IN PARTNERSHIP WITH:
CNRS

Institut national des sciences appliquées de Rennes

Université Rennes 1

Activity Report 2011

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: Video, Natural Language, Speech, Multimedia, Information Indexing And Retrieval

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]

External Collaborators

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

Technical Staff

Mathieu Ben [INRIA Technical Staff, Quaero project, until April 30th]
Rachid BenMokhtar [INRIA Technical Staff, Quaero project, since June 1st]
Morgan Bréhinier [INRIA Technical Staff, OpenSem project, since September 15th]
Sébastien Campion [INRIA Research Engineer]
Jonathan Delhumeau [INRIA Technical Staff, Quaero project since June 1st]
Carryn Hayward [INRIA Technical Staff, Quaero project, also with SAF, since April 11th]
Stacy Payne [INRIA Technical Staff, Quaero project, also with SAF, until April 21st]

PhD Students

Thanh Toan Do [MESR grant]
Thanh Nghi Doan [Vietnamese government grant and Brittany Council grant]
Ali Reza Ebadat [Quaero project]
Khaoula Elagouni [CIFRE grant with Orange]
Julien Fayolle [Quaero project and Brittany council grant]
Gylfi Gudmundsson [Quaero project]
Camille Guinaudeau [Quaero project and Brittany Council grant, until September 30th, Assistant professor at Univ. Rennes since Oct. 1st]
Mihir Jain [INRIA Cordis Grant, since February 1st]
Ludivine Kuznik [CIFRE grant with INA since April 18th]
Abir Ncibi [INRIA Grant, since November 2nd]
Cédric Penet [CIFRE grant with Technicolor]
Romain Tavenard [ENS Cachan grant, until August 31st]

Post-Doctoral Fellows

Josip Krapac [Postdoc on Quaero project since October 1st]
Bogdan Ludusan [CNRS postdoc since June 1st]

Peter Meerwald [INRIA Postdoc, until August 27th]
Anh Phuong Ta [INRIA Postdoc on Quaero project]

Visiting Scientists

Khang Nguyen Pham [Associate Professor at Cantho University, Vietnam]
Jiangbo Yuan [PhD at Florida State University, USA]

Administrative Assistants

Elodie Lequoc [INRIA Secretary, partial position, since November 11th]
Loïc Lesage [INRIA Secretary, partial position, until November 10th]

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 summarize our research project in one short sentence, let us say that 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, eventually 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

- TEXMEX has co-organized the MediaEval 2011 evaluation campaign tasks on Violent Scenes Detection and on Spoken Web Search.
- TEXMEX has successfully participated in two tasks of Trecvid 2011, the main benchmark in automatic video analysis and retrieval organized by the National Institute of Technology.
    1. In the Semantic Indexing task, we have contributed to the submission of the Quaero consortium, jointly with LIG and Karlsruhe Institute of Technology. This submission was ranked 3rd out of 19 participants.
    2. In the Copy Detection task, our joint submission with the LEAR project-team was ranked approximately 3rd out of 21 participants, with respect to search quality.
- Gwénolé Lecorvé was awarded the best Ph.D. award of the French Speech Communication Association.

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 [77] or the Difference of Gaussian method proposed by David Lowe [79].

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 [79], 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. A recent trend consists in computing a global descriptor from local ones, such as proposed in the so-called bag-of-visual-word approach [84]. 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 [83]; 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 [70]. 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, [82]) 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 [81]. 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 [73] 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 [80] 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 [71]. 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 [76], 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 [84] 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 most clever 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 great 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

5.1.1. New Software

5.1.1.1. Babaz

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

The deposit of this software at APP is currently being processed (submitted). The software is available from its homepage, namely http://babaz.gforge.inria.fr/.

Babaz is a audio database management system with an audio-based search function, which is intended for audio-based search in video archives.

It is licensed under the terms of the GNU General Public License v3.0.

5.1.1.2. Bag-of-colors

Participants: Sébastien Campion [correspondent], Hervé Jégou.

Joint work with Christian Wengert (Kooba Inc.) and Matthijs Douze (INRIA LEAR and SED project-teams)

This package implements the color descriptor proposed in our ACM Multimedia paper [48], which improves the previous color histogram representation.

The bag-of-colors software corresponds to two packages:

- The (reference) Matlab package, which was co-developed with Christian Wengert and Matthijs Douze ;
- The python package (translated) was translated from Matlab by Sébastien Campion.

The Matlab version of this package is available on Github at https://github.com/kooaba/bag-of-color/.

The python version is available on the gforge INRIA server, and might be available on request.

5.1.1.3. 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

5.1.1.4. Don Quixotte

Participant: Teddy Furon [correspondent].

This software was developed in collaboration with project-team TEMICS (P. Meerwald)

Don Quixotte a software suite in C for Tardos Fingerprinting code (Code generation, collusion, and accusation with single and/or joint decoding).

5.1.1.5. Rare Event

Participant: Teddy Furon [correspondent].

This software was developed in collaboration with project-team ASPI (F. Cérou, A. Guyader)

Rare Event is a Matlab package for rare event probabilities and extreme quantiles estimations

5.1.2. Most active software started before 2011

5.1.2.1. Bigimbaz

Participant: Hervé Jégou [correspondent].

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

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.

5.1.2.2. kertrack

Participant: Sébastien Campion [correspondent].

Visual graphical interface for tracking visual targets based on particle filter tracking or based on mean-shift. The deposit of this software at APP is currently being processed.

5.1.2.3. mozaic2d

Participant: Sébastien Campion [correspondent].

Creation of spatio-temporal mosaic based on dominant motion compensation. It depends on the Motion2D library, which computes the dominant motion, and then adjust the images by back-warping. The deposit of this software at APP is currently being processed.

5.1.2.4. PimPy

Participant: Sébastien Campion [correspondent].

The software homepage is available here: http://pim.gforge.inria.fr/pimpy/.

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

PimPy stands for Indexing Multimedia with Python (or Platform for Indexing Multimedia with Python). The aim of this module is to provide 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.

5.1.2.5. Pqcodes

Participant: Hervé Jégou [correspondent].

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

First APP deposit: IDDN.FR.001.220012.000.S.P.2010.000.10000

A new version of the software at APP is currently being processed.

Pqcodes is a library which implements the approximate k nearest neighbor search method of [18]. This software has been transferred to Technicolor in August 2011.

5.1.2.6. python-geohash

Participant: Sébastien Campion [correspondent].

The deposit of this software at APP is currently being processed.

Implementation of the Geometric Hashing algorithm of [85] to check if geometrical consistency between pairs of images.

5.1.2.7. Samusa

Participant: Sébastien Campion [correspondent].

This software is jointly maintained with Guillaume Gravier.

Samusa enable to detect speech and/or musical segment in multimedia content.

5.1.2.8. Yael

Participant: Hervé Jégou [correspondent].

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

APP deposit: IDDN.FR.001.220014.000.S.P.2010.000.10000

A new version of the software at APP is currently being processed.

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.

5.1.2.9. TVSearch

Participant: Sébastien Campion [correspondent].

TVSearch is a content based retrieval search engine used to search and propagate manual annotation such as advertisement in a TV corpora. Based on a binary DCT descriptor, it used GPU card to compute exhaustive Hamming distance between the query and database. For example, a query of 11 seconds in 21 days on television (504 hours) is done in 9 seconds. (*i.e.*, bit-rate of 2,3 days/second) TVSearch offer a web services API using the HTTP/REST protocol.

The deposit of this software at APP is currently being processed.

5.1.2.10. AVSST

Participant: Sébastien Campion [correspondent].

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.

5.1.3. Other softwares

Several software programs have been developed in the team over the years:

I-DESCRIPTION (APP deposit number: 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, build 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, Fishnet is an automatic web pages grabber associated with a specific theme.

MATCH MAKER, semantic relation extraction by statistical methods.

IRISA NEWS TOPIC SEGMENTER (IRINTS), automatically segments speech transcripts into topic-consistent parts.

IRISAPHON, produce phonetic words.

5.2. Demonstrations

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 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. At IRISA/INRIA Rennes, 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 (Quaero CTC workshop, Journée INRIA Industrie La Télévision du Futur) and a video has been made available online on the portal of the EIT ICT Labs OpenSEM project.

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 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.

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 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.

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

6. New Results

6.1. Advanced algorithms of data analysis, description

6.1.1. Advanced description techniques

6.1.1.1. Image joint description and compression

Participant: Ewa Kijak.

This is a joint work with the TEMICS *project-team (J. Zepeda and C. Guillemot).*

In the context of ANR project ICOS-HD ended at december 2010, in collaboration with Christine Guillemot from TEMICS, we investigated sparse representations methods for local image description. We have developed methods for learning dictionaries to be used for sparse signal representations. These methods lead to dictionaries which have been called Iteration-Tuned Dictionaries (ITDs), Basic ITD (BITD), Tree-Structured ITD (TSITD) and Iteration-Tuned and Aligned Dictionaries (ITAD). All three proposed ITD schemes (BITD, TSITD and ITAD) have been shown to outperform the state-of-the-art learned dictionaries in terms of PSNR versus sparsity. The performance of these dictionaries has also been assessed for both compression and de-noising applications. ITAD in particular has been used to produce a new image codec that outperforms JPEG2000 for a fixed image class and leads in 2011 to two new publications [49], [20].

