INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Project-Team Texmex

Efficient Exploitation of Multimedia Documents: Exploring, Indexing and Searching in Very Large Databases

Rennes - Bretagne-Atlantique

Theme : Vision, Perception and Multimedia Understanding

Table of contents

1. Team

Research Scientists

Patrick Gros [Team Leader, Senior Research Scientist, INRIA, HdR]
Laurent Amsaleg [Research Scientist, CNRS]
Vincent Claveau [Research Scientist, CNRS]
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]
Pierre Tirilly [Assistant Professor, Univ. Rennes 1, until August 31st]

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]
Florent Dutrech [INRIA Technical Staff, until August 31th]
Sébastien Campion [INRIA Research Engineer]
Stacy Payne [INRIA Technical Staff, also with SAF]

PhD Students

Thanh Toan Do [MESR grant]
Thanh Nghi Doan [Vietnam government grant and Brittany Council grant, since October 1st]
Ali Reza Ebadat [Quaero project]
Khaoula Elagouni [CIFRE grant with Orange]
Julien Fayolle [Quaero project and Brittany council grant]
Gylfi Gudmundsson [Quaero project, since March 25th]
Camille Guinaudeau [Quaero project and Brittany Council grant]
Gwénolé Lecorvé [MESR grant until September 30th, INRIA contract until December 31th]
Cédric Penet [CIFRE grant with Technicolor since September 15th]
Romain Tavenard [ENS Cachan grant]
Joaquin Zepeda [ICOS-HD project, until October 31th, also with TEMICS]

Administrative Assistant

Loïc Lesage [Secretary INRIA, partial position in the project-team]

2. Overall Objectives

2.1. Overall Objectives

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

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

Many applications exist in both professional and domestic areas. On the professional side, such applications include transforming a TV broadcast program into a web site, a DVD or a mobile phone service, switching from a traditional TV program to an interactive one, better exploiting TV and video archives, constructing new video services (video on demand, video edition...). 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 delinearize 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 of the Year

- We have participated to several image search engine evaluations this year. First, a joint participation with Exalead has obtained a bronze medal in the Multimedia Grand Challenge. Second, we have obtained excellent results in the copy detection task of the TRECVID copy detection task. Third, our image search demonstrator has received the best demonstration award at the RFIA conference.
- The start-up Powedia, which is a spin-off of our project-team, was officially created (March 2010).
- We have started studying the issue of security in large scale image indexing. Papers on this problem have been presented during the IEEE MMSP and the ACM Multimedia conference this year.

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 [69] or the Difference of Gaussian method proposed by David Lowe [70].

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 [70], 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 [75]. 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 [74]; 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 [63]. 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, [73]) 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 [72]. 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 [66] 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 [71] 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 [65]. 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 [64]. 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 [68], 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 [75] 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. kertrack

Participant: Sébastien Campion [correspondent].

Visual graphical interface for tracking visual targets based on particule filter tracking or based on mean-shift.

5.1.2. mozaic2d

Participants: Florent Dutrech, 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.

5.1.3. Samusa

Participant: Sébastien Campion [correspondent].

This software is jointly maintained with Guillaume Gravier (METISS project-team).

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

5.1.4. PimPy

Participant: Sébastien Campion [correspondent].

The deposit of this software at APP is currently being processed. The software homepage is available here: http://pim.gforge.inria.fr/pimpy/.

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.

5.1.5. python-geohash

Participant: Sébastien Campion [correspondent].

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

Implementation of the Geometric Hashing algorithm of [77] to check if geometrical consistency between pairs of images.

5.1.6. 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.

5.1.7. 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

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.

5.1.8. Pqcodes

Participant: Hervé Jégou [correspondent].

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

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

Pqcodes is a library which implements the approximate k nearest neighbor search method of [21]. This software is used, in particular, in our image search demonstrator.

5.1.9. 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.*, bitrate of 2,3 days/second) TVSearch offer a web services API using the HTTP/REST protocol.

5.1.10. 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.

5.1.11. Previous 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-consistant parts.

IRISAPHON, produce phonetic words.

5.2. Demonstrations

5.2.1. Automatic Generation of Hypervideos

Participants: Sébastien Campion [correspondent], Mathieu Ben, Camille Guinaudeau, Gwénolé Lecorvé.

This work was made with the help of Guillaume Gravier, from the METISS *project-team.*

We created a demonstrator to illustrate an application of video topic segmentation on a collection of TV news programs (INA corpus) in collaboration with Guillaume Gravier from the METISS project-team. The core of the system is our topic segmenter, which is fed by the output of an automatic speech transcriptor, and the output of the S2E module dedicated to automatic extraction of structuring events in the video. Behind the topic segmenter, the Kiwi module extracts a list of keywords from each topic segment. Using these keywords we then create links to web pages dealing with the same topic, and links to related video segments inside the collection. All the generated metadata for a given video are used to generate a web page, that we call a hypervideo, and which allows non-linear browsing of the video, according to topic segments. Furthermore, the user can jump to web pages related to his/her topic of interest, or to other reports in the video collection dealing with the same or similar topics. Each time, the user can play the corresponding video segment in a player fully integrated in the browser. To this aim, we used the last version (3.6) of the Firefox browser, which handles video HTML mark-ups.

This demo was presented at the NEM summit 2009, St-Malo, France.

5.2.2. Image Search Engines Comparator

Participants: Sébastien Campion [correspondent], Laurent Amsaleg.

This is joint work with Nguyen Khang Pham, a former PhD student of the team.

Using ISEC (Image Search Engine Comparator), we publish a website which gives the possibility to use and compare several CBIR search engines. Currently we can use NVTree search engine on several datasets (up to 10 millions of images) and IRCA (Image Retrieval by Correspondence Analysis) search engine.

5.2.3. Image search demonstrator

Participants: Hervé Jégou, Sébastien Campion [correspondent].

*This is joint development with INRIA/*LEAR.

This image search demonstrator is based on our work of [21] and [39]. The memory used per indexed image is of 21 bytes only. It performs search by similarity in 10 millions images in about 20ms. Thanks to this high computational and memory efficiency, the demonstrator works on a laptop. We have also designed an improved graphical interface.

The former version of this demonstrator received a best demo award at the RFIA'2010 conference.

5.3. Experimental Platform

Participants: Laurent Amsaleg, Mathieu Ben, 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 taks, 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.

5.4. Datasets

We have released a new public dataset, called BIGANN, of one billion 128-dimensional vectors and proposed an experimental setup to evaluate high dimensional indexing algorithms on a realistic scale. The ground-truth is pre-calculated and provided. The BIGANN dataset is available online: http://corpus-texmex.irisa.fr.

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

Participants: Ewa Kijak, Joaquin Zepeda.

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

The objective of the study initiated in 2007 is to design scalable signal representation and approximation methods amenable to both compression (that is with sparseness properties) and description. In this work, we investigate sparse representations methods for local image description. The sparsity of the signal representation indeed depends on how well the bases match with the local signal characteristics.

In 2010, we have developed three methods for learning dictionaries to be used for sparse signal representations. These design methods extend traditional overcomplete dictionaries to increase overcompleteness by better taking into account the iterative nature of the matching pursuit algorithm: in all our design methods, the dictionary is adapted at each iteration (selection of an atom).

