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

TEXMEX *is a common project-team with CNRS, University of Rennes 1 and INSA. It has been created on January the 1st, 2002 and became an INRIA project on November the 1st, 2002.*

1. Team

Research Scientist

Patrick Gros [Team Leader, Senior Research Scientist, INRIA, HdR]
Laurent Amsaleg [Research Scientist, CNRS]
Vincent Claveau [Research Scientist, CNRS]

Faculty Member

Ewa Kijak [Associate Professor, Univ. Rennes 1]
Annie Morin [Associate Professor, Univ. Rennes 1, HdR]
François Poulet [Associate Professor, Univ. Rennes 1, HdR]
Pierre Tirilly [Assistant Professor, Univ. Rennes 1, from September 1st]
Pascale Sébillot [Professor, INSA Rennes, HdR]
Christian Raymond [Associate Professor, INSA Rennes]

External Collaborator

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]
Sébastien Campion [INRIA Research Engineer]
Stacy Payne [INRIA Technical Staff, also with SAF]

PhD Student

Siwar Baghdadi [CIFRE grant with Thomson until March 31st]
Thanh Toan Do [MESR grant, from October 1st]
Ali Reza Ebadat [Quaero project, from October 1st]
Khaoula Elagouni [CIFRE grant with Orange, from October 1st]
Julien Fayolle [Quaero project and Brittany council grant, from October 12th]
Camille Guinaudeau [Quaero project and Brittany Council grant]
Gwénolé Lecorvé [MESR grant]
Gaël Manson [CIFRE grant with Orange, until October 31st]
Nguyen Khang Pham [also with *Institut francophone pour l'informatique*, Hanoï, Vietnam, until November 6th]
Romain Tavenard [ENS Cachan grant]
Pierre Tirilly [CNRS - Brittany Council grant, until August 31st]
Joaquin Zepeda [ICOS-HD project, also with TEMICS]

Post-Doctoral Fellow

Zein Al-Abidin Ibrahim [Semim@ges project, until June 30th]

Visiting Scientist

Sébastien Lefèvre [Associate Professor, Univ. Strasbourg, 1-year visiting from September 1st]

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 PCs 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 several media in a joint manner 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 characteristic 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

- In Nguyen Khang Pham’s thesis, we tried to use a classical data analysis tool, the factorial correspondance analysis, with images. This approach was very successful: The method appeared to be more efficient than the techniques used like LSA and pLSA or the TF.IDF weighting scheme. Moreover, these algorithms were developed for GPU chips and are very fast. See section 6.1.2
- We have launched in 2008 a new collaboration with the PILGRIM team of IRISA whose goal is to revisit the principle of search engines based on the fuzzy set theory. In this framework, queries and document are fuzzy set sets of words which are paired by the use of gradual implications. This work provided first very promising results and was published in ECIR 2009 and Fuzz-IEEE’09, two major conferences in respectively Information retrieval and Fuzzy Logic, a proof that the work, although still preliminary, is of interest for the community. See section 6.2.2

3. Scientific Foundations

3.1. Background

This section presents some of the basic techniques used in the team. They are basic blocks used in many research projects as well as research topics by themselves, since we usually use them in new contexts.

3.2. Local 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 can be re-sized, cropped... The solution consists in computing descriptors from the images and to turn image comparison into descriptor comparison. This can be done if, on the one hand, the descriptors contain some information on the image content, while, on the other hand, they do not depend on the image format, size or on transformations the image can undergo.

The most classical method associates 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. A second method consists in extracting regions in the image and to associate a descriptor with each of these regions. Most of the time, this is done by extracting points (called interest points) with a Harris-like detector [54], and by considering a circular or elliptic region around each of these points.

The differential invariants were among the first local descriptors used. Established by Florack [51], their use for image comparison was proposed by Schmid [66]. Each descriptor is a combination of the first derivatives of the signal at the interest point. These descriptors appeared experimentally to be very robust to geometric and photometric transforms. An even more powerful descriptor was then proposed by Lowe: the SIFT descriptor [57]. It is composed of 16 local histograms of gradient directions around the interest point.

Local descriptors can be used in many applications: image comparison for object recognition, image copy detection, detection of repeats in television streams... 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 can generate a database of one billion of descriptors. That is why specific indexing techniques are required. Up to now, most of them are computed from decompressed images, while most formats images are stored compressed. Thus it would be interesting to directly compute the descriptors in the compressed domain. Finally, their evaluation for very large image collection (several millions of images) is still an open and interesting problem.

