licence: G. Gravier, Databases, 27h, L2, INSA de Rennes
- Licence: G. Gravier, Programming in C, 8h, L3, INSA de Rennes,
- Licence: H. Jégou, Databases, 40h, L2, INSA de Rennes

9.2.2. Supervision

- PhD : Juan David Cruz-Gomez, Socio-semantic Networks Algorithm for a point of View Based Visualization of On-line Communities. December 10^{th} 2012, Cécile Bothorel (Télécom Bretagne), François Poulet.
- PhD: Thanh-Toan Do, Security Analysis of Copy Image Detection systems based on SIFT Descriptors, University Rennes 1, September 27^{th} 2012, Laurent Amsaleg, Ewa Kijak, Teddy Furon.
- PhD : Ali Reza Ebadat, Toward robust information extraction models for multimedia documents, INSA de Rennes, October 17^{th} 2012, Vincent Claveau and Pascale Sébillot
- PhD in progress: Petra Bosilj, Video description and retrieval, Started October 2012, Ewa Kijak and Sebastien Lefèvre (in collaboration with IRISA SEASIDE team)
- PhD in progress: Mohammed-Haykel Boukadida, Automatic video summarization based on constraint programming, January 2012, Patrick Gros, Sid-Ahmed Berrani (Orange Labs, Rennes)
- PhD in progress : Thanh Nghi Doan, Image Classification, Started November 2010, François Poulet
- PhD in progress: Khaoula Elagouni, Automatic image and video indexing using automatic recognition of embedded texts and natural language processing, October 10, 2009, Pascale Sébillot, Christophe Garcia (LIRIS, Lyon) and Franck Mamalet (Orange Labs, Rennes)
- PhD in progress: Julien Fayolle, Information retrieval in TV streams, Started October 2009, Fabienne Moreau, Christian Raymond, Guillaume Gravier, Patrick Gros.
- PhD in progress: Gylfi Gudmundsson, Towards parallel and distributed CBIR systems, started April 2010, Laurent Amsaleg
- PhD in progress: Mihir Jain, Video description and retrieval, started February 2011, Hervé Jégou, Patrick Gros and Patrick Bouthemy
- PhD in progress: Ludivine Kuznik, Structuring and navigating documentary archives, started April 2011, Guillaume Gravier and Pascale Sébillot, in collaboration with Institut National de l'Audiovisuel (Jean Carrive)
- PhD in progress: Abir Ncibi, Structure learning of TV streams, started October 2011, Vincent Claveau, Guillaume Gravier, Patrick Gros, Emmanuelle Martienne
- PhD in progress: Cédric Penet, Multimodal content based analysis for video on demand, started September 2010, Guillaume Gravier and Patrick Gros, in collaboration with Technicolor
- PhD in progress: Bingqing Qu, Structure discovery of TV programs from an homogeneous collection, started October 2012, Guillaume Gravier in collaboration with Institut National de l'Audiovisuel
- PhD in progress: Anca Roxana Simon, Hierarchical semantic structuring of video collections, started October 2012, Guillaume Gravier and Pascale Sébillot
- PhD in progress: Stefan Ziegler, Landmark driven speech recognition, started September 2010, Guillaume Gravier

9.2.3. Juries

- Patrick Gros, HDR, Benoît Huet (université de Nice-Sophia Antipolis)
- Guillaume Gravier, PhD, Mickaël Rouvier (LIA);
- Guillaume Gravier, PhD, Pierre Gotab (LIA);
- Guillaume Gravier, PhD, Sylvain Raybaud (LORIA);
- Guillaume Gravier, PhD, Christian Gillot (LORIA);
- Hervé Jégou, PhD, Safadi Bahjat (LIG);
- Ewa Kijak, PhD, Muneeb Ullah (Inria);
- Pascale Sébillot, HDR, Ludovic Tanguy (Toulouse university);
- Pascale Sébillot, PhD, Clémentine Adam (Toulouse university);
- Pascale Sébillot, PhD, Gert-Jan Poulisse (Katholieke Universiteit (KU) Leuven, Belgium) ;