The proposed schemes 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 denoising applications. In particular, the last method, called ITAD (Iteration-Tuned and Aligned Dictionaries), has been used to produce a new image codec that outperforms JPEG2000 for a fixed image class.

The corresponding paper [49] has received the **second best paper award** at the MMSP workshop.

6.1.1.2. NLP techniques for Image Description

Participants: Vincent Claveau, Patrick Gros, Pierre Tirilly.

Natural Language Processing (NLP) and text retrieval techniques can help to describe and retrieve images at two stages:

- low-level image description: if we rely on an image description that shares some properties with the usual text description, such as the visual word scheme proposed by Sivic and Zisserman [75], we can use NLP and text retrieval techniques to improve image retrieval;
- high level image description: NLP and text retrieval techniques can be used to mine textual information coming with images, such as the news articles that images illustrate, and extract textual information to describe the images.

Following the work initiated in 2009, we worked on each of these two stages.

First, we continued the work about the use of weighting schemes and different distances for visual word-based image retrieval [75] based on techniques used for textual information retrieval. We confirmed the results of our preliminary experiments by showing that the weighting scheme and the distance chiefly depends on the characteristics of the image dataset considered [46].

Then, we also carried on working on high level image description, using NLP techniques to extract textual image descriptors from the text accompanying images. The annotation schemes was evaluated for logos and faces in the framework of a large parallel text-image corpus of news articles and demonstrated the interest of such an approach [47].

The PhD defense of P. Tirilly, including all this work, has taken place in July 2010 [14].

6.1.1.3. Describing Sequences for Audio/Video Retrieval

Participants: Laurent Amsaleg, Romain Tavenard.

Our work on this topic is done in close collaboration with Guillaume Gravier from the METISS *project-team.*

Today, very large databases of still images can be efficiently indexed and queried. Several temporal description techniques also exist for audio and video, but the state of the art approaches taking into account the concept of sequences can only do it on a limited scale. We have started investigating this issue in 2007. The fundamental question we have to answer is: when do we need to use fine metrics that takes temporality into account to compare sequences and when can we avoid this? For a large set of tasks ranging from TV stream structuring to audio word spotting, rather simple metrics could be used that operate at a very local scale, ignoring the whole sequence structure. Yet, for a few applications, deciding whether two sequences of descriptors are similar requires costly methods. We have tried two very different approaches where elements to compare were either the descriptors themselves, or a new feature based on the whole sequence of descriptors.

Directly comparing sequences of descriptors is done using the traditional Dynamic Time Warping approach. Here, the similarity of sequences is directly related to the similarity of the descriptions. As computing optimal alignment is computationally costly, we investigated ways to approximate the alignment using few computations. These initial results suggest pushing forward the investigations. We will look on ways to insert these techniques into large-scale indexing schemes.

We also compared sequence models, where each sequence is modeled using Support Vector Machines. Each model is somehow a translation of the temporal behavior of its corresponding sequence. Overall, we have shown that relying on models (instead of relying on descriptors) provides a better robustness to severe modifications of sequences, like temporal distortions for example. These results were obtained using a sequence collection made of real audio data broadcast on radio. We used cross-similarity estimation based metrics to compare models, as direct comparison between models is impossible.

6.1.1.4. GPU-based local descriptor extraction

Participant: Laurent Amsaleg.

Our work on this topic is done in close collaboration with researchers from Reykjavík University.

Video analysis using local descriptors requires a high-throughput descriptor creation process. This speed can be obtained from modern GPUs. We have adapted the computation of the Eff2 descriptors, a SIFT variant, to the GPU. We have compared our GPU-Eff descriptors to SiftGPU and shown that while both variants yield similar results, the GPU-Eff descriptors require significantly less processing time.

6.1.1.5. Aggregating local descriptors into a compact image representation

Participant: Hervé Jégou.

This is joint work with Matthijs and Cordelia Schmid, from the LEAR *project-team, and Patrick Pérez, from Technicolor.*

To make an image index at web scale, a server has to handle 10 million to 1 billion images. At this scale, it is no longer possible to use a conventional approach based on local descriptors: the memory usage of the image representation is prohibitive (several kilo-bytes). More importantly, the amount of memory scanned to do a single search increases, slowing down the search below the acceptable for an interactive search.

Therefore, we have investigated a new method [39] to optimize the trade-off between search accuracy, efficiency, but also the memory usage, which is a critical parameter in practical systems. To do so, we have proposed to revisit the different steps involved in image indexing, namely 1) the aggregation step, which produces a single vector representation from a set of local descriptors, 2) dimensionality reduction and 3) multi-dimensional indexing, where we have used a recent method based on a source coding paradigm [21].

Overall, our approach is able to index an image using a few dozen bytes only. Our experiments exhibits search quality comparable to the reference bag-of-features approach and significantly better efficiency: querying an image database of 10 million images takes 20 milliseconds on a single processor core.

6.1.2. Advanced Data Analysis Techniques

6.1.2.1. Use of Factorial Analysis for Text and Textual Streams Mining

Participant: Annie Morin.

This is joint work with Monica Becue et Belchin Kostov, from Polytechnic University of Catalunya

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 are also using Canonical Correspondence analysis, a method frequently used in Ecology where they have several groups of variables (discrete and/or continuous) describing statistical units. In our case, these units are the documents. We first try to find the trend and the seasonal components in the documents and we then detect the exceptional events. We focus on the visualization of the results.

6.1.2.2. Browsing Personal Image Collections

Participant: Laurent Amsaleg.

Our work on this topic is done in close collaboration with researchers from Reykjavík University.

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 propose a multi-dimensional model for media browsing, called ObjectCube, based on the multi-dimensional model commonly used in OLAP applications. ObjectCube has objects, tags, tag-sets and hierarchies as well as with various filtering operations, overall instantiating the OLAP concepts of *dimensions* and *facts* and *pivot*, *drill-down*, etc. primitives. A first proof-of-concept implementation of ObjectCube is running. We are currently adding various low-level image-processing techniques to, for example, automatically detect and classify the faces found in images.

6.1.2.3. Intensive Use of SVM for Text Mining and Image Mining

Participant: François Poulet.

This joint joint work with Nguyen Khang Pham, from Vietnamese College of Information & Technology

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: a boosting of least squares SVM to classify very large datasets on standard personal computers and incremental and parallel SVMs. The incremental part of the algorithm avoids us to load the whole dataset in main memory; we only need to have a small part of the dataset in main memory to build a part of the data model. Then we put together the partial models to get the full one with the same accuracy as usual algorithm; it solves the memory capacity problem of SVM algorithms.

To solve the computational time problem we have distributed the computation of the data blocks on different computers by the way of parallel and distributed 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. The GPU version of the algorithm is 130 times faster than the CPU one. The time needed for usual SVM algorithms like libSVM, SVMPerf or CB-SVM is divided by at least 2500 with one GPU or 5000 with two GPU cards.

We have extended the least squares SVM algorithm (LS-SVM). The first step was to adapt the algorithm to deal with datasets having a very large number of dimensions (like in text or image mining). Then we have applied boosting to LS-SVM for mining huge datasets having simultaneously a very large number of vectors and dimensions on standard computers. The performance of the new algorithm has been evaluated on large datasets from Machine Learning repository like Reuters-21578 or Forest Cover Type and image datasets. The accuracy is increased in almost all datasets compared to LibSVM.