3.3. 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 [65]; 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 [43]. 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.

3.4. Stochastic Models for Multimodal Analysis

Describing multimedia documents, *i.e.*, documents that contain several modalities (e.g. text, images, sound) requires to take all these modalities into account since they can 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 to generate 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 (HMM, [62]) 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 HMM, 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 HMM 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 [61]. These models are an extension of HMM 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 [49] 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 [58] 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 HMM 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.

3.5. 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. Two classes of multi-dimensional indexing methods can be distinguished: exact nearest neighbor (NN) searches and approximate NN-search schemes.

Traditional multidimensional indexing techniques typically divide the data space into cells containing vectors [47]. 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. NN-algorithms typically use the geometrical properties of (minimum bounding) cells to eliminate cells that cannot have any impact on the result of the current query [48].

Many data-partitioning index methods derive from the seminal R-Tree [53], and their differences lie in the properties of the shapes used to build cells and/or in the degree of overlapping between cells. Well known space-partitioning techniques are somehow related to the K-D-B-Tree [63], and differ on the way space is split and cells encoded.

Unfortunately, the "curse of dimensionality" phenomenon makes these traditional approaches ineffective in high-dimensional spaces [46]. This phenomenon is particularly prevalent when performing *exact* NN-searches. There is therefore an increasing interest in performing *approximate* NN-searches, where result quality is traded for reduced query execution time. Many approaches to approximate NN-searches have been published; their description can be found in [46].

Some approaches simply rely on dimensionality reduction techniques, such as PCA, but their use remains problematic when facing very high-dimensional datasets. Other approaches abort the search process early, after having accessed an arbitrary and predetermined number of cells. While this is highly effective, it does not give any clue on the quality of the result returned to the user. Some other approaches consider an approximation of the sizes of cells instead of considering their exact sizes, making somehow cells "smaller". Shrunk cells increase efficiency of retrievals as they reduce overlap in space, but interesting vectors might be missed, however.

Recently, several approaches have transformed costly nearest neighbor searches in multidimensional space into efficient uni-dimensional accesses. One approach using locality sensitive hashing (LSH) techniques [52] uses several hash functions such that co-located vectors are likely to collide in buckets. Fagin *et al.* [50] proposed a framework based on projecting the descriptors onto a limited set of random lines, each line giving a ranking of the database descriptors with respect to the query descriptor.

3.6. 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 concept 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

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 camcording, 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

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 (IDDN.FR.001.0032.000.S.C.2005.000.20900), ANAMORPHO (IDDN.FR.001.050022.000.S.P.2008.000.20900), FAESTOS (IDDN.FR.001.470029.000.S.P.2006.000.40000), 2PAC (IDDN.FR.001.470028.000.S.P.2006.000.40000), DIVATEX (IDDN.FR.001.320006.000.S.P.2006.000.40000), NAVITEX (IDDN.FR.001.190034.000.S.P.2007.000.40000) and TELEMEX. Please refer to the previous activity reports for a description of these programs. In 2008, the following pieces of software were developed or updated.

5.1.1. Vidsig

Participants: Sébastien Campion [correspondent], Patrick Gros.

The application VIDSIG, developed in the framework of Xavier Naturel's thesis, enables to compute a video signature of various video formats (TV and Web) Based on Discrete Cosinus Transformation, it produces video signature small and robust (64 bits per image).

Registration at the "Agence pour la Protection des Programmes" (APP) is in progress.

5.1.2. Vidseg

Participants: Sébastien Campion [correspondent], Patrick Gros.

The VIDSEG software, developed in the framework of Manolis Delakis thesis, allows to detect video segment in various formats. Results are composed by significant segmentation features such as silences in audio track, changes of ratio aspect, monochrome images, cuts and dissolves.

Vidseg was registered at APP (IDDN.FR.001.250009.000.S.P.2009.000.40000)

5.1.3. Isec

Participant: Sébastien Campion [correspondent].

ISEC (*I*mage *S*earching *E*ngines *C*omparator) is a web application used as graphical interface for image searching engines based on retrieval by content. It was developed with the help of Pierre Cosquer during an internship.