9.3. Invited talks and lectures

- Vincent Claveau: Invited speaker at the ASAP off-site seminar, Quiberon, France, May 2012
- Vincent Claveau: Invited speaker at the Cordial seminar, December 2012
- Hervé Jégou: Keynote speaker at the ECCV workshop on web scale vision and social media, Firenze, Italy, October 2012
- Hervé Jégou: Lecturer at the CVPR conference "Large scale object recognition" tutorial (selected), Providence, USA, June 2012
- Hervé Jégou: Invited lecturer at Summer School on Social Media Modeling and Search (SSMS'12), Thira, Greece, September 2012
- Hervé Jégou: Invited tutorial at the BMVC conference, Surrey, UK, September 2012
- Hervé Jégou: Invited talk at Oxford University, Visual Geometry Group, UK, September 2012
- Hervé Jégou: Invited talk at the National Institute of Technology (NII), Tokyo, Japan, April 2012
- Hervé Jégou: Invited talk at Technicolor Palo Alto, California, USA, August 2012
- Hervé Jégou: Invited talk at the ETIS laboratory/ENSEA, Cergy-Pontoise, France, December 2012

9.4. Popularization

- Sébastien Campion and Teddy Furon: Participation to *Fête de la Science*, October 12-13, 2012;
- The image search technique based on Pqcodes [83] was popularized in the journal Futura Sciences;
- The Texmix demonstration was presented at Futur en Seine.

10. Bibliography

Major publications by the team in recent years

[1] L. AMSALEG, P. GROS. *Content-based Retrieval Using Local Descriptors: Problems and Issues from a Database Perspective*, in "Pattern Analysis and Applications", March 2001, vol. 2001, n^{o} 4, p. 108-124.

[2] V. CLAVEAU, P. SÉBILLOT, C. FABRE, P. BOUILLON. *Learning Semantic Lexicons from a Part-of-Speech and Semantically Tagged Corpus Using Inductive Logic Programming*, in "Journal of Machine Learning Research, special issue on Inductive Logic Programming", August 2003, vol. 4, p. 493–525.

[3] S. HUET, G. GRAVIER, P. SÉBILLOT. *Morpho-Syntactic Post-Processing with N-best Lists for Improved French Automatic Speech Recognition*, in "Computer Speech and Language", October 2010, vol. 24, n^{o} 4, p. 663–684.

[4] H. JÉGOU, M. DOUZE, C. SCHMID. *Product quantization for nearest neighbor search*, in "IEEE Transactions on Pattern Analysis & Machine Intelligence", January 2011, vol. 33, n^{o} 1, p. 117–128.

[5] E. KIJAK, G. GRAVIER, L. OISEL, P. GROS. *Audiovisual integration for sport broadcast structuring*, in "Multimedia Tools and Applications", 2006, vol. 30, p. 289-312, http://www.springerlink.com/content/24h61433843r474l/.

[6] H. LEJSEK, F. H. ASMUNDSSON, B. Þ. JÓNSSON, L. AMSALEG. *NV-tree: An Efficient Disk-Based Index for Approximate Search in Very Large High-Dimensional Collections*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", May 2009, vol. 31, n^{o} 5, p. 869–883.

[7] X. NATUREL, P. GROS. *Detecting Repeats for Video Structuring*, in "Multimedia Tools and Applications", May 2008, vol. 38, n^{o} 2, p. 233–252.

Publications of the year

Doctoral Dissertations and Habilitation Theses

[8] T.-T. DO. *Security analysis of image copy detection systems based on SIFT descriptors*, Université Rennes 1, September 2012, http://hal.inria.fr/tel-00766932.

[9] A.-R. EBADAT. *Toward Robust Information Extraction Models for Multimedia Documents*, INSA de Rennes, October 2012, http://hal.inria.fr/tel-00760383.

Articles in International Peer-Reviewed Journals

[10] J. D. CRUZ GOMEZ, C. BOTHOREL, F. POULET. *Community detection and visualization in social networks: integrating structural and semantic information*, in "ACM Transactions on Intelligent Systems and Technology", 2013, http://hal.inria.fr/hal-00763931.