We have used the same kind of principles (incremental and parallel) with another classification algorithm, an incremental and parallel k-means clustering algorithm has been developed to deal with very large vocabulary size in image categorization based on a bag of visual words [42]. We investigate other possible use of the same idea.

6.1.2.4. Large scale clustering

Participants: Laurent Amsaleg, Gylfi Gudmundsson.

Our work on this topic is done in close collaboration with researchers from Reykjavík University and from the University of Ioannina.

High-dimensional clustering is used by some content-based image retrieval systems to partition the data into groups; the groups (clusters) are then indexed to accelerate processing of queries. As clustering is central to many high-dimensional indexing strategies, we investigated several issues raised when clustering large collections of high-dimensional data. We basically tried to improve the performance of the clustering by either over-simplifying the algorithm or by relying on parallelism.

We extended a simplified version of the k-means algorithm and evaluated its behavior in an image-indexing context at a quite large scale. We proposed three extensions improving its performance and scalability, accelerating both query processing and the construction of clusters.

We also designed a high performance parallel implementation of a hierarchical data-clustering algorithm. The OpenMP programming model deals with the high irregularity of the algorithm and allows for efficient exploitation of the inherent loop-level nested parallelism. Thorough experimental evaluation demonstrates the performance scalability of our parallelization and the effective utilization of computational resources, which results in a clustering approach able to provide high quality clustering of very large datasets.

6.1.3. Security of Media

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

Over the years, the level of maturity reached by content-based retrieval systems (CBRSs) has significantly increased. We have now in research labs and also on the market various solutions that can process the contents of photos, of videos, of audio streams, etc. Of course, there are still many unsolved problems; yet, such systems are slowly entering our lives.

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. Overall, an abundant literature assesses that today's CBRSs are robust against general-purpose attacks, but almost no study address the security of content-based retrieval systems.

Challenging the security of CBRSs is a very targeted process. A security hacker typically attacks one system that uses a particular set of technology blocks, in order to delude one particular content-based task. It is the in-depth knowledge of the techniques used in one system that challenges security.

Because of our expertise in content-based systems, we are getting concerned by understanding the security side of CBRSs. 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. This very serious threat is a strong motivation for investigating in greater depth the many issues related to the security of content-based systems.

See [31], [32], [30], [61] for our work on this issue.

6.2. Multi-dimensional Indexing and clustering

6.2.1. Approximate nearest neighbor search using sparse coding techniques

Participants: Ewa Kijak, Joaquin Zepeda.

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

We introduced a new method [50] to search for approximate nearest neighbors under the normalized inner product similarity, using sparse image representations. The approach relies on the construction of new sparse image vectors designed to approximate the normalized inner product between underlying signal vectors. The resulting ANN search algorithm shows significant improvement compared to querying with the original sparse query vectors used in the literature for content-based image search.

6.2.2. Reducing the search time variability in nearest neighbor search

Participants: Laurent Amsaleg, Hervé Jégou, Romain Tavenard.

Many algorithms for approximate nearest neighbor search in high-dimensional spaces partition the data into clusters. At query time, for efficiency, an index selects the few (or a single) clusters nearest to the query point. Clusters are often produced by the well-known k-means approach since it has several desirable properties. On the downside, it tends to produce clusters having quite different cardinalities. Imbalanced clusters negatively impact both the variance and the expectation of query response times. This work proposes to modify k-means centroids to produce clusters with more comparable sizes without sacrificing the desirable properties. Experiments with a large-scale collection of image descriptors show that our algorithm significantly reduces the variance of response times without severely impacting the search quality.

6.2.3. Source coding techniques for nearest neighbor search

Participants: Laurent Amsaleg, Hervé Jégou, Romain Tavenard.

*This work was done in cooperation with Matthijs Douze and Cordelia Schmid (INRIA/*LEAR*).*

We have developed indexing techniques inspired by source coding [44], [21]. They can successfully index billions of high-dimensional vectors in memory by usage semi-structured quantization, which allows the computation of the distances in the compressed domain, without explicitly decoding the indexing codes.

Furthermore, we propose an approach that re-ranks the neighbor hypotheses obtained by these compressed-domain indexing methods. In contrast to the usual post-verification scheme, which performs exact distance calculation on the short-list of hypotheses, the estimated distances are refined based on short quantization codes, to avoid reading the full vectors from disk.

6.2.4. Video indexing structure

Participant: Hervé Jégou.

This is joint work with Matthijs Douze, Cordelia Schmid (INRIA/LEAR) and Patrick Pérez (Technicolor).

This work proposes a way to index videos with a very compact yet discriminative indexing algorithm, which allows example-based search in a large number of frames corresponding to thousands of hours of video. The description extracts one descriptor per indexed video frame by aggregating a set of local descriptors. These frame descriptors are encoded using a time-aware hierarchical indexing structure. A modified temporal Hough voting scheme is used to rank the retrieved database videos and estimate segments in them that match the query. Using temporal description of the videos, matched video segments are localized with an excellent precision.

Experimental results on the TRECVID 2008 copy detection task and a set of 38,000 videos from YouTube show that our method offers an excellent trade-off between search accuracy, efficiency and memory usage.

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.

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 2010, we have built an effective terminology extraction system based on an existing tool called TermoStat [67]. More specifically, we have developed new pre-processing schemes to handle noisy data. This whole system was tested in the framework of a Quaero evaluation campaign and ranked first.

This year, we also have developed a completely new approach to structure biomedical terminologies [28]. This approach relies on the decomposition of terms into morphemes and the translation of these morphemes into Japanese (kanji) subwords. The kanji characters thus offer a semantic way to access the semantics of the morpheme and allow us to detect semantic relations between them. This whole approach relies on a new forward-backward alignment technique improved by using analogies at the subword level.

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 2010, we developed several approaches based on machine learning techniques (SVM, Random forests, Naive Bayes) and a simple bag-of-words representation for the relations. These techniques were tested in the framework of a Quaero evaluation campaign on gene interaction detection; 4 runs were submitted and ranked in the 4 first places.

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... All these sources offer a way to exploit complementary information that can be used to semantically annotate multimedia documents. During this year, we developed a football multimedia corpus. It contains the video of several matches, the speech transcript, associated textual data from specialized websites... The manual annotation of the events, named entities and other relevant information of this corpus is under progress.

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 2010, we developed a complete video Optical Character Recognition (OCR) system, specifically adapted to detect and recognize embedded texts in video. Based on a neural approach, this method outperforms related work, especially in terms of robustness to style and size variability, to background complexity and to low image resolution. Moreover to reduce segmentation errors, a language model is introduced, that drives several steps of the video OCR in order to remove ambiguities associated with a local letter-by-letter recognition. The approach has been evaluated on a database of French TV news videos and achieves a character recognition rate of 95%. This work has been submitted to ICMR 2011.

6.3.2. Oral and Textual Information Retrieval

6.3.2.1. Efficient information retrieval using Pivots

Participants: Laurent Amsaleg, Vincent Claveau, Romain Tavenard.