6.1.1.2. Bag-of-colors

Participant: Hervé Jégou.

This is joint work with Christian Wengert (Kooaba) and Matthijs Douze (INRIA LEAR and SED project-teams).

This work investigates [48] the use of color information when used within a state-of-the-art large scale image search system. We introduce a simple color signature generation procedure, used either to produce global or local descriptors. As a global descriptor, it outperforms several state-of-the-art color description methods, in particular the bag-of-words method based on color SIFT. As a local descriptor, our signature is used jointly with SIFT descriptors (no color) to provide complementary information.

6.1.1.3. Aggregating local image descriptors into compact codes

Participant: Hervé Jégou.

This is joint work with Matthijs Douze (INRIA LEAR and SED project-teams), Patrick Pérez (Technicolor), Florent Perronnin (Xérox Research Center Europe) and Cordelia Schmid (INRIA LEAR).

This work [19] addresses the problem of large-scale image search and consolidates and extends results from a previous work [78]. Different ways of aggregating local image descriptors into a vector are compared, and the Fisher vector is shown to achieves better performance than the reference bag-of-visual words approach for any given vector dimension. We then jointly optimize dimensionality reduction and indexing in order 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. Searching a 100 million image dataset takes about 250 ms on one processor core.

6.1.2. Browsing multimedia databases

Participant: Laurent Amsaleg.

This is a joint work with Björn Þór Jónsson and Grímur Tómasson from the School of Computer Science, Reykjavik University, Iceland.

Since the introduction of personal computers, personal collections of digital media have been growing ever larger. It is therefore increasingly important to provide effective browsing tools for such collections. We have proposed a multi-dimensional model for media browsing, called ObjectCube, based on the multi-dimensional model commonly used in OLAP applications. We implemented a prototype of a media browser based on the ObjectCube model. We then ran evaluations of its performance using three different underlying data stores and photo collections of up to one million photos.

6.1.3. Advanced data analysis techniques

6.1.3.1. Factorial analysis and output display for text and textual streams mining

Participant: Annie Morin.

Textual data can be easily transformed in frequency tables and any method working on contingency tables can be used to process them. Besides, with the important amount of available textual data, we need to find convenient ways to process the data and to get invaluable information. It appears that the use of factorial correspondence analysis allows us to get most of the information included in the data. We start exploring temporal changes in textual data and mainly focus on the visualization of results: we try to detect the topics if they have not already been identified and to study the evolution of the previous vocabulary inside a topic through time. In fact, as with economical datasets, we try to find seasonal components and cycling components in the documents and to characterize these components.

6.1.3.2. Intensive use of SVM for text and image mining

Participants: François Poulet, Thanh Nghi Doan.

Support Vector Machines (SVM) and kernel methods are known to provide accurate models but the learning task usually needs a quadratic program, so this task for very large datasets requires a large memory capacity and a long time. We have developed new algorithms. The first versions of the algorithms were based on a CPU distributed software program, then we have used GP-GPU (General Purpose GPU) versions to significantly improve the algorithm speed (130 times faster than the CPU one, 2500 times faster than libSVM, SVMPerf or CB-SVM). We have extended the least squares SVM algorithm (LS-SVM) to adapt the algorithm to datasets having a very large number of dimensions and have applied boosting to LS-SVM for datasets having simultaneously a very large number of vectors and dimensions on standard computers. In image classification, the usual frameworks involve three steps : feature extraction, building codebook by feature quantization and training the classifier with a standard classification algorithm (eg. SVM). However, task complexity becomes very large when applying this approach on large scale datasets like the ImageNet dataset containing more than 14 million images and 21,000 classes. The complexity is both about the time needed to perform each task and the memory and disk usage (eg. 11TB are needed to store the SIFT descriptors computed on the full datasets). Efficient algorithms must be used into these three steps: - obviously, the descriptors computed for one image are independant of the other image ones, so they can be computed in a parallel way, - the quantization step usually uses a k-means algorithm, we have developed different versions of parallel k-means algorithms to use on GPU or a cluster of CPUs, - for the learning task, we have developed a parallel version of LibSVM. The first results on the ten largest classes of ImageNet dataset are promising [55], we have developed a fast and efficient framework for large scale image classification.

6.1.4. Security of media

6.1.4.1. Security of content based image retrieval

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

Over the years, the level of maturity reached by content-based retrieval systems (CBRSs) has significantly increased. CBRSs have so far been used in very friendly settings where cultural enrichments are paramount. CBRSs are also used in quite different settings where the control, the surveillance and the filtering of multimedia information are central, such as for copyright enforcement systems. While an abundant literature assesses that today's CBRSs are robust against general-purpose attacks, we address in this work the security of content-based retrieval systems. Because of our expertise, we focus on security of content-based image retrieval, where images are described by SIFT descriptors and indexed by NV-Tree. We proved in one preliminary study that a real system fails to match a specifically attacked image and its quasi-copy, breaking its otherwise excellent copyright protection performances. After proposing specific attacks that aim to disturb the descriptor detection stage by both prevent some key-points of being detected and create new ones [75], [74], we pursue the work by considering attacks dedicated to the description computation stage.

6.1.4.2. Estimation of the false alarm probability in watermarking and fingerprinting

Participant: Teddy Furon.

A key issue in watermarking and fingerprinting applications is to satisfy the requirement on the probability of false detection or false accusation. Assume commercial contents are encrypted and watermarked and that future consumer electronics storage devices have a watermark detector. These devices refuse to record a watermarked content since it is copyrighted material. The probability of false alarm is the probability that the detector considers an original piece of content (which has not been watermarked) as protected. The movie that a user shot during his holidays could be rejected by his storage device. This absolutely non user-friendly behavior really scares consumer electronics manufacturers.

In fingerprinting, users' identifiers are embedded in purchased contents. When this content is found in an illegal place (e.g. a P2P network), the copyright holders decode the hidden message, find an identifier, and thus they can trace the traitor, i.e. the customer who has illegally broadcast his copy. However, the task is not that simple because dishonest users might collude. For security reason, anti-collusion codes have to be employed. Yet, these solutions have a non-zero probability of error (defined as the probability of accusing an innocent). This probability should be, of course, extremely low, but it is also a very sensitive parameter: anti-collusion codes get longer (in terms of the number of bits to be hidden in content) as the probability of error decreases. Fingerprint designers have to strike a trade-off, which is hard to conceive when only rough estimation of the probability of error is known. The major issue for fingerprinting algorithms is the fact that embedding large sequences implies also assessing reliability on a huge amount of data, which may be practically unachievable without using rare event analysis.

In collaboration with the team-projects ASPI and ALEA, we developed a novel strategy for simulating rare events and an associated Monte Carlo estimation of tail probabilities. Our method uses a system of interacting particles and exploits a Feynman-Kac representation of that system to analyze their fluctuations. Our precise analysis of the variance of a standard multilevel splitting algorithm reveals an opportunity for improvement. This leads to a novel method that relies on adaptive levels and produces, in the limit of an idealized version of the algorithm, estimates with optimal variance. Some numerical results show performance close to the idealized version of our technique for these practical applications. This work has been published in the journal Statistics and computing [13]. Algorithms for estimating extreme probabilities and quantiles are implemented as a Matlab package.

6.1.4.3. New decoders for fingerprinting

Participant: Teddy Furon.

So far, the accusation process of a Tardos fingerprinting code is based on single decoders which compute a score per user. Users with the highest score or whose scores is above a threshold are then deemed guilty. In the past years, we have contributed to this approach bringing two improvements: the 'learn and match' strategy aims at estimating the collusion process and using the matched score function; a rare event analysis translates this score into a more meaningful probability of being guilty. A fast implementation computes the scores of one million of users within 0.2 second on a regular laptop. Therefore, contrary to common belief, although a single decoder is exhaustive with a linear complexity in $O(n)$, it is not slow.

This fast implementation allows us to propose iterative decoders. A first idea is that conditioning by the identities of some colluders bring more discrimination power to the score function. The first iteration is thus a single decoder, users we are extremely confident to accuse are enrolled as side information. The next iteration computes new scores for the remaining users etc. A second idea is that information theory proves that a joint decoder computing scores for pairs, triplets, or in general t-tuples 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 $O(n^t)$. We propose in a first iteration to use a single decoder, to prune out users who are definitively innocents (because their scores are low) and keeping $O(\sqrt{n})$ individual suspects. The second iteration is a joint decoding working on pairs of users etc. Iteratively, we prune out enough users such that it is manageable to run a joint decoder on bigger t-tuples. This work has been done under a collaboration of TEMICS, and published in a series of conference papers [37], [38], [36]. A journal version has been submitted to IEEE Trans. on Information Forensics and Security. A Tardos code software suite (generation of code, collusion attacks, accusation algorithms) is available as a C package.

6.1.4.4. Protocols for fingerprinting

Participant: Teddy Furon.