[11] F. CÉROU, P. DEL MORAL, T. FURON, A. GUYADER. *Sequential Monte Carlo for rare event estimation*, in "Statistics and Computing", 2012, vol. 22, n^{o} 3, p. 795-908 [*DOI :* 10.1007/s11222-011-9231-6], http://hal.inria.fr/inria-00584352.

[12] G. GRAVIER, C.-H. DEMARTY, S. BAGHDADI, P. GROS. *Classification-oriented structure learning in Bayesian networks for multimodal event detection in videos*, in "Journal of Multimedia Tools and Applications", 2012, http://hal.inria.fr/hal-00712589.

[13] C. GUINAUDEAU, G. GRAVIER, P. SÉBILLOT. *Enhancing lexical cohesion measure with confidence measures, semantic relations and language model interpolation for multimedia spoken content topic segmentation*, in "Computer Speech and Language", 2012, vol. 26, n^{o} 2, p. 90-104, http://hal.inria.fr/hal-00645705.

[14] H. JÉGOU, F. PERRONNIN, M. DOUZE, J. SÁNCHEZ, P. PÉREZ, C. SCHMID. *Aggregating local image descriptors into compact codes*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", September 2012, http://hal.inria.fr/inria-00633013.

[15] B. LECOUTEUX, G. LINARES, Y. ESTÈVE, G. GRAVIER. *Dynamic Combination of Automatic Speech Recognition Systems by Driven Decoding*, in "IEEE Transactions on Audio, Speech and Language Processing", 2012, http://hal.inria.fr/hal-00758626.

[16] P. MEERWALD, T. FURON. *Towards practical joint decoding of binary Tardos fingerprinting codes*, in "IEEE Transactions on Information Forensics and Security", April 2012, vol. 7, n^{o} 4, p. 1168-1180 [*DOI :* 10.1109/TIFS.2012.2195655], http://hal.inria.fr/hal-00740964.

[17] A. MUSCARIELLO, F. BIMBOT, G. GRAVIER. *Unsupervised Motif Acquisition in Speech via Seeded Discovery and Template Matching Combination*, in "IEEE Transactions on Audio, Speech and Language Processing", September 2012, vol. 20, n^{o} 7, p. 2031 - 2044 [*DOI :* 10.1109/TASL.2012.2194283], http://hal.inria.fr/hal-00740978.

[18] A. ŠILIĆ, A. MORIN, J.-H. CHAUCHAT, B. D. BAŠIĆ. *Visualization of temporal text collections based on Correspondence Analysis*, in "Expert Systems with Applications", April 2012, vol. 39, n^{o} 15, p. .12143-12157 [*DOI :* 10.1016/J.ESWA.2012.04.040], http://hal.inria.fr/hal-00712590.

Articles in National Peer-Reviewed Journals

[19] J. D. CRUZ GOMEZ, C. BOTHOREL, F. POULET. *Détection et visualisation des communautés dans les réseaux sociaux*, in "Revue d'Intelligence Artificielle", 2012, vol. 26, n^{o} 4, p. 369-392, http://hal.inria.fr/hal-00746090.

International Conferences with Proceedings

[20] L. AMSALEG, G. T. GUDMUNDSSON, B. Þ. JÓNSSON. *Distributed High-Dimensional Index Creation using Hadoop, HDFS and C++*, in "CBMI'13 - 10th International Workshop on Content-Based Multimedia Indexing", Annecy, France, IEEE, June 2012, p. 1-6 [*DOI :* 10.1109/CBMI.2012.6269848], http://hal.inria.fr/hal-00746012.

[21] N. BALLAS, B. LABBÉ, A. SHABOU, H. LE BORGNE, P. GOSSELIN, M. REDI, B. MERIALDO, H. JÉGOU, J. DELHUMEAU, R. VIEUX, B. MANSENCAL, J. BENOIS-PINEAU, S. AYACHE, A. HAMADI, B. SAFADI, F. THOLLARD, N. DERBAS, G. QUENOT, H. BREDIN, M. CORD, B. GAO, C. ZHU, Y. TANG, E. DELLANDREA, C.-E. BICHOT, L. CHEN, A. BENOIT, P. LAMBERT, T. STRAT, J. RAZIK, S. PARIS, H. GLOTIN, T. N. TRUNG, D. PETROVSKA-DELACRÉTAZ, G. CHOLLET, A. STOIAN, M. CRUCIANU. *IRIM at TRECVID 2012: Semantic Indexing and Instance Search*, in "TRECVID 2012 - TREC Video Retrieval Evaluation workshop", Gaithersburg, MD, United States, November 2012, 12, http://hal.inria.fr/hal-00770258.