This year, we initiated a new work about efficient information retrieval (IR). We developed a new embedding technique allowing a complexity reduction of the matching step between a query and the collection of documents. It relies on the building of pivot document which are used to build a vectorial representation for documents and queries. The comparison between a query and a document is thus based on a *second order* affinity (a document and a query are said similar if they are close or not from the same pivot documents). The experiments conducted in the framework on textual IR [51] shows the interest of this approach in terms of

complexity but also in term of performance. The second order allows us to retrieve documents even if they do not share any term of the query.

6.3.2.2. Information Retrieval in the TV context

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

This work is done in close collaboration with Guillaume Gravier from the METISS project-team.

The main focus of this research is to conceive new generation of IR systems capable of retrieving information from TV data. Directly indexing speech automatic transcripts remains nevertheless a difficult task. Transcriptions may contain many word recognition errors –in particular in the TV context where error rates can be high for some programs– that affect particularly very significant words such as named entities (*e.g.*, name of persons, places, organizations).

The main challenge of our work is therefore to investigate IR approaches robust to transcription errors. As an initial step, we are studying a new hybrid representation of transcripts whose aim is both to rely on the words that are correctly recognized and to ensure more flexibility for the portions of transcripts most likely to contain errors. To this end, we need: (i) to detect in transcripts the erroneous words. We have proposed a new word-level confidence measure that may efficiently ensure the reliability of transcribed words [35], focusing on words that are relevant for the IR task such as named entities [34], (ii) to define and locate the portions of transcripts containing errors (iii) and to propose an alternative (phonetic) representation of these erroneous areas. Exploitation of this representation in information retrieval requires to propose new index structures that are well suited for hybrid representation and to adapt the textual IR mechanisms to the TV context where the notion of document is not clearly defined.

6.3.2.3. Graded-Inclusion-Based Information Retrieval Systems

Participants: Vincent Claveau, Laurent Ughetto.

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

Databases (DB) querying mechanisms, and more particularly the division of relations was at the origin of the Boolean model for IR Systems. This model has rapidly shown its limitations and is no more used in IR. Among the reasons, the Boolean approach does 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 [48]. It has been shown that our fuzzy implication-based model can be shown as a logical model in IR, even if in the literature one writes $q \Rightarrow d$ and the other $d \Rightarrow q$. In the framework of a master internship, it has also been shown that our model can be seen as a language model in IR.

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, Patrick Gros, Emmanuelle Martienne, Sébastien Campion.

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 vs. commercial breaks). 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, our goal was to study 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.

6.4.2. Program Structuring

6.4.2.1. Audiovisual models for event detection in videos

Participants: Cédric Penet, Patrick Gros.

Our work on this topic is done in close collaboration with Guillaume Gravier from the METISS *project-team and Technicolor as external partner.*

We investigated the use of the audio modality for the detection of violent scenes in videos. A first approach based on SVM classification of short audio frames into four classes of sounds associated with violence (gunshot, screams, explosions and the rest) highlighted the difficulty of the task. This difficulty arises principally from the high variability of such sound classes between movies. This first approach however opens the door to further investigation for multimodal integration in the framework of violence detection in movies. In particular, we are currently focusing on the design of robust statistical approaches to deal with variability across movies.

6.4.2.2. Unsupervised mining of audiovisually consistent segments in videos

Participant: Mathieu Ben.

Our work on this topic is done in close collaboration with Guillaume Gravier from the METISS *project-team.*

Extraction of characteristic events in video programs is a crucial pre-processing step for video content-based analysis. However most current techniques rely on supervised approaches specifically dedicated to a given target event, for example detection of anchor person shots in TV news programs or specific actions in sports.

To overcome this genericity issue, we have developed a multimodal event mining technique to discover repeating video segments exhibiting audio and visual consistency in a totally unsupervised manner. The mining strategy first exploits independent audio and visual cluster analysis to provide segments which are consistent in both their visual modality and their audio modality, thus likely corresponding to a unique underlying event. A subsequent modeling stage using discriminative models enables accurate detection of the underlying event throughout the video. Event mining is applied to an unsupervised video-structuring task, using simple heuristics on occurrence patterns of the events discovered to select those relevant to the video's structure.

Results on TV programs ranging from news to talk shows and games, show that structurally relevant events are discovered with precisions ranging from 87 % to 98 % and recalls from 59 % to 94 %.

We will now focus on the exploitation of the results from this discovery module for higher level tasks like full structure matching of TV programs or topic segmentation where the discovered events could be used as anchor marks to guide the segmentation process.

6.4.3. Using Speech to Describe and Structure Video

Participants: Julien Fayolle, Camille Guinaudeau, Gwénolé Lecorvé, Christian Raymond, Pascale Sébillot.

Our work on this topic is done in close collaboration with Guillaume Gravier from the METISS *project-team.*

Speech can be used to structure and organize large collections of spoken documents (videos, audio streams...) 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 2010, we mostly focused on domain-robust transcription, named entity extraction and topic segmentation.

Automatically adapting ASR systems to various topics is a crucial issue in multimedia applications dealing with large collections of multi-topic documents. We worked on two aspects of the problem: language model adaptation and adding words to the vocabulary of the ASR system [12]. Firstly, we pursued our work on MDI adaptation of the language model using terminologies, exploiting constraints based on simple or complex terms. Best results are obtained with a few simple terms and diagnosis experiments have shown that most of the benefit of LM adaptation is lost during the transcription process [54]. Secondly, we proposed an original method to add out-of-vocabulary (OOV) words to the ASR system, combining syntactic and semantic aspects to define equivalences between the OOV word to add and in-vocabulary words.

Regarding information extraction from speech, we compared the robustness of several algorithms in [17]. Three of them were used, namely CRF, SVM and FSM, for named entity (NE) recognition in automatic transcripts [43]. All methods perform decently in spite of transcription specifics. CRFs perform the best on the single-best transcription while FSM allow us to process word-graph. Using different systems producing different errors opens the door to combination and to the use of the output of different NE systems as a feature to determine transcripts' quality.

Finally, transcripts are exploited for topic segmentation. We pursued our work on extending Utiyama and Isahara's probabilistic method [76] to account for confidence measures, semantic relations and, in collaboration with Columbia University, acoustic cues [37]. We proposed new lexical cohesion measures including all these information. Confidence measures and semantic relations were shown to be useful in different contexts. Though useless for topic segmentation, acoustic cues turned out interesting for keyword selection.

7. Contracts and Grants with Industry

7.1. Contracts with industry

7.1.1. Pôle de Compétitivité

Participant: Patrick Gros.

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). We are part, through our participation to the two projects Semim@ges and ICOS-HD, to the pole called "Images and networks" whose main actors are Technicolor and Orange Labs and which is located in Brittany and Pays de la Loire. Patrick Gros is also deputy member of the executive committee and the project selection committee.

7.2. Grants with industry

7.2.1. Contract with Technicolor

Participants: 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.

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.

7.3. 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, 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 2010, TEXMEX's participation in QUAERO roughly stabilized with respect to previous year. A Phd student (Gylfi Gudmundsson) joined the team in March while an engineer (Florent Dutrech) left in August. An independent annotator has been hired in February to work on annotation of our large-scale video corpus. She was joined by two interns during the summer to do this annotation work. Another intern did research work in the framework of Quaero during a few months before summer.