A key assumption of the fingerprinting schemes developed so far is that the colluders may know their own codewords but they ignore the codeword of any other innocent user. Otherwise, the collusion can very easily forge a pirated content framing an innocent user because it contains a sequence close enough to his/her codeword. This puts a lot of pressure on the versioning mechanism which creates the personal copy of the content in accordance to a codeword. For instance, suppose that the versioning is done in the user's setup box, the unique codeword being loaded into this device at the manufacture. If the code matrix ends up in the hands of an untrustworthy employee, then the whole fingerprinting system is pulled down. This is one argument of the motivation for designing cryptographic protocols for the construction, the versioning and the accusation. We have proposed a new asymmetric fingerprinting protocol dedicated to the state-of the-art Tardos codes. We believe that this is the first such protocol, and that it is practically efficient. The construction of the fingerprints and their embedding within pieces of content is based on oblivious transfer and do not need a trusted third party. Note, however, that during the accusation stage, a trusted third party, like a Judge, is necessary like in any asymmetric fingerprinting scheme we are aware of. This works has been done in collaboration with the team-project TEMICS, Lab-STICC Telecom Bretagne and University College London, and presented at Information Hiding [22]. Ana Charpentier defended her PhD. thesis in October 2011 [72].

6.1.4.5. Reconstructing an image from its local descriptors

Participant: Hervé Jégou.

We show [47] that an image can be approximately reconstructed based on the output of a black-box local description software such as those classically used for image indexing. Our approach consists first in using an off-the-shelf image database to find patches which are visually similar to each region of interest of the unknown input image, according to associated local descriptors. These patches are then warped into input image domain according to interest region geometry and seamlessly stitched together. Final completion of still missing texture-free regions is obtained by smooth interpolation. As demonstrated in our experiments, visually meaningful reconstructions are obtained just based on image local descriptors like SIFT, provided the geometry of regions of interest is known. The reconstruction allows most often the clear interpretation of the semantic image content. As a result, this work raises critical issues of privacy and rights when local descriptors of photos or videos are given away for indexing and search purpose.

6.2. Multi-dimensional indexing and clustering

6.2.1. Improved NV-tree

Participant: Laurent Amsaleg.

This is a joint work with Björn Þór Jónsson from the School of Computer Science, Reykjavik University, Iceland and with Herwig Lejsek, Videntifier Technologies, Iceland.

We have further improved the NV-Tree (Nearest Vector Tree) indexing techniques. It addresses the specific, yet important, problem of efficiently and effectively finding the approximate k-nearest neighbors within a collection of a few billion high-dimensional data points. The NV-Tree is a very compact index, as only six bytes are kept in the index for each high-dimensional descriptor. It thus scales extremely well when indexing large collections of high-dimensional descriptors. The NV-Tree efficiently produces results of good quality, even at such a large scale that the indices cannot be kept entirely in main memory any more. We have demonstrated this with extensive experiments using a collection of 2.5 billion SIFT (Scale Invariant Feature Transform) descriptors. Additional experiments involving more than 30 billion SIFT descriptors show results are still of a good quality and that disks are handled as efficiently as they can be.

6.2.2. Indexation of time series

Participants: Laurent Amsaleg, Romain Tavenard.

Dynamic Time Warping (DTW) is the most popular approach for evaluating the similarity of time series, but its computation is costly. Therefore, simple functions lower bounding DTW distances have been designed, accelerating searches by quickly pruning sequences that could not possibly be best matches. The tighter the bounds, the more they prune and the better the performance. Designing new functions that are even tighter is difficult because their computation is likely to become complex, canceling the benefits of their pruning. It is possible, however, to design simple functions with a higher pruning power by relaxing the *no false dismissal* assumption, resulting in approximate lower bound functions. We have discover how very popular approaches accelerating DTW such as LB_Keogh and LB_PAA can be made more efficient *via* approximations. The accuracy of approximations can be tuned, ranging from no false dismissal to potential losses when aggressively set for great response time savings. At very large scale, indexing time series is mandatory. These approximate lower bound functions can be used with iSAX. Furthermore, we have also observed that a k-means-based quantization step for iSAX gives significant performance gains.

6.2.3. Improved image indexing with asymmetric Hamming embedding

Participants: Patrick Gros, Mihir Jain, Hervé Jégou.

We have proposed [28] an improved asymmetric Hamming Embedding scheme for large scale image search based on local descriptors. The comparison of two descriptors relies on an vector-to-binary code comparison, which limits the quantization error associated with the query compared with the original Hamming Embedding method. The approach is used in combination with an inverted file structure that offers high efficiency, comparable to that of a regular bag-of-features retrieval systems, and consistently improves the search quality over the symmetric version on the two datasets used for the evaluation.

6.2.4. Compression techniques for nearest neighbor search

Participants: Laurent Amsaleg, Teddy Furon, Hervé Jégou, Romain Tavenard.

*Part of this work on this topic was done in cooperation with Matthijs Douze and Cordelia Schmid (INRIA/*LEAR*).*

6.2.4.1. Re-ranking with source coding

An extension of our previous work on source coding techniques for high-dimensional indexing has been proposed [29]. The goal is to index a large set of vectors, as large as 1 billion vectors, with limited CPU and memory usage. Based on the product quantization-based indexing technique [18], we show that it is interesting to add an additional level of processing to refine the estimated distances. It consists in quantizing the difference vector between a point and the corresponding centroid. When combined with an inverted file, this gives three levels of quantization. Experiments performed on SIFT and GIST image descriptors show excellent search accuracy outperforming three state-of-the-art approaches. Compared with the original work [18], the proposed re-ranking technique is shown to obtain better trade-off with respect to memory, efficiency and search quality.

6.2.4.2. Anti-sparse coding for approximate nearest neighbor search

Following recent works on Hamming Embedding techniques, we propose [67] 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.

6.2.5. Architecture-aware indexing techniques for solid state disks

Participants: Laurent Amsaleg, Gylfi Gudmundsson.

This is a joint work with Björn Þór Jónsson from the School of Computer Science, Reykjavik University, Iceland.

The scale of multimedia data collections is expanding at a very fast rate. In order to cope with this growth, the high-dimensional indexing methods used for content-based multimedia retrieval must adapt gracefully to secondary storage. Recent progress in storage technology, however, means that algorithm designers must now cope with a spectrum of secondary storage solutions, ranging from traditional magnetic hard drives to state-of-the-art solid state disks. We have analyzed the impact of storage technology on a simple, prototypical high-dimensional indexing method for large scale query processing. We found that while the algorithm implementation deeply impacts the performance of the indexing method, the setup of the underlying storage technology is equally important.

6.3. New techniques for linguistic information acquisition and use

6.3.1. NLP for document description

6.3.1.1. Semantic annotation of multimedia documents based on textual data

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

This work is done in the framework of the Quaero project (see below).

On this subject, TEXMEX is implied in three tasks of the Quaero project.

The first task concerns the extraction of terminology from document. The objective of this work is to study the development and the adaptation of methods to automate the acquisition and the structuring of terminologies . In this context, in 2011, we have undergone a new evaluation of terminology extraction systems. Here again, our system, relying on TermoStat (see previous reports) ranked first for the tracks in which we participated. We have also continued our work the use of morphology for biomedical terminologies. This approach relies on the decomposition of terms into morphemes and the translation of these morphemes into japanese (kanji) sub-words. The kanji characters thus offer a semantic way to access the semantics of the morpheme and allow us to detect semantic relations between them. We have tested this approach on more languages and have proved its relevance for information retrieval problems.

The second task aims at extracting semantic and ontological relations from documents. Indeed, detecting semantic and ontological relations in texts is a key to describe a domain and thus manipulate cleverly documents. In 2011, we developed a new relation extraction system based on k-nearest-neighbors and language modeling. It has been tested in the framework of the Quaero evaluation campaign and ranked first or second for all tracks. We have also developed a clustering technique for named entities. It relies on new representation schemes called bag-of-vectors (or bag-of-bags-of-features), which perform better than the classical bag-of-word approach.

The last task directly deals with the semantic annotation of multimedia documents based on textual data, for, very often, many textual or language-related data can be found in multimedia documents or come along such documents. For example, a TV-broadcast, contains speech that can transcribed, Electronic Program Guide and standard program guide information, closed captions, associated websites, etc. All these sources offers a way to exploit complementary information that can be used to semantically annotate multimedia document. During this year, we finished the development of a football multimedia corpus. It contains the video of several matches, the speech transcripts, associated textual data from specialized websites. All these media have been manually annotated in terms of events, named entities, specialized relations (fouls, replacements, etc) and other relevant information. This corpus will be distributed under LGPL-LR license.

6.3.1.2. Text recognition in videos

Participants: Khaoula Elagouni, Pascale Sébillot.

This work is done in the context of a joint TEXMEX*/Orange Ph.D. thesis supported by a CIFRE grant with Orange Labs.*

We aim at helping multimedia content understanding by obtaining benefit from textual clues embedded in digital video data. In 2011, we proposed an Optical Character Recognition-based method to recognize natural scene texts in images, avoiding the conventional character segmentation step. The text image is scanned with multi-scale windows and a robust recognition model is applied on each window, relying on a neural classification approach, to identify non valid characters and recognize valid ones. A graph model is used to represent spatial constraint between recognition results, and to determine the best sequence of characters. Some linguistic knowledge is also incorporated in the graph to remove errors due to recognition confusions. The method was evaluated on the ICDAR 2003 database of scene text images and outperforms state-of-the-art approaches. This work will be presented at DAS2012.

6.3.1.3. DEFT evaluation campaign participation

Participants: Vincent Claveau, Christian Raymond.

Christian Raymond and Vincent Claveau participated to DEFT (http://deft2011.limsi.fr/). Two tasks were proposed: the first one was called "the diachronic variation task" whose objective was to identify the writing year of some OCR newspapers from 1801 to 1944. The second one was a abstract/article pairing task. Their approaches based on boosting and k-nearest neighbors was ranked first on the difficult diachronic task.