[22] P. BAS, T. FURON. *Key Length Estimation of Zero-Bit Watermarking Schemes*, in "EUSIPCO 2012", Romania, August 2012, http://hal.inria.fr/hal-00728159.

[23] P. BAS, T. FURON, F. CAYRE. *Practical Key Length of Watermarking Systems*, in "IEEE ICASSP", Kyoto, Japan, IEEE, March 2012, http://hal.inria.fr/hal-00686813.

[24] M. BRÉHINIER, S. CAMPION, G. GRAVIER. *Texmix: an automatically generated news navigation portal*, in "ICMR - ACM International Conference on Multimedia Retrieval", Hong-Kong, China, ACM, June 2012 [*DOI :* 10.1145/2324796.2324868], http://hal.inria.fr/hal-00767253.

[25] V. CLAVEAU. *IRISA Participation in JRS 2012 Data-Mining Challenge: Lazy-Learning with Vectorization*, in "JRS 2012 Data Mining Competition: Topical Classification of Biomedical Research Papers, special event of Joint Rough Sets Symposium", Chengdu, China, September 2012, http://hal.inria.fr/hal-00760145.

[26] V. CLAVEAU. *Unsupervised and semi-supervised morphological analysis for Information Retrieval in the biomedical domain*, in "Computational Linguistics (COLING)", Mumbai, India, December 2012, http://hal.inria.fr/hal-00760114.

[27] V. CLAVEAU, C. RAYMOND. *Participation de l'IRISA à DeFT2012 : recherche d'information et apprentissage pour la génération de mots-clés*, in "Défi Fouille de textes (DEFT)", Grenoble, France, June 2012, http://hal.inria.fr/hal-00760132.

[28] J. D. CRUZ GOMEZ, C. BOTHOREL, F. POULET. *Layout algorithm for clustered graphs to analyze community interactions in social networks*, in "IEEE/ACM ASONAM 2012: International Conference on Advances in Social Networks Analysis and Mining", Istambul, Turkey, 2012, http://hal.inria.fr/hal-00739637.

[29] T.-T. DO, L. AMSALEG, E. KIJAK, T. FURON. *Security-Oriented Picture-In-Picture Visual Modifications*, in "ACM International Conference on Multimedia Retrieval", Hong-Kong, China, 2012, http://hal.inria.fr/hal-00764431.

[30] T.-T. DO, E. KIJAK, L. AMSALEG, T. FURON. *Enlarging hacker's toolbox: deluding image recognition by attacking keypoint orientations*, in "IEEE ICASSP", Kyoto, Japan, IEEE, March 2012, http://hal.inria.fr/hal-00686809.

[31] T.-T. DO, E. KIJAK. *Face recognition using co-occurrence histograms of oriented gradients*, in "IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)", Japan, March 2012, p. 1301-1304 [*DOI :* 10.1109/ICASSP.2012.6288128], http://hal.inria.fr/hal-00766960.

[32] A.-R. EBADAT, V. CLAVEAU, P. SÉBILLOT. *Proper Noun Semantic Clustering using Bag-Of-Vectors*, in "Applied Natural Language Processing (ANLP) conference. Special track at the 25th International FLAIRS Conference.", Marco Island, FL, United States, May 2012, http://hal.inria.fr/hal-00760105.

[33] K. ELAGOUNI, C. GARCIA, F. MAMALET, P. SÉBILLOT. *Combining Multi-Scale Character Recognition and Linguistic Knowledge for Natural Scene Text OCR*, in "10th IAPR International Workshop on Document Analysis Systems, DAS 2012", Gold Coast, Queensland, Australia, 2012, p. 120-124, http://hal.inria.fr/hal-00753908.