7.4. Start-up Creation

Participant: Patrick Gros.

The start-up Powedia, which is a spin-off of our project-team, was officially created (March 2010).

See http://www.powedia.com/.

8. Other Grants and Activities

8.1. Regional Initiatives

8.1.1. Support from Brittany General Council

Participant: Laurent Amsaleg.

Laurent Amsaleg received a grant from a joint effort between the Brittany General Council and CNRS to help setting up FP7 European projects. Laurent received 16,000 Euros, used in part to organize a two days workshop on the security issues of multimedia search engines with colleagues from Italy, Switzerland, Austria, England and Iceland. A proposal has subsequently been sent to the European community.

8.1.2. Support from University of Rennes I

Participant: Annie Morin.

Annie Morin received a grant from the University of Rennes to help setting up FP7 European projects. Annie received 16,700 Euros, used in part to invite Artur Silic from the University of Zagreb, Monica Becue and Belchin Kostov from the University Polytechnic de Catalunya. During their venue, we will start writing a proposal to be sent subsequently to EC and will finish the redaction of one scientific paper.

8.2. National Initiatives

8.2.1. ANR project ICOS-HD

Participants: Hervé Jégou, Ewa Kijak, Joaquin Zepeda.

Duration : 4 years, starting in January 2007. Partners: University of Bordeaux 1, CNRS-I3S.

This project concerns scalable indexing and compression for high definition video content management. Recent solutions for achieving high-quality compression of images/video result in scalable bit streams. The objective of the project is to propose new solutions of scalable description to facilitate editing, manipulation and access of HD contents via heterogeneous infrastructures. TEXMEX project-team is involved in studying new signal representations amenable to both compression and image description, as well as descriptor adaptation for image retrieval in large databases.

8.3. International Initiatives

8.3.1. Collaboration with Reykjavík University, Iceland

Participant: Laurent Amsaleg.

This collaboration is done in the context of the INRIA Associate Teams program. This program links two research teams (one INRIA, one foreign) willing to cross-leverage their respective excellence and their complementarity. Björn Þór Jónsson (Associate Professor) leads the team of researchers involved in Iceland.

This long-term collaboration, as old as the Texmex team itself, was done in the context of the INRIA Associate Teams program. The goal of this project was to research and develop new database support that integrates efficiency and effectiveness for modern, large-scale, computer-vision related applications and problems. This collaboration proved to be successful, with many papers accepted in journals and conferences. The creation of the Videntifier Technologies startup is another indicator of success. The Egide program also supported in part this collaboration. An European proposal has recently been submitted with both TEXMEX and Reykjavík University.

8.3.2. Collaboration with Croatia and Slovenia

Participant: Annie Morin.

Medical School, University of Zagreb, department of Electronics, Microelectronics, Computer and Intelligent systems, University of Zagreb, Zagreb, Croatia; Faculty of Computer and Information Science, University of Ljubljana, Slovenia; ERIC lab., University of Lyon2

We keep on the collaboration with the University of Zagreb, department of Electronics, Microelectronics, Computer and Intelligent systems.

The concerned research teams have different expertise on the same subject: machine learning for the Croatian team, statistics for the French team and common abilities such as development of open source data mining software and visualization tools. They have been in touch since a first meeting in 2004 on intelligent data mining. We have already implemented a new prototype for visualization of textual streams. Proposed collaboration includes sharing of a number of Ph.D. students.

8.4. Visits of foreign researchers, Invitations to foreign labs

8.4.1. Visits to and from Polytechnic University of Catalunya

Participant: Annie Morin.

Annie Morin was invited to visit the Polytechnic University of Catalunya. She gave a seminar, met several researchers and discussed on-going projects. Future work on exploratory text streams mining is foreseen.

Monica Becue spent a week in IRISA at the end of August to prepare a European project and to discuss about thesis in co-supervision.

8.4.2. Visit to the Spoken Language Processing Group at Columbia University

Participant: Camille Guinaudeau.

Spoken Language Processing Group - Department of Computer Science - Columbia University - New York, New York, USA

C. Guinaudeau spent three months, from July to September 2010, in the Spoken Language Processing Group at Columbia University to work on the use of acoustic information for TV stream structuring. Most methods developed for user browsing of a TV stream, to follow the evolution of a particular story, *e.g.*, are based on the transcripts of the speech contained in the stream. However, non-textual data is important as well, in particular the way the speech is pronounced in the program.

The objective of the visit was to collaborate with Julia Hirschberg, on the integration of acoustic information in a topic segmentation and a topic tracking systems developed for TV stream structuring.

8.4.3. Visit of members of the University of Reykjavík

Participant: Laurent Amsaleg.

Björn Por Jónsson and Grímur Tómasson spent one week within the team. They came to push the work initiated on Objectcube (personal photo browser) and to start investigating the security issues related to maliciously attacking the indexing and retrieval steps of multidimensional search engines.

9. Dissemination

9.1. Conference, Workshop and Seminar Organization

- F. Poulet and B. Le Grand organized and edited the proceedings of the 8th Workshop Visualisation et Extraction de Connaissances co-located with Extraction et Gestion de Connaissances, (EGC'10), Hammamet, Tunisia, Jan. 2010.

9.2. Involvement with the Scientific Community

- L. Amsaleg:
  - was a program committee member of BDA 2010, Toulouse, France;
  - was a program committee member of CIVR 2010, Xi'an, China;
  - was a program committee member of CORIA 2010, Sousse, Tunisia;
  - was a program committee member of EDBT 2010, Lausanne, Switzerland;
  - was a program committee member of LIVA 2010, Tsukuba, Japan;
  - was a program committee member of VLDB 2010, Singapore;
  - was in the reading committee of the EURASIP Journal on Advances in Signal Processing;
  - was the co-organizer of a GRD Isis special day "Passage à l'échelle de la recherche et de la fouille de contenus multimédia".
- V. Claveau:
  - was a reviewing committee member of TALN'10 (17^{e} conférence nationale Traitement automatique des langues naturelles), Montreal, Canada, July 2010;
  - was a program committee member of RECITAL'10, Montreal, Canada, July 2010;
  - was a program committee member of RFIA'10, 17ème conférence en Reconnaissance des Formes et Intelligence Artificielle, Caen, France, January 2010;
  - was a program committee member of Conférence en Recherche d'Information et Applications, CORIA 2010, Sousse, Tunisia, March 2010;
  - was a reviewing committee member for the journal TAL, Traitement Automatique des Langues ;
  - was a reviewing committee member for the journal Documents numériques.