6.3.2. Oral and textual information retrieval

6.3.2.1. Graded-inclusion-based Information retrieval systems

Participants: Vincent Claveau, Laurent Ughetto.

Our work on this topic is done in close collaboration with Olivier Pivert from the PILGRIM *project-team of IRISA Lannion.*

Databases (DB) querying mechanisms, and more particularly the division of relations was at the origin of the Boolean model for Information Retrieval Systems (IRSs). This model has rapidly shown its limitations and is no more used in Information retrieval (IR). Among the reasons, the Boolean approach do not allow to represent and use the relative importance of terms indexing the documents or representing the queries. However, this notion of importance can be captured by the division of fuzzy relations. This division, modeled by fuzzy implications, corresponds to graded inclusions. Theoretical work conducted by the PILGRIM project-team have shown the interest of this operator in IR.

Our first work was to investigate the use of graded inclusions to model the information retrieval process. In this framework, documents and queries are represented by fuzzy sets, which are paired with operations like fuzzy implications and T-norms. Through different experiments, we have shown that only some among the wide range of fuzzy operations are relevant for information retrieval. When appropriate settings are chosen, it is possible to mimic classical systems, thus yielding results rivaling those of state-of-the-art systems. These positive results have validated the proposed approach, while negative ones have given some insights on the properties needed by such a model.

More recently, the links between our fuzzy model and other classical IR models have been studied. It has been shown that our fuzzy implication-based model can be interpreted as several classical models: an Extended Boolean Model, a Logical Model, a Vector Space Model or a Language Model in IR.

6.3.2.2. Information retrieval in TV streams using automatic speech recognition

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

Automatic speech recognition outputs are by nature incomplete and uncertain, so much that lexical indexes of speech are not sufficient to overcome the errors due to out-of-vocabulary words and to most of the named entities, consisting in important semantic information from the discourse. Using if necessary a phonetic index is a solution to retrieve partially the mis-recognized words but at the price of a lower precision because the phonetic representation is also noisy. We proposed this year (still to be submitted) an indexation method which jointly model lexical and phonetic levels with finite-state transducers, offering the possibility to take a lexical path or a phonetic path between two synchronization nodes. The edges are weighted by a vector of features (edition scores, confidence measures, durations) that will be used in a supervised manner to estimate

the reliability of the returned result at the search step. The experiments have shown the complementary of lexical-phonetic representations and their contribution for a task of spoken utterance retrieval using named entity queries.

6.4. New processing tools for audiovisual documents

6.4.1. TV stream structuring

6.4.1.1. Repetition detection-based TV structuring

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

We work on the issue of structuring large TV streams. More precisely, we focus on the problem of labeling the segments of a stream according to their types (*e.g.*, programs, commercial breaks, sponsoring, etc). Contrary to existing techniques, we wanted to take into account the sequential aspect of the data, and thus we used Conditional Random Fields (CRF), a classifier which has proved useful to handle sequential data in other domains like computational linguistics or computational biology. During this year, we proved the relevance of CRF in the framework of TV segments labeling. We conducted different experiments, either on manually or automatically segmented streams, with different label granularities, and demonstrated that this approach rivals existing ones. The use of this model for semi-supervised and unsupervised learning are under study.

6.4.2. Program structuring

6.4.2.1. Audiovisual models for event detection in videos

Participants: Guillaume Gravier, Patrick Gros, Cédric Penet.

This work was performed in close collaboration with Technicolor as external partner.

Following our work on the detection of audio concepts related to violence in movie soundtracks [58], we developed a system for the detection of violent scenes in movies, combining multimodal features. We investigated multimodal fusion strategies and temporal integration exploiting Bayesian networks as a joint distribution model. Several strategies for learning the structure of the Bayesian networks were compared, resulting in a complete system for violence detection. The system was evaluated on the Violent Scenes Detection task of the MediaEval 2011 international evaluation [42] that we co-organized with Technicolor and the University of Geneva [62]. A fair amount of time was dedicated this year to the organization of the evaluation campaign which includes defining the task and metrics, supervising the annotation, recruiting participants, analyzing the results and organizing the corresponding workshop session.

6.4.2.2. Unsupervised multimedia content mining

Participants: Guillaume Gravier, Anh Phuong Ta.

This work on audio content discovery was partially carried out in collaboration with Armando Muscariello and Frédéric Bimbot from the Metiss project-team.

As an alternative to supervised approaches for multimedia content analysis, where predefined concepts are searched for in the data, we investigate content discovery approaches where knowledge emerge from the data. Following this general philosophy, we pursued work on motif discovery in audio and video content.

Audio motif discovery is the task of finding out, without any prior knowledge, all pieces of signals that repeat, eventually allowing variability. In 2011, we extended our recent work on seeded discovery to near duplicate detection and spoken document retrieval from examples. First, we proposed algorithmic speed ups for the discovery of near duplicate motifs (low variability) in large (several days long) audio streams, exploiting subsampling strategies [39]. Second, we investigated the use of previously proposed efficient pattern matching techniques to deal with motif variability in speech data [40] in a different setting, that of spoken document retrieval from an audio example. We demonstrated the potential of model-free approaches for efficient spoken document retrieval on a variety of data sets, in particular in the framework of the Spoken Web Search task of the MediaEval 2011 international evaluation [41].

Video structure is often enforced through editing rules which result in a set of shots defining an event that repeats throughout the video with a high visual and audio similarity. Typical such shots are anchor persons and close-up on guests in talk-shows. We recently proposed an unsupervised multimodal approach to discover such events exploiting audio and visual consistency between two sets of independent nested clusters, one for each modality [21]. In 2011, we extended the approach in two directions. First, we improved the selection of consistent audio and visual clusters and the unsupervised selection of positive and negative examples exploiting redundancy between nested clusters. Second, we extended the method to discover several audio-visually consistent events rather than a single one in our previous work, thus enabling the use of unsupervised mining as a pre-processing step for video structure analysis.

6.4.2.3. Topic segmentation with vectorization and morpho-mathematics

Participant: Vincent Claveau.

Our work on this topic is done in close collaboration with Sébastien Lefèvre from the SEASIDE *project-team of IRISA Vannes.*

Segmenting a program into topics is an important step for fine-grained structuring of TV streams. Based on our work on vectorization (see previous reports), we have developed a new segmentation technique using speech transcripts. Making an analogy with image segmentation, we have adapted the watershed transform to handle these textual data and more precisely the distances computed by vectorization between possible segments.

This method has been tested on different TV collections (news, reports) as well as more usual text collection used for segmentation evaluation. In every cases, our technique has outperformed any state-of-the-art approaches.

6.4.3. Using speech to describe and structure video

Participants: Camille Guinaudeau, Guillaume Gravier, Ludivine Kuznik, Bogdan Ludusan, Pascale Sébillot.

Speech can be used to structure and organize large collections of spoken documents (videos, audio streams, etc) based on semantics. This is typically achieved by first transforming speech into text using automatic speech recognition (ASR), before applying natural language processing (NLP) techniques on the transcripts. Our research focuses firstly on the adaptation of NLP methods designed for regular texts to account for the specific aspects of automatic transcripts. In particular, we investigate a deeper integration between ASR and NLP, i.e., between the transcription phase and the semantic analysis phase.

In 2011, we mostly focused on robust transcription, hierarchical topic segmentation and collection structuring.

On the one hand, we investigated the use of broad phonetic landmarks and syllable prominence to improve large vocabulary speech recognition by guiding the Viterbi search process. Several mechanisms to incorporate landmarks into the search space were studied. Significant improvements were observed on radio broadcast news data in the French language. On the other hand, we pursued our work on unsupervised topic adaptation, focusing on the automatic selection of out-of-vocabulary words combining phonetic and morpho-syntactic criteria.

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. We proposed new probabilistic measures of the lexical cohesion to emphasize the contribution of words that appears only locally, thus attenuating the impact of words which contributed to the segments at an upper level of the hierarchy [11].

Finally, we initiated work in collaboration with INA on structuring a large collection of news reports. The idea is to automatically create links and threads between reports in several months of broadcast news shows, based either on the documentary records of the shows and/or on the automatic transcripts. As preliminary step towards this goal, we investigated distances between documentary records in an information retrieval setting so as to construct a nearest neighbor graph. The next step consists in exploiting graph clustering methods.

Our research in speech for TV content structuring was illustrated through the Texmix demonstration (see Section 5.2) which exploits most of our achievements in the field, including transcription, topic segmentation and collection structuring.

7. Contracts and Grants with Industry

7.1. Contracts with industry

7.1.1. Pôle de compétitivité

Participants: Patrick Gros, Sébastien Campion.

The French government organized in 2005 competitiveness poles (*pôles de compétitivité*) in France to strengthen ties in given regions between industries (big and small companies), research labs (both public and private ones) and teaching institutions (universities and schools of engineering). In 2011, the pole actively prepared a proposal to build an "IRT" (Institut de Recherche Technologique), a new tool proposed by the government to foster innovation and transfer between academic and industrial partners. Texmex is involved in this project, and is responsible for one of its experimental platform. Until Oct 1st, Patrick Gros was also deputy member of the executive committee and the project selection committee.