[34] K. ELAGOUNI, C. GARCIA, F. MAMALET, P. SÉBILLOT. *Text Recognition in Videos using a Recurrent Connectionist Approach*, in "22th International Conference on Artificial Neural Networks, ICANN 2012", Lausanne, Switzerland, Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2012, vol. 7553, p. 172-179 [*DOI :* 10.1007/978-3-642-33266-1_22], http://hal.inria.fr/hal-00753906.

[35] J. FAYOLLE, F. MOREAU, C. RAYMOND, G. GRAVIER. *Automates lexico-phonétiques pour l'indexation et la recherche de segments de parole*, in "Journées d'Études sur la Parole", Grenoble, France, 2012, vol. 1, p. 49-56, http://hal.inria.fr/hal-00742848.

[36] J. FAYOLLE, M. SARACLAR, F. MOREAU, C. RAYMOND, G. GRAVIER. *Lexical-phonetic automata for spoken utterance indexing and retrieval*, in "International Conference on Speech Communication and Technologies", Portland, United States, 2012, http://hal.inria.fr/hal-00757765.

[37] K. FORT, V. CLAVEAU. *Annotating Football Matches: Influence of the Source Medium on Manual Annotation*, in "International Conference on Language Resources and Evaluation (LREC)", Istanbul, Turkey, May 2012, http://hal.inria.fr/hal-00709170.

[38] T. FURON, P. BAS. *A New Measure of Watermarking Security Applied on DC-DM QIM*, in "Information Hiding 2012", Berkeley, United States, May 2012, http://hal.inria.fr/hal-00702689.

[39] T. FURON, A. GUYADER, F. CÉROU. *Decoding Fingerprinting Using the Markov Chain Monte Carlo Method*, in "IEEE Workshop on Information Forensics and Security", Tenerife, Spain, IEEE, December 2012, http://hal.inria.fr/hal-00757152.

[40] G. GRAVIER, G. ADDA, N. PAULSON, M. CARRÉ, A. GIRAUDEL, O. GALIBERT. *The ETAPE corpus for the evaluation of speech-based TV content processing in the French language*, in "International Conference on Language Resources, Evaluation and Corpora", Turkey, 2012, http://hal.inria.fr/hal-00712591.

[41] P. GROS. *Recent advances and challenges in TV structuring*, in "EUSIPCO - 20th European Signal Processing Conference", Bucharest, Romania, EURASIP, August 2012, http://hal.inria.fr/hal-00766732.

[42] C. GUINAUDEAU, G. GRAVIER, P. SÉBILLOT. *IRISA at MediaEval 2012: Search and Hyperlinking Task*, in "MediaEval 2012 Workshop", Pisa, Italy, 2012, http://hal.inria.fr/hal-00753909.

[43] G. Þ. GUÐMUNDSSON, L. AMSALEG, B. Þ. JÓNSSON. *Distributed High-Dimensional Index Creation using Hadoop, HDFS and C++*, in "Content-Based Multimedia Indexing", Annecy, France, 2012, http://hal.inria.fr/hal-00764434.

[44] M. JAIN, R. BENMOKHTAR, P. GROS, H. JÉGOU. *Hamming Embedding Similarity-based Image Classification*, in "ICMR - ACM International Conference on Multimedia Retrieval", Hong-Kong, China, June 2012, http://hal.inria.fr/hal-00688169.

[45] H. JÉGOU, O. CHUM. *Negative evidences and co-occurrences in image retrieval: the benefit of PCA and whitening*, in "ECCV - European Conference on Computer Vision", Firenze, Italy, October 2012, http://hal.inria.fr/hal-00722622.

[46] H. JÉGOU, J. DELHUMEAU, J. YUAN, G. GRAVIER, P. GROS. *Babaz: a large scale audio search system for video copy detection*, in "ICASSP - 37th International Conference on Acoustics, Speech, and Signal Processing", Kyoto, Japan, January 2012, http://hal.inria.fr/hal-00661581.

[47] H. JÉGOU, T. FURON, J.-J. FUCHS. *Anti-sparse coding for approximate nearest neighbor search*, in "ICASSP - 37th International Conference on Acoustics, Speech, and Signal Processing", Kyoto, Japan, January 2012, http://hal.inria.fr/hal-00661591.