- E. Kijak:
  - was an evaluator for the French ANR, 2010.
- P. Gros:
  - was a program committee member of the eight International Workshop on Content Based Multimedia Indexing (CBMI) Which was held in Grenoble, France in June 2010;
  - is a member of the steering board of the Content Based Multimedia Indexing (CBMI) workshop series;
  - was a program committee member of RFIA'10, 17ème conférence en Reconnaissance des Formes et Intelligence Artificielle, Caen, France, January 2010;
  - was a program committee member of the Second International Conference on Creative Content Technologies CONTENT, Lisbon, Portugal, November 2010;
  - was an associate editor for the special issue of EURASIP Journal on Image and Video Processing on video Analysis for Novel TV Services.
- H. Jégou:
  - was a program committee member of CVPR'2010, San Francisco, USA, June 2010 ;
  - was a program committee member of ECCV'2010, Heraklion, Greece, September 2010 ;
  - was a program committee member of CORESA'2010, Lyon, France, October 2010 ;
  - was a program technical program committee of MMSP'2010, Saint-Malo, France, October 2010.
- A. Morin:
  - was a program committee member of ITI 2010 (Information technology interfaces);
  - is vice-president of the CNU (National Council of the University) in the computer science section.
- F. Poulet:
  - was a program committee member of VINCI'10, Visual INformation Communications International, Beijing, China, September 2010;
  - was a program committee member of EGC'10, Extraction et Gestion de Connaissances, Hammamet, Tunisia, January 2010;
  - was co-organizer of the 8th workshop Visualisation et Extraction de Connaissances, (AVEC-EGC'10), Hammamet, Tunisia, January 2010;
  - was a reviewing committee member I3, Information-Interaction-Intelligence.
- C. Raymond:
  - is a member of the editorial board of the e-journal "Discours", http://discours.revues.org.

P. Sébillot:

- was a member of the editorial committee of RFIA 2010 (17e congrès francophone AFRIF-AFIA Reconnaissance des Formes et Intelligence Artificielle), Caen, France, January 2010;
- was a member of the program committee of CORIA 2010 (7e conférence en recherche d'information et applications), Sousse, Tunisia, March 2010;
- was a member of the program committee of LREC 2010 (7th international conference on Language Resources and Evaluation), Valletta, Malta, May 2010;

  - was a member of the program committee of JADT 2010 (10th international conference on the Statistical Analysis of Textual Data), Rome, Italie, June 2010;
  - was a member of the program committee of TALN 2010 (17e conférence francophone Traitement automatique des langues naturelles), Montréal, Canada, July 2010;
  - was a member of the program committee of DEFT 2010 (6e défi fouille de textes), Montréal, Canada, July 2010;
  - is an editorial committee member of the Journal TAL (Traitement automatique des langues; since July 2009)
  - was a member of the reading committee of the special issue "Le texte : objet d'analyse et vecteur de connaissances" of the Journal Document Numérique, and of several issues of the Journal TAL (Traitement automatique des langues) in 2010.

- P. Tirilly:
  - was a program committee member of EGC 2010, Hammamet, Tunisie.

9.3. Teaching Activities

- L. Amsaleg, H. Jégou and F. Poulet: Managing Large Collections of Digital Data. Master by research in computer science (2^{nd} year), University of Rennes 1.
- L. Amsaleg: Advanced Databases, ENSAI.
- V. Claveau: Symbolic Sequential Data, Master by research in computer science (2^{nd} year), University of Rennes 1.
- P. 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.
- E. Kijak is head of the Image engineering track of the engineering cursus of University of Rennes 1
- E. Kijak: Analysis of audiovisual documents and flows for indexing, Master by research in computer science (2^{nd} year), University of Rennes 1.
- E. Kijak and C. Guinaudeau: Digital Documents Indexing and Retrieval, Professional Master in Computer Science, 2^{nd} year, IFSIC, University of Rennes 1.
- A. Morin : Data Mining, Institut de la Francophonie pour l'Informatique, Hanoi, Master
- A. Morin : Data Mining, University of Rennes 1, Miage 2, Master.
- A. Morin: Statistical process Control and Reliability, International Master in Electronics and Telecommunication, SEU, Nanjing, China, University of Rennes 1.
- F. Poulet is in charge of the Master in computer science (2^{nd} year), MITIC, Computer Science Methods and Information and Communication Technologies, ISTIC,University of Rennes 1.
- F. Poulet: Supervised Learning. Master by research in computer science (2^{nd} year), ISTIC, University of Rennes 1.
- F. Poulet: Introduction to Data Mining. Professionnal Master in Computer Science, 2^{nd} year, ISTIC, University of Rennes 1.
- F. Poulet: Mining Symbolic Data. Professionnal Master in Computer Science, 2^{nd} year, ISTIC, University of Rennes 1.
- F. Poulet: Data Warehouses. Professionnal Master in Computer Science, 2^{nd} year, ISTIC, University of Rennes 1.

- F. Poulet: Applications and Problem Solving. Professionnal Master in Computer Science, 2nd year, ISTIC, University of Rennes 1.
- F. Poulet: Learning Methods for Multimedia Data. Professionnal Master in Computer Science, 2nd year, ISTIC, University of Rennes 1.
- P. Sébillot is course co-director of the Research in Computer Science specialism of the Master's in Computer Science (2nd year), University of Rennes 1.
- P. Sébillot: Advanced Databases and Modern Information Systems, 5th year, Computer Science, INSA Rennes.

9.4. Invited talks

- L. Amsaleg. Talk at MiFoR 2010.
- H. Jégou. Talk at the ERMITES summer school, September 2010.
- H. Jégou. Talk at Xerox Research Center Europe, May 2010.

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° 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° 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° 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. P. 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° 5, p. 869–883.

[8] X. NATUREL, P. GROS. *Detecting Repeats for Video Structuring*, in "Multimedia Tools and Applications", May 2008, vol. 38, n° 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° 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° 1, p. 105-127.

Publications of the year

Doctoral Dissertations and Habilitation Theses

[11] S. BAGHDADI. *Extraction multimodale de métadonnées de séquences vidéo dans un cadre bayésien*, Université de Rennes 1, February 2010, http://tel.archives-ouvertes.fr/docs/00/51/27/06/PDF/vf_these.pdf.

[12] G. LECORVÉ. *Adaptation thématique non supervisée d'un sytème de reconnaissance automatique de la parole*, Institut National des Sciences Appliquées de Rennes, December 2010.

[13] G. MANSON. *Délinéarisation automatique de flux de télévision*, Université de Rennes 1, July 2010, http://tel.archives-ouvertes.fr/docs/00/52/33/61/PDF/these_Gael_Manson.pdf.

[14] P. TIRILLY. *Traitement automatique des langues pour l'indexation d'images*, Université de Rennes 1, July 2010, http://tel.archives-ouvertes.fr/docs/00/51/64/22/PDF/these.pdf.

Articles in International Peer-Reviewed Journal

[15] M. DOUZE, H. JÉGOU, C. SCHMID. *An image-based approach to video copy detection with spatio-temporal post-filtering*, in "IEEE Transactions on Multimedia", June 2010, vol. 12, n° 4, p. 257-266 [*DOI :* 10.1109/TMM.2010.2046265], http://ieeexplore.ieee.org/stamp/5437235.pdf.

[16] P. E. HADJIDOUKAS, L. AMSALEG. *Nested OpenMP Parallelization of a Hierarchical Data Clustering Algorithm*, in "Parallel Processing Letters", June 2010, vol. 20, n° 2, p. 187-208 [*DOI :* 10.1142/S0129626410000144], http://hal.inria.fr/inria-00514758/en/.

[17] S. HAHN, M. DINARELLI, C. RAYMOND, F. LEFÈVRE, P. LEHNEN, R. DE MORI, H. NEY, G. RICCARDI, A. MOSCHITTI. *Comparing Stochastic Approaches to Spoken Language Understanding in Multiple Languages*, in "IEEE Transactions on Audio, Speech and Language Processing", 2010, Accepted for publication.