7.2. Grants with industry

7.2.1. Contract with Technicolor

Participants: Guillaume Gravier, Patrick Gros, Cédric Penet.

Duration: 36 months, since September 15th 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 this work is to study and develop techniques based on stochastic models to analyze the content of movies. The application developed in Technicolor consists in detecting violent scenes in movies in order to facilitate parental supervision.

7.2.2. 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.

7.2.3. 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.

7.3. European initiatives

7.3.1. Quaero

Participants: Laurent Amsaleg, Mathieu Ben, Sébastien Campion, Vincent Claveau, Ali Reza Ebadat, Julien Fayolle, Patrick Gros, Gylfi Gudmundsson, Camille Guinaudeau, Carryn Hayward, Hervé Jégou, Ewa Kijak, Fabienne Moreau, Stacy Payne, 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 2011, the Quaero team of TEXMEX was mainly affected by the leave of Mathieu Ben, our technical coordinator and of Stacy Payne our financial coordinator. S. Payne was replaced by Carryn Hayward. Among the key fact of our participation this year is our participation to Trecvid.

7.3.2. IET ICT Labs - Opensem project

Participants: Morgan Bréhinier, Sébastien Campion, Guillaume Gravier, Teddy Furon, Patrick Gros, Hervé Jégou.

Duration: 1 year, starting January 2011.

OpenSEM is a project of the EIT KIC ICT Labs grouping 5 academic partners: TU Delft (The Netherlands), VTT (Finland), TU Berlin (Germany), Institut Eurecom (France) and INRIA Rennes.

The project (See http://www.opensem.eu) builds a virtual center of excellence in order to speed up and maximize the potential for innovation in semantic media:

- Maximizing the open dissemination and impact of existing knowledge, tangible results (software, tools, demonstrations, field trial results), and rich social content (multimedia, plus metadata such as tag and ratings, plus social network information).
- Driving the immediate potential for the triple synergy between content-based analysis, user-based collaborative analysis and social networks and community building through large scale benchmarking competitions (MediaEval).

Participation to the project includes contributing software and demonstration to the OpenSEM portal as well as organizing and participating to the MediaEval 2011 benchmark initiative. As a particularly visible action, we developed the Texmix demo interface that allows the demonstration, on a corpora of news reports provided by INA, the work that was developed in the team on topic segmentation, keyword extraction, image retrieval, named entity extraction and classification. This demo was demonstrated during the fall Quaero plenary and during the INRIA-industry special day on future TV.

7.3.3. FIIA: Forensic image identifier and analyzer

Participants: Laurent Amsaleg, Ewa Kijak.

Duration: 32 months, starting November 2011. Prime: Videntifier Technologies.

FIIA is an innovative software service for the Forensic market that automatically identifies and analyzes the content of images on web sites and seized computers. The service saves time and money, gathers better evidence, and builds stronger court cases. We are in charge of helping with the technology needed to identify the logos from terrorist organizations that are inserted in images or videos. Challenges are related to the poor resolution and small size of logos as well as to the very strict efficiency constraints that the logo detector must match.

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) 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.

8.2. International initiatives

8.2.1. Visits of international scientists

8.2.1.1. Visit to Delft University of Technology
Participant: Guillaume Gravier.

Guillaume Gravier was invited to the Multimedia Information Retrieval Lab at Delft University of Technology for one week in May 2011. He gave a seminar entitled Speech, Language and Multimedia at IRISA/INRIA Rennes and participated in the organization of MediaEval 2011, an international benchmark initiative in multimedia processing. A EU project involving TU Delft and IRISA/INRIA Rennes, initiated during his visit in Delft, has been submitted to the FET Open program.

8.2.1.2. Visit to the BUSIM speech processing group at Bogazici University
Participant: Julien Fayolle.

Julien Fayolle spent three months, from May to July 2011, in the BUSIM speech processing group at Bogazici University (Istanbul, Turkey) to work on lexical-phonetic automata for spoken utterance retrieval in collaboration with Murat Saraçlar. Whereas the state-of-the-art approaches consist in a late fusion of the results of phonetic and lexical searches, the idea was to adapt the Murat Saraclar's spoken utterance retrieval methods to a new representation combining both lexical and phonetic levels earlier than the retrieval step.

8.2.1.3. Doctoral Internship from Florida State University
Participants: Guillaume Gravier, Patrick Gros, Hervé Jégou.

Jiangbo Yuan spent five months in the TEXMEX project-team to work on audio indexing for video copy detection. He has contributed to the submission of INRIA at copy detection task, working on our audio indexing engine, more precisely on the post-verification step. He then worked on the detection of near-duplicate patterns in very large datasets of vectors.

His venue was funded by the EIT ICT Labs OpenSEM project.

8.2.1.4. Visit to the Czech Technical University in Prague
Participant: Hervé Jégou.

Hervé Jégou spent one week (October 2011) in the Center of Machine Perception at the Czech Technical University, Prague, to initiate a collaboration with Pr. Jiri Matas and Dr. Ondrej Chum. During this visit, he gave an invited talk at the 29th Pattern Recognition and Computer Vision Colloquium, entitled "Approximate search as a source coding problem, with application to large scale image retrieval". This visit was the opportunity to start a joined work on image search.

8.2.1.5. Visit of members of the University of Reykjavík and Videntifier Technologies
Participant: Laurent Amsaleg.

Björn Þór Jónsson from the School of Computer Science, Reykjavik University, Iceland and with Herwig Lejsek, Videntifier Technologies, Iceland spent one week in the team. They came to participate to the large scale high dimensional indexing experiments involving more than 30 billion SIFT descriptors.

9. Dissemination

9.1. Animation of the scientific community

- Laurent Amsaleg
  - was a program committee member of CIVR 2010;
  - was a program committee member of ACM Multimedia 2011;
  - was a program committee member of ICMR 2011;
  - was a program committee member of MMM 2011;
  - was a program committee member of VLDB 2011 PhD Forum;
  - was a reviewer for IET Information Security;
  - was a reviewer for EURASIP Journal on Advances in Signal Processing;
  - was the publicity chairman of ACM Multimedia 2011;
  - was a member of the "commission de spécialistes, Toulon";
  - was a member of the "commission de spécialistes, Nantes".
- Vincent Claveau
  - was a program committee member of WI'11 (International conference on Web Intelligence), Grenoble, France, July 2011;
  - was a program committee member of TALN'11 (17^{e} conférence nationale Traitement automatique des langues naturelles), Montpellier, France, July 2011;
  - was a program committee member of RECITAL'11, Montpellier, France, July 2011;
  - was a program committee member of SIIM, (Symposium sur l'Ingénierie de l'Information Médicale), Toulouse, France, June 2011;
  - was a program committee member of Conférence en Recherche d'Information et Applications, CORIA 2011, Avignon, France, March 2011;
  - is a member of the editorial board of the journal TAL, Traitement Automatique des Langues;
  - was a reviewing committee member for the journal Multimedia Tools and Applications (MTAP).
- Teddy Furon
  - was a program committee member of Information Hiding 2011, Prague, Czech Republic;
  - was a program committee member of CMS 2011, Ghent, Belgium;
  - was a program committee member of ISPA 2011, Dubrovnik, Croatia;
  - was a program committee member of IEEE ICME 2011, Barcelona, Spain;
  - was a program committee member of EUSIPCO 2011, Barcelona, Spain;
  - was a program committee member of IEEE WIFS 2011, Foz do Iguaçu, Brazil;
  - was an evaluator for the French ANR, 2011;
  - was an associate editor of EURASIP Journal on Information Security, 2011;
  - was an associate editor of IET Journal of Information security, 2011;
  - was a member of the technical committee of IEEE Information Forensics and Security subsociety, 2011.
- Guillaume Gravier

  - was a technical chair of CBMI 2011, Madrid, Spain;
  - was a program committee member of MediaEval 2011;
  - is vice-president of the French Speech Communication Association (AFCP);
  - is the scientific leader of the French ANR project ETAPE targeting an evaluation campaign on speech technologies for multimedia;
  - co-founded in 2011 the Speech and Language in Multimedia (SLIM) Special Interest Group of the Intl. Speech Communication Association (ISCA).

- Patrick Gros
  - was a co-organizer of the International ACM Workshop on Automated Media Analysis and Production for Novel TV Services (AIEMPro 2011) that took place during the ACM International conference on Multimedia in Scottsdale, Arizona, USA, in December 2011;
  - was a program committee member of the ninht International Workshop on Content Based Multimedia Indexing (CBMI) Which was held in MAdrid, Spain, France in June 2011;
  - is a member of the steering board of the Content Based Multimedia Indexing (CBMI) workshop series;
  - was a program committee member of RFIA'12, 18ème conférence en Reconnaissance des Formes et Intelligence Artificielle, Lyon, France, January 2012;
  - was a scientific committee member of the "Conférence en Recherche d'Information et Applications (CORIA) 2011 - Avignon, March 2011;
  - was a program committee member of the Second International Conference on Creative Content Technologies CONTENT, Roma, Italy, September 2011.

- Hervé Jégou
  - was a program committee member of CVPR 2011;
  - was a program committee member of ICCV 2011;
  - was an evaluator for the French ANR, 2011;
  - was an expert for the program "Futur et Rupture" of the Institut Télécom;
  - was reviewer for several journals, in particular for the International Journal of Computer Vision, IEEE Transactions on Pattern Analysis and Machine Intelligence and IEEE Transactions on Multimedia.