[48] F. METZE, N. RAJPUT, X. ANGUERA, M. DAVEL, G. GRAVIER, C. VAN HEERDEN, G. MANTENA, A. MUSCARIELLO, K. PRADHALLAD, I. SZÖKE, J. TEJEDOR. *The Spoken Web Search task at MediaEval 2011*, in "IEEE International Conference on Acoustics, Speech and Signal Processing", France, 2012, http://hal.inria.fr/hal-00671011.

[49] C. PENET, C.-H. DEMARTY, G. GRAVIER, P. GROS. *Multimodal information fusion and temporal integration for violence detection in movies*, in "ICASSP - 37th International Conference on Acoustics, Speech, and Signal Processing (2012)", Kyoto, Japan, March 2012, http://hal.inria.fr/hal-00671016.

[50] B. SAFADI, N. DERBAS, A. HAMADI, F. THOLLARD, G. QUENOT, J. DELHUMEAU, H. JÉGOU, T. GEHRIG, H. K. EKENEL, R. STIFELHAGEN. *Quaero at TRECVID 2012: Semantic Indexing*, in "TRECVID 2012 - TREC Video Retrieval Evaluation workshop", Gaithersburg, MD, United States, November 2012, 6, http://hal.inria.fr/hal-00770240.

[51] A.-P. TA, M. BEN, G. GRAVIER. *Improving Cluster Selection and Event Modeling in Unsupervised Mining for Automatic Audiovisual Video Structuring*, in "International Conference on MultiMedia Modeling", Austria, 2012, http://hal.inria.fr/hal-00671157.

[52] A.-P. TA, G. GRAVIER. *Unsupervised mining of multiple audiovisually consistent clusters for video structure analysis*, in "Intl. Conf. on Multimedia and Exhibition", Australia, 2012, http://hal.inria.fr/hal-00718985.

[53] G. TÓMASSON, H. SIGURÞÓRSSON, K. RUNARSSON, G. K. OLAFSSON, B. Þ. JÓNSSON, L. AMSALEG. *Using PhotoCube as an Extensible Demonstration Platform for Advanced Image Analysis Techniques*, in "Content-Based Multimedia Indexing", annecy, France, 2012, http://hal.inria.fr/hal-00764447.

[54] J. YUAN, G. GRAVIER, S. CAMPION, X. LIU, H. JÉGOU. *Efficient Mining of Repetitions in Large-Scale TV Streams with Product Quantization Hashing*, in "Workshop on Web-scale Vision and Social Media, in conjunction with ECCV", Firenze, Italy, August 2012, http://hal.inria.fr/hal-00731090.

National Conferences with Proceeding

[55] M. BRÉHINIER. *TexMix - Navigation dans des archives de journaux télévisés*, in "RFIA 2012 (Reconnaissance des Formes et Intelligence Artificielle)", Lyon, France, January 2012, p. 978-2-9539515-2-3, Session "Démo", http://hal.inria.fr/hal-00660955.

[56] V. CLAVEAU. *Vectorisation, Okapi et calcul de similarité pour le TAL : pour oublier enfin le TF-IDF*, in "Traitement Automatique des Langues Naturelles (TALN)", Grenoble, France, June 2012, http://hal.inria.fr/hal-00760158.

[57] V. CLAVEAU, E. KIJAK. *Analyse morphologique fine pour la recherche d'information biomédicale*, in "conférence sur la recherche d'information et applications CORIA", Bordeaux, France, March 2012, http://hal.inria.fr/hal-00760124.

[58] T.-N. DOAN, F. POULET. *Un environnement efficace pour la classification d'images à grande échelle*, in "12èmes Journées Francophones "Extraction et Gestion des Connaissances"", Bordeaux, France, Hermann, January 2012, vol. E, p. 471-482, http://hal.inria.fr/hal-00763926.

[59] A.-R. EBADAT, V. CLAVEAU, P. SÉBILLOT. *Semantic Clustering using Bag-of-Bag-of-Features*, in "9e conférence en recherche d'information et applications, CORIA 2012", Bordeaux, France, 2012, p. 229-244, http://hal.inria.fr/hal-00753912.