[18] S. HAHN, M. DINARELLI, C. RAYMOND, F. LEFÈVRE, P. LEHNEN, R. DE MORI, H. NEY, G. RICCARDI. *Comparing Stochastic Approaches to Spoken Language Understanding in Multiple Languages*, in "IEEE Transactions on Audio, Speech and Language Processing", 2011.

[19] 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° 4, p. 663-684.

[20] H. JÉGOU, M. DOUZE, C. SCHMID. *Improving bag-of-features for large scale image search*, in "International Journal of Computer Vision", February 2010, vol. 87, n° 3, p. 316-336 [*DOI :* 10.1007/s11263-009-0285-2], http://www.springerlink.com/content/wh52x87315697752/fulltext.pdf.

[21] H. JÉGOU, M. DOUZE, C. SCHMID. *Product Quantization for Nearest Neighbor Search*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", January 2011 [*DOI :* 10.1109/TPAMI.2010.57], http://ieeexplore.ieee.org/stamp/5432202.pdf.

[22] H. JÉGOU, C. SCHMID, H. HARZALLAH, J. VERBEEK. *Accurate image search using the contextual dissimilarity measure*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", January 2010, vol. 32, n^{o} 1, p. 2-11 [*DOI :* 10.1109/TPAMI.2008.285], http://ieeexplore.ieee.org/stamp/4695831.pdf.

[23] G. MANSON, S.-A. BERRANI. *Automatic TV Broadcast Structuring*, in "International Journal of Digital Multimedia Broadcasting", 2010 [*DOI :* 10.1155/2010/153160], http://www.hindawi.com/journals/ijdmb/2010/153160.html.

[24] C. MORAND, J. BENOIS-PINEAU, J.-P. DOMENGER, J. ZEPEDA, E. KIJAK, C. GUILLEMOT. *Scalable Object-based Video Retrieval in HD Video DataBases*, in "Signal Processing: Image Communication", July 2010, vol. 25, n^{o} 6, p. 450-465 [*DOI :* 10.1016/J.IMAGE.2010.04.004], http://www.sciencedirect.com/science/article/B6V08-4YYRMMF-1/2/de16b6c1e4f3e19f3beb1e76be9bc1db.

[25] A. OLAFSSON, B. P. JÓNSSON, L. AMSALEG, H. LEJSEK. *Dynamic behavior of balanced NV-trees*, in "Multimedia Systems", 2010 [*DOI :* 10.1007/s00530-010-0199-4], http://www.springerlink.com/content/e303136755270314/fulltext.html.

[26] L. PAULEVÉ, H. JÉGOU, L. AMSALEG. *Locality sensitive hashing: A comparison of hash function types and querying mechanisms*, in "Pattern Recognition Letters", 2010, vol. 31, n^{o} 11, p. 1348-1358.

International Peer-Reviewed Conference/Proceedings

[27] F. BÉCHET, C. RAYMOND, F. DUVERT, R. DE MORI. *Frame Based Interpretation Of Conversational Speech*, in "Spoken Language Technologies Workshop", Berkeley, California, U.S.A, December 2010.

[28] V. CLAVEAU, E. KIJAK. *Analyse morphologique en terminologie biomédicale par alignement et apprentissage non-supervisé*, in "Conférence Traitement automatique des langues naturelles, TALN'10", Montréal, Québec, Canada, July 2010.

[29] K. DADASON, Á. P. JÓHANNSSON, H. LEJSEK, B. P. JÓNSSON, L. AMSALEG. *GPU Acceleration of Eff2 Descriptors using CUDA*, in "18th ACM International Conference on Multimedia", Florence, Italy, October 2010.

[30] T.-T. DO, E. KIJAK, T. FURON, L. AMSALEG. *Challenging the Security of Content-Based Image Retrieval Systems*, in "IEEE International Workshop on Multimedia Signal Processing, MMSP'10", Saint-Malo, France, October 2010.

[31] T.-T. DO, E. KIJAK, T. FURON, L. AMSALEG. *Deluding Image Recognition in SIFT-based CBIR Systems*, in "18th ACM International Conference on Multimedia - Workshop on Multimedia in Forensics, Security and Intelligence", Florence, Italy, October 2010.

[32] T.-T. DO, E. KIJAK, T. FURON, L. AMSALEG. *Understanding the Security and Robustness of SIFT*, in "18th ACM International Conference on Multimedia", Florence, Italy, October 2010.

[33] M. DOUZE, H. JÉGOU, C. SCHMID, P. PÉREZ. *Compact video description for copy detection with precise temporal alignment*, in "European Conference on Computer Vision, ECCV'10", Heraklion, Greece, September 2010.

[34] J. FAYOLLE, F. MOREAU, C. RAYMOND, G. GRAVIER. *Reshaping Automatic Speech Transcripts for Robust High-level Spoken Document Analysis*, in "4th Workshop on Analytics for Noisy Unstructured Text Data, AND'10", Toronto, Canada, October 2010, http://www.irisa.fr/texmex/publications/versionElect/2010/fayolle10b.pdf.

[35] J. FAYOLLE, F. MOREAU, C. RAYMOND, G. GRAVIER, P. GROS. *CRF-based Combination of Contextual Features to Improve A Posteriori Word-level Confidence Measures*, in "International Conference on Speech Communication and Technologies, Interspeech'10", Makuari, Japan, September 2010, http://www.irisa.fr/texmex/publications/versionElect/2010/fayolle10a.pdf.

[36] G. GUDMUNDSSON, B. P. JÓNSSON, L. AMSALEG. *A Large-Scale Performance Study of Cluster-Based High-Dimensional indexing*, in "18th ACM International Conference on Multimedia - Workshop on Very-Large-Scale Multimedia Corpus, Mining and Retrieval", Florence, Italy, October 2010.

[37] C. GUINAUDEAU, G. GRAVIER, P. SÉBILLOT. *Improving ASR-based topic segmentation of TV programs with confidence measures and semantic relations*, in "11th Annual Conference of the International Speech Communication Association, Interspeech'10", Makuhari, Japan, September 2010, p. 1365-1368.

[38] C. GUINAUDEAU, G. GRAVIER, P. SÉBILLOT. *Utilisation de relations sémantiques pour améliorer la segmentation thématique de documents télévisuels*, in "17e conférence sur le traitement automatique des langues naturelles, TALN'10", Montréal, Québec, Canada, July 2010, http://hal.inria.fr/docs/00/53/33/89/PDF/guinaudeau_taln2010.pdf.

[39] H. JÉGOU, M. DOUZE, C. SCHMID, P. PÉREZ. *Aggregating local descriptors into a compact image representation*, in "IEEE Conference on Computer Vision and Pattern Recognition, CVPR'10", San Fransisco, USA, June 2010.

[40] H. LEJSEK, H. PÓRMÓDSDÓTTIR, F. H. ASMUNDSSON, K. DADASON, Á. P. JÓHANNSSON, B. P. JÓNSSON, L. AMSALEG. *VidentifierTM Forensic: Large-Scale Video Identification in Practise*, in "18th ACM International Conference on Multimedia - Workshop on Multimedia in Forensics, Security and intelligence", Florence, Italy, October 2010.