- Annie Morin
  - was a vice -president of the CNU, Conseil national des Universités, in Computer Science;
  - was a member of ITI (Information technology Interfaces) 2011 conference.

- François Poulet
  - was a program committee member of VINCI'11, Visual INformation Communications International, Hong-Kong, China, August 2011;
  - was a program committee member of AusDM'11, Autralasian Data Mining Conference, Ballarat, Australia, December 2011;
  - was a program committee member of EGC'11, Extraction et Gestion de Connaissances, Brest, France, January 2011;
  - was a reviewer of ESANN 2001, European Symposium of Artificial Neural Network, Bruges, Belgique, April 2011;
  - was a reviewer of NeuroComputing;
  - was a reviewer of EJOR, European Journal of Operational Research;

  - was co-organizer of the 9th workshop Visualisation et Extraction de Connaissances, (AVEC-EGC'11), Brest, France, January 2011.

- Christian Raymond
  - is a member of the editorial board of the e-journal "Discours", http://discours.revues.org;
  - was a reviewing committee member for the journal CSL, Computer Speech and Language;
  - was a reviewing committee member of Interspeech (12th Annual Conference of the International Speech Communication Association), Florence, Italy;
  - was a reviewing committee member of ICMLA (The tenth International Conference on Machine Learning and Applications), Honolulu Hawaii, USA.

- Pascale Sébillot
  - was a member of the program committee of CORIA 2011 (8e conférence en recherche d'information et applications), Avignon, France, March 2011;
  - was a member of the program committee of TALN 2011 (18e conférence francophone Traitement automatique des langues naturelles), Montpellier, France, June-July 2011;
  - was a member of the program committee of TIA 2011 (9e conférence Terminologie et intelligence artificielle), Paris, France, November 2011;
  - 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 2011.

- Laurent Ughetto
  - was a program committee member of the Rencontres Francophones sur la Logique Floue et ses Applications (LFA'11).

9.2. Invited talks and prizes

9.2.1. Invited talks

- Teddy Furon. Talk at GdR ISIS French national day on fingerprinting, July 2011
- Hervé Jégou. Talk at GdR ISIS French national day on fingerprinting, July 2011
- Hervé Jégou. Invited talk at INESC Porto, Portugal, July 2011
- Hervé Jégou, Invited talk at CVUT Pragua at the 29th Pattern Recognition and Computer Vision Colloquium, Czech Republic, October 13th 2011
- Hervé Jégou, Talk at the Trecvid Workshop, Gaithersburg, USA, December 2011

9.2.2. Prizes

- Gwénolé Lecorvé was awarded the best Ph.D. award of the French Speech Communication Association.
- Hervé Jégou was awarded an Outstanding Reviewer Award at CVPR '11.
- TEXMEX has participated to the Trecvid copy detection task in 2011. This joint submission [61] with LEAR project-team was ranked roughly 3rd out of 21 participants. On the "No-False alarm" profile, our submission was best for 23 of the 56 transformation types for the optimal NDCR measure.
- TRECVID semantic indexing task: Hervé Jégou has contributed to the Quaero submission [63], jointly with LIG (main contributor) and Karlsruhe Institute of Technology. This submission was ranked 3rd out of 19 participants.

- TEXMEX was ranked first on the diachronic task (identification of writing year of OCR newpapers) at the DEFT evaluation campaign.

9.3. Teaching

9.3.1. Main teaching activities and responsabilities

Laurent Amsaleg: Managing Large Collections of Digital Data, 14h, M2 R, University Rennes 1, France.

Laurent Amsaleg: Advanced Databases, 8h, Master, ENSAI, France.

Vincent Claveau: Symbolic Sequential Data, Master by research in computer science 2nd year (8 students, 7 hours), University of Rennes 1, France.

Vincent Claveau: Multimedia Databases, engineer diploma 3rd year, (14 students, 10 hours), ENSSAT - University of Rennes 1, Lannion, France.

Patrick Gros: coordinates the track "From Data to Knowledge: Machine Learning, Modeling and Indexing Multimedia Contents and Symbolic Data" of the Master by research in computer science (2^{nd} year), University of Rennes 1, Rennes. He is responsible of the Math workshop of the master (10h).

Camille Guinaudeau: Analysis of audiovisual documents and flows for indexing (7h), Master by research in computer science 2nd year (M2), University of Rennes 1, France.

Annie Morin: Data Analysis L3 MIAGE ISTIC University of Rennes 1 (38 hours, 60 students).

Annie Morin: Short term forecast M1 MIAGE ISTIC University of Rennes 1 (40 hours, 20 students).

Annie Morin: Statistical Data Mining M2 MIAGE ISTIC University of Rennes 1 (38 hours, 20 students).

Annie Morin: Statistical Process control and Reliability M2 Micro Electronics ISTIC University of Rennes 1 (37 hours, 10 students).

Annie Morin: Statistical Process control and Reliability International M2 Telecom and Electronics, South East University, Nanjing, China (40 hours, 15 students).

Ewa Kijak is head of the Image engineering track of ESIR, the engineering formation of University of Rennes 1, France.

Ewa Kijak: Image processing (50h), Image analysis and classification (24h), engineer diploma 2nd year (M1), ESIR - University of Rennes, France.

Ewa Kijak: Computer vision: Image indexing and retrieval (10h), engineer diploma 3rd year (M2), ESIR - University of Rennes, France.

Ewa Kijak: Multimedia Databases (10h), engineer diploma 3rd year (M2), ENSSAT - University of Rennes 1, Lannion, France.

Ewa Kijak and Camille Guinaudeau: Digital Documents Indexing and Retrieval (22h and 10h), Professional Master in Computer Science 2nd year (M2), ISTIC, University of Rennes 1, France.

François Poulet is in charge of the Master in computer science, M2, MITIC, Computer Science Methods and Information and Communication Technologies, ISTIC, University of Rennes 1.

François Poulet: Managing Large Collections of Digital Data. Master by research in computer science, M2, ISTIC, University of Rennes 1, 10h EqTD.

François Poulet: Supervised Learning. Master by research in computer science, M2, ISTIC, University of Rennes 1, 16h EqTD.

François Poulet: Introduction to Data Mining. Professional Master in Computer Science, M2, ISTIC, University of Rennes 1, 15h EqTD.

François Poulet: Mining Symbolic Data. Professional Master in Computer Science, M2, ISTIC, University of Rennes 1, 25h EqTD.

François Poulet: Applications and Problem Solving. Professional Master in Computer Science, M2, ISTIC, University of Rennes 1, 10h EqTD.

François Poulet: Learning Methods for Multimedia Data. Professional Master in Computer Science, M2, ISTIC, University of Rennes 1, 26h EqTD.

François Poulet: Algorithms and Functional Programming. Computer Science Licence, L1, ISTIC, University of Rennes 1, 60h EqTD.

Pascale Sébillot was course co-director of the Research in Computer Science specialism of the Master's in Computer Science (2nd year), University of Rennes 1, till September 12, 2011.

Pascale Sébillot: : Advanced Databases and Modern Information Systems, 69 hours, M2, INSA de Rennes, France.

Pascale Sébillot: : Data-Based Knowledge Acquisition 2: Symbolic Methods, 18 hours, M1, INSA de Rennes, France.

9.3.2. PhD

PhD : Romain Tavenard, Indexation de séquences de descripteurs, University Rennes 1, defended July 4th 2011, Laurent Amsaleg & Patrick Gros.

PhD : Camille Guinaudeau, Structuration automatique de flux télévisuels, INSA de Rennes, defended December 7th, 2011, Guillaume Gravier and Pascale Sébillot.

PhD in progress : Juan David Cruz-Gomez, Algorithmique de réseaux socio-sémantiques pour la Visualisation par point de vue de communautés en ligne, since December 2009, Cécile Bothorel (Télécom Bretagne), François Poulet.

PhD in progress : Thanh Toan Do, Challenging the security of CBIR systems, 2nd year, Laurent Amsaleg, Ewa Kijak, Teddy Furon.

PhD in progress : Thanh-Nghi Doan, Image Classification, since November 2010, François Poulet.

PhD in progress : Ali Reza Ebadat, Annotation de documents multimédias à partir d'indices textuels, October 10th, 2009, Vincent Claveau and Pascale Sébillot.

PhD in progress : Khaoula Elagouni, Indexation automatique d'images et de vidéos par reconnaissance automatique de textes incrustés et traitement automatique des langues, October 10th, 2009, Pascale Sébillot, Christophe Garcia (LIRIS, Lyon) and Franck Mamalet (Orange Labs, Rennes).

PhD in progress : Gylfi Gudmundsson, Towards parallel and distributed CBIR systems, 2nd year, Laurent Amsaleg.

PhD in progress : Mihir Jain, Video description and indexing, since February 2011, Hervé Jégou and Patrick Gros.

PhD in progress : Ludivine Kuznik, Structuration et navigation dans des archives documentaires, April 18th, 2011, Guillaume Gravier, Pascale Sébillot and Jean Carrive (INA).

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] M. DELAKIS, G. GRAVIER, P. GROS. *Audiovisual Integration with Segment Models for Tennis Video Parsing*, in "Computer Vision and Image Understanding", August 2008, vol. 111, n^{o} 2, p. 142–154.