[60] K. FORT, V. CLAVEAU. *Annotation manuelle de matchs de foot : Oh la la la ! l'accord inter-annotateurs ! et c'est le but !*, in "Traitement Automatique des Langues Naturelles", Grenoble, France, June 2012, p. 383-390, http://hal.inria.fr/hal-00709181.

[61] E. MARTIENNE, V. CLAVEAU, P. GROS. *Application des Champs Conditionnels Aléatoires à l'étiquetage de flux télévisuel*, in "RFIA 2012 (Reconnaissance des Formes et Intelligence Artificielle)", Lyon, France, January 2012, p. 978-2-9539515-2-3, Session "Posters", http://hal.inria.fr/hal-00656547.

Conferences without Proceedings

[62] M. BRÉHINIER, G. GRAVIER. *Texmix: An automatically generated news navigation portal*, in "3rd International Workshop on Future Television", Germany, 2012, http://hal.inria.fr/hal-00767158.

[63] C.-H. DEMARTY, C. PENET, G. GRAVIER, M. SOLEYMANI. *A benchmarking campaign for the multimodal detection of violent scenes in movies*, in "European Conference on Computer Vision, Workshop on Information Fusion in Computer Vision for Concept Recognition", Italy, 2012, http://hal.inria.fr/hal-00767036.

[64] C.-H. DEMARTY, C. PENET, G. GRAVIER, M. SOLEYMANI. *The MediaEval 2012 Affect Task: Violent Scenes Detection*, in "Working Notes Proceedings of the MediaEval 2012 Workshop", Italy, 2012, http://hal.inria.fr/hal-00757577.

[65] B. LUDUSAN, S. ZIEGLER, G. GRAVIER. *Integrating Stress Information in Large Vocabulary Continuous Speech Recognition*, in "Interspeech", United States, 2012, http://hal.inria.fr/hal-00758622.

[66] F. METZE, E. BARNARD, M. DAVEL, C. VAN HEERDEN, X. ANGUERA, G. GRAVIER, N. RAJPUT. *The Spoken Web Search Task*, in "Working Notes Proceedings of the MediaEval 2012 Workshop", Italy, 2012, http://hal.inria.fr/hal-00757594.

[67] C. PENET, C.-H. DEMARTY, M. SOLEYMANI, G. GRAVIER, P. GROS. *Technicolor/Inria/Imperial College London at the MediaEval 2012 Violent Scene Detection Task*, in "Working Notes Proceedings of the MediaEval 2012 Workshop", Italy, 2012, http://hal.inria.fr/hal-00757584.

[68] M. TREVISIOL, J. DELHUMEAU, H. JÉGOU, G. GRAVIER. *How Inria/IRISA identifies Geographic Location of a Video*, in "Working Notes Proceedings of the MediaEval 2012 Workshop", Italy, 2012, http://hal.inria.fr/hal-00757453.

[69] S. ZIEGLER, B. LUDUSAN, G. GRAVIER. *Towards a new speech event detection approach for landmark-based speech recognition*, in "IEEE Workshop on Spoken Language Technology", United States, 2012, http://hal.inria.fr/hal-00758424.

[70] S. ZIEGLER, B. LUDUSAN, G. GRAVIER. *Using Broad Phonetic Classes to Guide Search in Automatic Speech Recognition*, in "Interspeech", United States, 2012, http://hal.inria.fr/hal-00758427.

Scientific Books (or Scientific Book chapters)

[71] H. AZZAG, B. LE GRAND, M. NOIRHOMME-FRAITURE, F. PICAROUGNE, F. POULET. *Atelier fouille visuelle de données (EGC'2012)*, N/A, January 2012, 64, http://hal.inria.fr/hal-00746092.

[72] V. CLAVEAU, P. SÉBILLOT. *Automatic Acquisition of GL Resources, Using an Explanatory, Symbolic Technique*, in "Advances in Generative Lexicon Theory", Spinger, December 2012, chap. 19, http://hal.inria.fr/hal-00760258.