[41] S. OZDOWSKA, V. CLAVEAU. *Inferring syntactic rules for word alignment through Inductive Logic Programming*, in "7th Language Resources and Evaluation Conference, LREC'10", Valetta, Malta, May 2010.

[42] F. POULET, N.-K. PHAM. *High Dimensional Image Categorization*, in "Advanced Data Mining and Applications, ADMA'10", Chongqin, Chine, L. CAO, Z. JIANG, F. YONG (editors), Lecture Notes in Computer Science, Springer-Verlag, 2010, vol. 6440, p. 465-476.

[43] C. RAYMOND, J. FAYOLLE. *Reconnaissance robuste d'entités nommées sur de la parole transcrite automatiquement*, in "17e conférence sur le traitement automatique des langues naturelles, TALN'10", Montréal, Québec, Canada, July 2010, http://www.irisa.fr/texmex/publications/versionElect/2010/raymond10a.pdf.

[44] H. SANDHAWALIA, H. JÉGOU. *Searching with expectations*, in "IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP'10", Dallas, USA, March 2010.

[45] C. SERVAN, N. CAMELIN, C. RAYMOND, F. BÉCHET, R. DE MORI. *On the Use of Machine Translation for Spoken Language Understanding Portability*, in "IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP'10", Dallas, Texas, USA, March 2010, p. 5330-5333 [*DOI :* 10.1109/ICASSP.2010.5494960], http://ieeexplore.ieee.org/iel5/5487364/5494886/05494960.pdf.

[46] P. TIRILLY, V. CLAVEAU, P. GROS. *Distances and weighting schemes for bag of visual words image retrieval*, in "ACM International Conference on Multimedia Information Retrieval, MIR'10", Philadelphia, Pennsylvania, USA, March 2010, p. 323-332 [*DOI :* 10.1145/1743384.1743438], http://portal.acm.org/ft_gateway.cfm?id=1743438&type=pdf&coll=GUIDE&dl=GUIDE&CFID=104652799&CFTOKEN=29699755.

[47] P. TIRILLY, V. CLAVEAU, P. GROS. *News image annotation on a large parallel text-image corpus*, in "7th Language Resources and Evaluation Conference, LREC'10", Valletta, Malta, May 2010.

[48] L. UGHETTO, G. PASI, V. CLAVEAU, O. PIVERT, P. BOSC. *Implication in Information Retrieval Systems*, in "9th International Conference on Adaptivity, Personalization and Fusion of Heterogeneous Information, RIAO'10", Paris, France, April 2010.

[49] J. ZEPEDA, C. GUILLEMOT, E. KIJAK. *The Iteration-Tuned Dictionary for Sparse Representations*, in "IEEE International Workshop on Multimedia Signal Processing, MMSP'10", Saint-Malo, France, October 2010.

[50] J. ZEPEDA, E. KIJAK, C. GUILLEMOT. *Approximate nearest neighbors using sparse representations*, in "IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP'10", Dallas, Texas, USA, March 2010.

National Peer-Reviewed Conference/Proceedings

[51] V. CLAVEAU, R. TAVENARD, L. AMSALEG. *Vectorisation des processus d'appariement document-requête*, in "7e conférence en recherche d'informations et applications, CORIA'10", Sousse, Tunisia, March 2010, p. 313-324, http://asso-aria.org/coria/2010/313.pdf.

[52] C. GUINAUDEAU, G. GRAVIER, P. SÉBILLOT. *Indices utiles à la cohésion lexicale pour la segmentation thématique de documents oraux*, in "28es journées d'étude sur la parole, JEP'10", Mons, Belgique, May 2010, http://hal.inria.fr/docs/00/53/33/88/PDF/guinaudeau_jep2010.pdf.

[53] H. JÉGOU, M. DOUZE, C. SCHMID. *Représentation compacte des sacs de mots pour l'indexation d'images*, in "Congrès francophone AFRIF-AFIA de reconnaissance des formes et d'intelligence artificielle, RFIA'10", January 2010.

[54] G. LECORVÉ, G. GRAVIER, P. SÉBILLOT. *L'adaptation thématique d'un modèle de langue fait-elle apparaître des mots thématiques ?*, in "28es journées d'étude sur la parole, JEP'10", Mons, Belgique, May 2010.

[55] N.-K. PHAM, A. MORIN, P. GROS, F. POULET. *Analyse des correspondances hiérarchiques pour la fouille d'images*, in "8e atelier visualisation et extraction de connaissances - 10es journées d'extraction et de gestion des connaissances, EGC'10", Hammamet, Tunisia, January 2010.

[56] N.-K. PHAM, F. POULET, A. MORIN, P. GROS. *Indexation et recherche d'images à très grande échelle avec une AFC incrémentale et parallèle sur GPU*, in "10es journées d'extraction et de gestion des connaissances, EGC'10", Hammamet, Tunisia, Revue des nouvelles technologies de l'information, January 2010, vol. RNTI-E.

[57] P. TIRILLY, V. CLAVEAU, P. GROS. *Détection de logos pour l'annotation d'images de presse*, in "Congrès francophone AFRIF-AFIA de reconnaissance de formes et d'intelligence artificielle, RFIA'10", Caen, France, 2010.

Workshops without Proceedings

[58] J. D. CRUZ GOMEZ, C. BOTHOREL, F. POULET. *Détection de communautés dans les réseaux socio-sémantiques par point de vue*, in "Journée fouille de grands graphes, JFGG'10", Toulouse, France, October 2010.

[59] H. JÉGOU, M. DOUZE, G. GRAVIER, C. SCHMID, P. GROS. *INRIA LEAR-TEXMEX: video copy detection task*, in "TRECVID Workshop", Gaithersburg, USA, November 2010.

Books or Proceedings Editing

[60] F. POULET, B. LE GRAND (editors). *Actes du 8e atelier visualisation et extraction des connaissances - 10es journées d'extraction et de visualisation des connaissances, EGC'10*, January 2010.

Research Reports

[61] T.-T. DO, E. KIJAK, T. FURON, L. AMSALEG. *Understanding the security and robustness of SIFT*, INRIA, May 2010, n^{o} 7280, http://hal.inria.fr/inria-00482502/en/.

[62] X. NATUREL, P. GROS. *Dealing with Television Archives: Television Structuring*, INRIA, Rennes, France, May 2010, n^{o} 7301, http://hal.archives-ouvertes.fr/docs/00/50/77/52/PDF/RR-7301.pdf.

References in notes

[63] 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.

[64] L. AMSALEG. *Indexation multidimensionnelle*, in "L'indexation multimédia. Description et recherche automatiques", P. GROS (editor), Hermes, 2007, p. 215-244.

[65] 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.

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

[67] P. DROUIN. *Term extraction using non-technical corpora as a point of leverage*, in "Terminology", 2003, vol. 9, n^{o} 1, p. 99-117.

[68] 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.

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

[70] 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.

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

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

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

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

[75] 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.

[76] M. UTIYAMA, H. ISAHARA. *A Statistical Model for Domain-Independent Text Segmentation*, in "Proceedings of the 39th Annual Meeting of Association for Computational Linguistics, ACL'01", Toulouse, France, July 2001, p. 491-498.

[77] 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.