[4] M. DOUZE, H. JÉGOU, H. SINGH, L. AMSALEG, C. SCHMID. *Evaluation of GIST descriptors for web-scale image search*, in "8th ACM International Conference on Image and Video Retrieval, CIVR'09", Santorin, Greece, July 2009.

[5] 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.

[6] 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/.

[7] 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.

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

[9] S. PETROVIC, B. DALBELO BASIC, A. MORIN, B. ZUPAN, J.-H. CHAUCHAT. *Textual features for corpus visualization using correspondence analysis*, in "Intelligent Data Analysis", 2009, vol. 13, n^{o} 5, p. 795–813.

[10] M. ROSSIGNOL, P. SÉBILLOT. *Combining Statistical Data Analysis Techniques to Extract Topical Keyword Classes from Corpora*, in "Intelligent Data Analysis", 2005, vol. 9, n^{o} 1, p. 105-127.

Publications of the year

Doctoral Dissertations and Habilitation Theses

[11] C. GUINAUDEAU. *Structuration automatique de flux télévisuels*, INSA de Rennes, December 2011, http://hal.inria.fr/tel-00646522/en.

[12] R. TAVENARD. *Indexation de séquences de descripteurs*, Université Rennes 1, July 2011, http://hal.inria.fr/tel-00639225/en.

Articles in International Peer-Reviewed Journal

[13] F. CÉROU, P. DEL MORAL, T. FURON, A. GUYADER. *Sequential Monte Carlo for rare event estimation*, in "Statistics and Computing", April 2011, p. 1-14 [*DOI :* 10.1007/s11222-011-9231-6], http://hal.inria.fr/inria-00584352/en.

[14] G. GRAVIER, C. GUINAUDEAU, G. LECORVÉ, P. SÉBILLOT. *Exploiting Speech for Automatic TV Delinearization: From Streams to Cross-Media Semantic Navigation*, in "EURASIP Journal on Image and Video

Processing", January 2011, vol. 2011, 689780, 17 [*DOI :* 10.1155/2011/689780], http://hal.inria.fr/hal-00645216/en.

[15] 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/en.

[16] Z. A. A. IBRAHIM, P. GROS. *TV Stream Structuring*, in "ISRN Signal Processing", April 2011, vol. 2011 [*DOI :* 10.5402/2011/975145], http://hal.inria.fr/inria-00601845/en.

[17] V. JOUHET, G. DEFOSSEZ, A. BURGUN, P. LE BEUX, P. LEVILLAIN, P. INGRAND, V. CLAVEAU. *Automated Classification of Free-text Pathology Reports for Registration of Incident Cases of Cancer*, in "Methods of Information in Medecine", 2011, vol. 50, http://hal.inria.fr/hal-00643819/en.

[18] 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.

[19] 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", 2011, http://hal.inria.fr/inria-00633013/en.

[20] J. ZEPEDA, C. GUILLEMOT, E. KIJAK. *Image Compression Using Sparse Representations and the Iteration-Tuned and Aligned Dictionary*, in "IEEE Journal of Selected Topics in Signal Processing", September 2011, vol. 5, n^{o} 5, p. 1061 -1073, http://hal.inria.fr/hal-00647264/en.

International Conferences with Proceedings

[21] M. BEN, G. GRAVIER. *Unsupervised mining of audiovisually consistent segments in videos with application to structure analysis*, in "IEEE Intl. Conf. on Multimedia and Exhibition", Spain, 2011, http://hal.inria.fr/hal-00646603/en.

[22] A. CHARPENTIER, C. FONTAINE, T. FURON, I. COX. *An Asymmetric Fingerprinting Scheme based on Tardos Codes*, in "Information Hiding", Prague, Czech Republic, T. PEVNY, T. FILLER (editors), Springer-Verlag, 2011, http://hal.inria.fr/inria-00581156/en.

[23] V. CLAVEAU, E. KIJAK. *Morphological Analysis of Biomedical Terminology with Analogy-Based Alignment*, in "Recent Advances in Natural Language Processing", Bulgaria, 2011, http://hal.inria.fr/hal-00644041/en.

[24] V. CLAVEAU, S. LEFÈVRE. *Topic Segmentation of TV-streams by mathematical morphology and vectorization*, in "InterSpeech", Italy, 2011, http://hal.inria.fr/hal-00643905/en.

[25] J. D. CRUZ GOMEZ, C. BOTHOREL, F. POULET. *Entropy based community detection in augmented social networks*, in "International Conference on Computational Aspects of Social Networks", Salamanca, Spain, 2011, p. 163-168, http://hal.inria.fr/hal-00640722/en.

[26] K. ELAGOUNI, C. GARCIA, P. SÉBILLOT. *A Comprehensive Neural-Based Approach for Text Recognition in Videos using Natural Language Processing*, in "ICMR", Trento, Italy, April 2011, 8, http://hal.inria.fr/hal-00645219/en.

[27] C. GUINAUDEAU, J. HIRSCHBERG. *Accounting for Prosodic Information to Improve ASR-Based Topic Tracking for TV Broadcast News*, in "Interspeech'11", Florence, Italy, August 2011, 4, http://hal.inria.fr/hal-00646626/en.

[28] M. JAIN, H. JÉGOU, P. GROS. *Asymmetric Hamming Embedding*, in "ACM Multimedia", Scottsdale, United States, October 2011, http://hal.inria.fr/inria-00607278/en.

[29] H. JÉGOU, R. TAVENARD, M. DOUZE, L. AMSALEG. *Searching in one billion vectors: re-rank with source coding*, in "International Conference on Acoustics, Speech and Signal Processing", Prague, Czech Republic, May 2011, http://hal.inria.fr/inria-00566883/en.

[30] J. LAWTO, J.-L. GAUVAIN, L. LAMEL, G. GREFENSTETTE, G. GRAVIER, J. DESPRES, C. GUINAUDEAU, P. SÉBILLOT. *A Scalable Video Search Engine Based on Audio Content Indexing and Topic Segmentation*, in "NEM Summit", Torino, Italy, September 2011, 6, http://hal.inria.fr/hal-00645228/en.

[31] G. LE-JAN, Y. BENEZETH, G. GRAVIER, F. BIMBOT. *A study on auditory feature spaces for speech-driven lip animation*, in "Interspeech", Florence, Italy, August 2011, http://hal.inria.fr/inria-00598314/en.

[32] G. LECORVÉ, G. GRAVIER, P. SÉBILLOT. *Automatically Finding Semantically Consistent N-grams to Add New Words in LVCSR Systems*, in "ICASSP", Prague, Czech Republic, May 2011, 4, http://hal.inria.fr/hal-00645223/en.

[33] S. LEFÈVRE, V. CLAVEAU. *Topic segmentation: application of mathematical morphology to textual data*, in "ISMM, Internation Symposium on Mathematical Morphology", Italy, 2011, http://hal.inria.fr/hal-00643913/en.

[34] H. LEJSEK, B. Þ. JÓNSSON, L. AMSALEG. *NV-Tree: nearest neighbors at the billion scale*, in "1st ACM International Conference on Multimedia Retrieval", Trento, Italy, 2011 [*DOI :* 10.1145/1991996.1992050], http://hal.inria.fr/hal-00644939/en.

[35] E. MARTIENNE, V. CLAVEAU, P. GROS. *Labeling TV stream segments with Conditional Random Fields*, in "MUSCLE International Workshop on Computational Intelligence for Multimedia Understanding", Italy, 2011, http://hal.inria.fr/hal-00644150/en.

[36] P. MEERWALD, T. FURON. *Group testing meets traitor tracing*, in "ICASSP", Prague, Czech Republic, IEEE, 2011 [*DOI :* 10.1109/ICASSP.2011.5947280], http://hal.inria.fr/inria-00580899/en.

[37] P. MEERWALD, T. FURON. *Iterative single Tardos decoder with controlled probability of false positive*, in "International Conference on Multimedia and Expo", Barcelona, Spain, IEEE, 2011, http://hal.inria.fr/inria-00581159/en.

[38] P. MEERWALD, T. FURON. *Towards Joint Tardos Decoding: The 'Don Quixote' Algorithm*, in "Information Hiding", Prague, Czech Republic, T. PEVNY, T. FILLER (editors), Springer-Verlag, 2011, http://hal.inria.fr/inria-00581148/en.

[39] A. MUSCARIELLO, G. GRAVIER, F. BIMBOT. *An efficient method for the unsupervised discovery of signalling motifs in large audio streams*, in "International Workshop on Content-Based Multimedia Indexing", Madrid, Spain, June 2011, http://hal.inria.fr/inria-00572817/en.

[40] A. MUSCARIELLO, G. GRAVIER, F. BIMBOT. *Towards robust word discovery by self similarity matrix comparison*, in "IEEE International Conference on Acoustics, Speech and Signal Processing", Pragua, Czech Republic, May 2011, http://hal.inria.fr/inria-00563418/en.

[41] A. MUSCARIELLO, G. GRAVIER, F. BIMBOT. *Zero-resource audio-only spoken term detection based on a combination of template matching techniques*, in "INTERSPEECH 2011: 12th Annual Conference of the International Speech Communication Association", Firenze, Italy, 2011, http://hal.inria.fr/inria-00597907/en.

[42] C. PENET, C.-H. DEMARTY, G. GRAVIER, P. GROS. *Technicolor and INRIA/IRISA at MediaEval 2011: learning temporal modality integration with Bayesian Networks*, in "MediaEval 2011, Multimedia Benchmark Workshop", Pisa, Italy, http://ceur-ws.org, November 2011, vol. 807, http://hal.inria.fr/hal-00643645/en.