Its architecture allows to plug different searching engines and enables to compare their results. It also permits to edit submitted images with several operations (cropping, blurring, etc.). These attacks make it possible to test the robustness of searching engines.

Registration at the "Agence pour la Protection des Programmes" (APP) is in progress.

5.1.4. GPU-KMeans

Participants: Nguyen Khang Pham, Sébastien Campion [correspondent], François Poulet.

GPU-KMEANS is an implementation of k-means algorithm on graphical process unit (graphic cards) using CuBLAS library. These hardware properties (data-parallel computing) allow GPU-KMeans to reduce compute operations by a factor 10.

Registration at the "Agence pour la Protection des Programmes" (APP) is in progress.

5.1.5. Correspondence Analysis

Participants: Nguyen Khang Pham, Annie Morin, Sébastien Campion [correspondent].

CORRESPONDENCE ANALYSIS software is designed to compute a factorial correspondence analysis (FCA) for image retrieval. FCA analyzes a contingency table crossing terms/words and documents. To adapt FCA on images, CORRESPONDENCE ANALYSIS uses "visual words" computed from scalable invariant feature transform (SIFT) descriptors in images and use them for image quantization.

Registration at the "Agence pour la Protection des Programmes" (APP) is in progress.

5.1.6. GPU Correspondence Analysis

Participants: Nguyen Khang Pham, Sébastien Campion [correspondent].

GPU CORRESPONDENCE ANALYSIS is an implementation of the previous software CORRESPONDENCE ANALYSIS on graphical processing unit (graphical card) which reduces execution time of compute operations by a factor 20.

Registration at the "Agence pour la Protection des Programmes" (APP) is in progress.

5.1.7. CAVIZ

Participants: Nguyen Khang Pham, Annie Morin, Sébastien Campion [correspondent].

CAVIZ (*C*orrespondance *A*nalysis *VIZ*ualisation) is an interactive graphical tool, which allows to display and to extract knowledge from the results of a Correspondence Analysis on images.

Registration at the “Agence pour la Protection des Programmes” (APP) is in progress.

5.1.8. Kiwi

Participants: Gwénolé Lecorvé, Pascale Sébillot [correspondent].

KIWI (standing for *Keyw*ords *E*xtrator) is a software which seeks to convert textual documents into weighted word vectors using the TF-IDF method, widely used in information retrieval. Hence, Kiwi is mostly dedicated to indexing and keyword extraction purposes. Besides raw texts, the standard TF-IDF method has been extended to process transcriptions and word meshes, to return word n-uple vectors, and to integrate external knowledge on words such as confidence measure or mutual information.

Registration at the “Agence pour la Protection des Programmes” (APP) is in progress.

5.1.9. Topic Segmenter

Participants: Camille Guinaudeau, Pascale Sébillot [correspondent].

In collaboration with G. Gravier from the Metiss project-team.

The TOPIC-SEGMENTER program is a software dedicated to topic segmentation of texts and (automatic) transcripts, mostly based on lexical cohesion, implementing (and extending) a method described in the paper from Masao Utiyama and Hitoshi Isahara, "A Statistical Model for Domain-Independent Text Segmentation", ACL, 491–498, 2001.

Registration at the “Agence pour la Protection des Programmes” (APP) is in progress.

5.1.10. S2E

Participants: Sébastien Campion [correspondent], Mathieu Ben.

S2E (*S*tructuring *E*vents Extractor) is a Python module which allows the automatic discovery of audiovisual structuring events in videos. The algorithm is based on a cross-modal mutual information measure between audio and video clusters. These mono-modal clusters are obtained from two separate hierarchical clustering processes, one on audio segments, the other one on video shots. The module outputs a segmentation of the video with a score for each segment indicating how confident we are that a structuring event occurs during the corresponding time slot.

Registration at the “Agence pour la Protection des Programmes” (APP) is in progress.

5.2. Demonstrations

5.2.1. Automatic Generation of Hypervideos

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

This work was made with the help of Guillame 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 (3.5) 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 [correpondent], Laurent Amsaleg, Nguyen Khang Pham.

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 Retrival by Correspondence Analysis) search engine.

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 (up to 70TB), and contains 4 acquisition devices (for Digital Terrestrial TV), 3 video servers, and 15 computing servers partially included in the local cluster architecture. In 2008, we acquire a new server with 96 GB of memory which enable to improve the speed of building index or language model. A memory upgrade was also done on servers.