[73] Z. A. A. IBRAHIM, P. GROS. *About TV Stream Macro-Segmentation: Approaches and Results*, in "TV Content Analysis: Techniques and Applications", Y. KOMPATSIARIS, B. MERIALDO, S. LIAN (editors), CRC Press, March 2012 [*DOI :* 10.1201/B11723-10], http://hal.inria.fr/hal-00740992.

[74] C. RAYMOND, V. CLAVEAU. *Apprentissage supervisé et paresseux pour la fouille de textes*, in "Expérimentations et évaluations en fouille de textes", Systèmes d'information et organisations documentaires, Hermes - Lavoisier, November 2012, ch11, http://hal.inria.fr/hal-00760620.

Books or Proceedings Editing

[75] B. L. GRAND, F. POULET (editors). *Revue d'intelligence artificielle RSTI série RIA Vol.26 (4) 2012 Extraction de connaissances et visualisation de grands réseaux*, Revue d'intelligence artificielle, Hermes Lavoisier, 2012, 116, http://hal.inria.fr/hal-00746088.

Research Reports

[76] A. BOURRIER, F. PERRONNIN, R. GRIBONVAL, P. PÉREZ, H. JÉGOU. *Nearest neighbor search for arbitrary kernels with explicit embeddings*, Inria, August 2012, n^{o} RR-8040, http://hal.inria.fr/hal-00722635.

Other Publications

[77] A.-R. SIMON. *Hierarchical Topic Segmentation of TV shows Automatic Transcripts*, June 2012, http://dumas.ccsd.cnrs.fr/dumas-00725338.

References in notes

[78] S. WERMTER, E. RILOFF, G. SCHELER (editors). *Connectionist, Statistical and Symbolic Approaches to Learning for Natural Language Processing*, Lecture Notes in Computer Science, Vol. 1040, Springer Verlag, 1996.

[79] S.-A. BERRANI, L. AMSALEG, P. GROS. *Recherche par similarités dans les bases de données multidimensionnelles : panorama des techniques d'indexation*, in "Ingénierie des Systèmes d'Information", 2002, vol. 7, n^{o} 5/6.

[80] T. DEAN, K. KANAZAWA. *A model for reasoning about persistence and causation*, in "Articial Intelligence Journal", 1989, vol. 93, n^{o} 1.

[81] A. GIONIS, P. INDYK, R. MOTWANI. *Similarity Search in High Dimensions via Hashing*, in "Proceedings of the 25th International Conference on Very Large Data Bases", Edinburgh, Scotland, United Kingdom, September 1999, p. 518–529.

[82] C. HARRIS, M. STEPHENS. *A Combined Corner and Edge Detector*, in "Proceedings of the 4th Alvey Vision Conference", 1988, p. 147-151.

[83] H. JÉGOU, M. DOUZE, C. SCHMID. *Product Quantization for Nearest Neighbor Search*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", January 2011, vol. 33, n^{o} 1, p. 117–128 [*DOI :* 10.1109/TPAMI.2010.57], http://hal.inria.fr/inria-00514462/en.

[84] D. G. LOWE. *Distinctive image features from scale-invariant keypoints*, in "International Journal of Computer Vision", 2004, vol. 60, n^{o} 2, p. 91–110.

[85] K. MURPHY. *Dynamic Bayesian Networks: Representation, Inference and Learning*, University of California, Berkeley, 2002.

[86] M. OSTENDORF. *From HMMs to Segment Models*, in "Automatic Speech and Speaker Recognition - Advanced Topics", Kluwer Academic Publishers, 1996, chap. 8.

[87] L. RABINER, B.-H. JUANG. *Fundamentals of speech recognition*, Prentice Hall, Englewood Cliffs, NJ, 1993.

[88] G. SALTON. *Automatic Text Processing*, Addison-Wesley, 1989.

[89] J. SIVIC, A. ZISSERMAN. *Video Google: A Text Retrieval Approach to Object Matching in Videos*, in "Proceedings of the International Conference on Computer Vision", October 2003, vol. 2, p. 1470–1477.

[90] H. J. WOLFSON, I. RIGOUTSOS. *Geometric Hashing: An Overview*, in "Computing in Science and Engineering", 1997, vol. 4, p. 10-21, http://doi.ieeecomputersociety.org/10.1109/99.641604.