[43] C. RAYMOND, V. CLAVEAU. *Participation de l'IRISA à DEFT 2011: expériences avec des approches d'apprentissage supervisé et non-supervisé*, in "Challenge DeFT (défi fouille de texte)", France, 2011, http://hal.inria.fr/hal-00643724/en.

[44] R. TAVENARD, H. JÉGOU, L. AMSALEG. *Balancing clusters to reduce response time variability in large scale image search*, in "International Workshop on Content-Based Multimedia Indexing (CBMI 2011)", Madrid, Spain, June 2011, http://hal.inria.fr/inria-00576886/en.

[45] G. TÓMASSON, H. SIGURÞÓRSSON, B. Þ. JÓNSSON, L. AMSALEG. *PhotoCube: effective and efficient multi-dimensional browsing of personal photo collections*, in "1st ACM International Conference on Multimedia Retrieval", trento, Italy, 2011 [*DOI :* 10.1145/1991996.1992066], http://hal.inria.fr/hal-00644942/en.

[46] L. UGHETTO, V. CLAVEAU. *Implication in Information Retrieval Systems*, in "European Society for Fuzzy Logic and Technology (EUSFLAT'2011)", Aix-Les-Bains, France, 2011, p. 431–438, http://hal.inria.fr/hal-00647594/en.

[47] P. WEINZAEPFEL, H. JÉGOU, P. PÉREZ. *Reconstructing an image from its local descriptors*, in "Computer Vision and Pattern Recognition", Colorado Springs, United States, IEEE, June 2011, http://hal.inria.fr/inria-00566718/en.

[48] C. WENGERT, M. DOUZE, H. JÉGOU. *Bag-of-colors for improved image search*, in "ACM Multimedia", Scottsdale, United States, October 2011, http://hal.inria.fr/inria-00614523/en.

[49] J. ZEPEDA, C. GUILLEMOT, E. KIJAK. *Image compression using the Iteration-Tuned and Aligned Dictionary*, in "IEEE International Conference on Acoustics, Speech and Signal Processing", Czech Republic, May 2011, p. 793 - 796 [*DOI :* 10.1109/ICASSP.2011.5946523], http://hal.inria.fr/hal-00647253/en.

National Conferences with Proceeding

[50] Y. BENEZETH, G. GRAVIER, F. BIMBOT. *Étude comparative des différentes unités acoustiques pour la synchronisation labiale*, in "GRETSI, Groupe d'Etudes du Traitement du Signal et des Images", Bordeaux, France, 2011, http://hal.inria.fr/inria-00593756/en.

[51] V. CLAVEAU, S. LEFÈVRE. *Segmentation thématique : apport de la vectorisation*, in "Conférence en recherche d'information et applications", France, 2011, http://hal.inria.fr/hal-00643688/en.

[52] J. D. CRUZ GOMEZ, C. BOTHOREL, F. POULET. *Idenfication et visualisation des partitions de réseaux sociaux à l'aide de points de vue sémantiques*, in "AVEC, Atelier Visualisation et Extraction de Connaissances", Brest, France, January 2011, p. 25-36.

[53] J. D. CRUZ GOMEZ, C. BOTHOREL, F. POULET. *Point of View Based Clustering of Socio-Semantic Network*, in "11èmes journées d'extraction et de gestion des connaissances, EGC'11", Brest, France, Revue des nouvelles technologies de l'information, January 2011, vol. RNTI-E, p. 309-310.

[54] J. D. CRUZ GOMEZ, C. BOTHOREL, F. POULET. *Semantic clustering of social networks using points of view*, in "CORIA: conférence en recherche d'information et applications 2011", Avignon, France, 2011, http://hal.inria.fr/hal-00609291/en.

[55] T.-N. DOAN, F. POULET. *Un environnement efficace pour la classification d'images à grande échelle*, in "12es journées d'extraction et de gestion des connaissances, EGC'12", Bordeaux, France, Revue des nouvelles technologies de l'information, January 2012, vol. RNTI-E.

[56] A.-R. EBADAT, V. CLAVEAU, P. SÉBILLOT. *Using shallow linguistic features for relation extraction in biomedical texts*, in "Traitement Automatique des Langues Naturelles, TALN", France, 2011, http://hal.inria.fr/hal-00644070/en.

[57] K. ELAGOUNI, C. GARCIA, P. SÉBILLOT. *Reconnaissance automatique de texte dans des vidéos à l'aide d'un OCR et de connaissances linguistiques*, in "GRETSI", Bordeaux, France, September 2011, 4, http://hal.inria.fr/hal-00645217/en.

[58] C. PENET, C.-H. DEMARTY, G. GRAVIER, P. GROS. *De la détection d'évènements sonores violents par SVM dans les films*, in "ORASIS - Congrès des jeunes chercheurs en vision par ordinateur", Praz-sur-Arly, France, INRIA Grenoble Rhône-Alpes, 2011, http://hal.inria.fr/inria-00595480/en.

[59] N.-K. PHAM, F. POULET, A. MORIN, P. GROS. *Analyse factorielle des correspondances hiérarchique pour la fouille d'images*, in "11es journées d'extraction et de gestion des connaissances, EGC'11", Brest, France, Revue des nouvelles technologies de l'information, January 2011, vol. RNTI-E, p. 175-182.

[60] L. UGHETTO, V. CLAVEAU, R. HARASTANI. *Différentes interprétations d'un modèle de RI à base d'inclusion graduelle*, in "Conférence en recherche d'information et applications", France, 2011, http://hal.inria.fr/hal-00643675/en.

Conferences without Proceedings

[61] M. AYARI, J. DELHUMEAU, M. DOUZE, H. JÉGOU, D. POTAPOV, J. REVAUD, C. SCHMID, J. YUAN. *INRIA@TRECVID'2011: Copy Detection & Multimedia Event Detection*, in "TRECVID 2011 Workshop", USA, 2011, http://hal.inria.fr/hal-00648016/en.

[62] C.-H. DEMARTY, C. PENET, M. SOLEYMANI, G. GRAVIER. *The Mediaeval 2011 affect task: Violent scene detection in Hollywood movies*, in "MediaEval 2011 Workshop", Italy, 2011, http://hal.inria.fr/hal-00646606/en.

[63] B. SAFADI, N. DERBAS, A. HAMADI, F. THOLLARD, G. QUÉNOT, H. JÉGOU, T. GEHRIG, H. K. EKENEL, R. STIFELHAGEN. *Quaero at TRECVID 2011: Semantic Indexing and Multimedia Event Detection*, in "TRECVID 2011 Workshop", USA, 2011.

Books or Proceedings Editing

[64] F. POULET, B. LE GRAND (editors). *Actes du 9e atelier visualisation et extraction des connaissances - 11es journées d'extraction et de visualisation des connaissances, EGC'11*, January 2011.

Research Reports

[65] G. Þ. GUÐMUNDSSON, L. AMSALEG, B. Þ. JÓNSSON. *Impact of Storage Technology on Cluster-Based High-Dimensional Indexing*, INRIA, July 2011, n^{o} RR-7681, http://hal.inria.fr/inria-00610446/en.

[66] H. JÉGOU, M. DOUZE, C. SCHMID. *Exploiting descriptor distances for precise image search*, INRIA, June 2011, http://hal.inria.fr/inria-00602325/en.

[67] H. JÉGOU, T. FURON, J.-J. FUCHS. *Anti-sparse coding for approximate nearest neighbor search*, INRIA, October 2011, n^{o} RR-7771, submitted to ICASSP'2012, http://hal.inria.fr/inria-00633193/en.

[68] R. TAVENARD, L. AMSALEG. *Improving the Efficiency of Traditional DTW Accelerators*, INRIA, November 2011, http://hal.inria.fr/hal-00639215/en.

Other Publications

[69] J. D. CRUZ GOMEZ, C. BOTHOREL, F. POULET. *Point of view based clustering of socio-semantic networks*, January 2011, http://hal.inria.fr/hal-00609268/en.

References in notes

[70] 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.

[71] 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.

[72] A. CHARPENTIER. *Identification de copies de documents multimedia grâce aux codes de Tardos*, Université Rennes 1, 10 2011, http://tel.archives-ouvertes.fr/tel-00646028/en/.

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

[74] T.-T. DO, E. KIJAK, T. FURON, L. AMSALEG. *Deluding Image Recognition in SIFT-based CBIR Systems*, in "ACM Multimedia in Forensics, Security and Intelligence", Firenze, Italie, ACM, October 2010, http://hal.inria.fr/inria-00505845/en/.

[75] T.-T. DO, E. KIJAK, T. FURON, L. AMSALEG. *Understanding the Security and Robustness of SIFT*, in "ACM Multimedia", Firenze, Italie, October 2010, http://hal.inria.fr/inria-00505843/en/.

[76] 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.

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

[78] H. JÉGOU, M. DOUZE, C. SCHMID, P. PÉREZ. *Aggregating local descriptors into a compact image representation*, in "IEEE Conference on Computer Vision & Pattern Recognition", jun 2010, p. 3304–3311, http://lear.inrialpes.fr/pubs/2010/JDSP10.

[79] 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.

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

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

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

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

[84] 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.

[85] 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.