A dedicated website has been developed in 2009 to provide a user support. It contains useful informations 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 practises and cookbooks for treatments of large datasets.

The platform will be completed with dedicated software to manage all the metadata associated with the data.

In 2008, we build up a corpus of multimedia data. It consists in a continuous recording (6 months) of two TV channels and three radios. It also includes web pages related to these contents captured on broadcaster's website. This corpus is to be used for different studies like the treatment of news along the time and to provide sub-corpus like TV news within the Quaero project (see below). The manual annotation of all the TV programs is under progress.

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

6. New Results

6.1. Advanced Algorithms of Data Analysis, Description and Indexing

6.1.1. Advanced Image Description Techniques

6.1.1.1. Morphological processing of images

Participant: Sébastien Lefèvre.

Despite its widespread use in image processing, mathematical morphology has mainly been ignored in the context of multimedia analysis. However, this toolbox makes possible the design of scale-spaces and subsequent global or local image descriptors, which may offer several advantages over the state-of-the-art: efficient algorithms to produce local descriptors with a limited computation time, strategies to ensure compactness of the image description, preservation of edge information through the multiresolution scheme, and finally intrinsic ability to offer invariance to several image transforms. So we have started to study the potential interest of mathematical morphology for image description.

6.1.1.2. 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, in collaboration with Christine Guillemot from TEMICS 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. These methods provide several advantages. First, unlike critically-sampled transforms as separable wavelet transforms, curvelets and bandlets that have been considered for compression and de-noising applications, they allow the extraction of low level signal features (points, edges, ridges, blobs) or of local descriptors. Then, sparse representations allow the use of inverted files, which provide a solution to the indexation of high dimensional data by taking advantage of sparse vectors properties. Indeed, document similarity calculations are thus carried out efficiently using the scalar products between these sparse vectors. In this context, two approaches have been considered: descriptors sparse representation and signal sparse description.

Concerning description sparse representation, we proposed an approach that applies a pursuit-based sparse decomposition to each SIFT descriptor of an image to obtain a sparse vector [28]. The descriptors sparsity still enables the use of inverted file type indices. The aim is to tackle the problem of SIFT descriptor high dimensionality, while retaining the local property of the input descriptors. We compare our approach in the context of local querying to the *Video Google* one, where multiple input SIFT descriptors are aggregated into a single sparse descriptor, resulting in the loss of description locality.

Concerning signal sparse description, our aim is to adapt existing work in sparse decomposition (designed with compression and prediction in mind) to the construction of image descriptors displaying covariance to the set of admissible transformations. In this context we have identified three problems to address: 1) dictionary design, 2) atom selection method and 3) descriptor comparison method.

Regarding descriptor comparison method, we introduced a new method to search for approximate nearest neighbors (ANN) under the normalized inner product distance, 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, approach considered in the literature for content-based image search.

Regarding dictionary design, we study a method to construct dictonaries in a way that a different dictionary is used at each iteration of the decomposition, and that these *iteration tuned dictionaries* satisfy some desirable properties. This method and associated algorithm gave promising results and should be validated next year for compression and indexing purposes.

6.1.1.3. 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 [68], we can use NLP and text retrieval techniques to improve image retrieval;
- high-level image description: NLP and text retrieval techniques can be use to mine textual information coming with images, such as the news articles that images illustrate, and extract textual information to describe the images.

We worked in 2009 on each of these two stages.

First, we continued the work about the use of weighting schemes [64], [45] and Minkowski distances for visual word-based image retrieval [68] that we initiated in 2008. We showed that the optimal parameters (weighting scheme and distance) of visual word-based systems strongly depend on the kind of query considered. It questions some common habits in visual word-based retrieval that consider tf.idf as the state-of-the-art weighting scheme and L1 as the state-of-the-art distance. It also shows that using only one dataset to evaluate

visual word-based systems, as it is the case most of the time, should be avoided because the query's properties can sharply change from one dataset to another, making the results unstable. At last, this work highlightens a correlation between the use of certain weighting schemes and the effect of the Minkoswki distance parameter. This result might be particulary interesting in the case of fractional distances [44], which generally offer the best performance but do not respect the triangle inequality: using L1 distance with adapted weights could overcome the triangle inequality limitation and offer the same performance as a fractional distance at the same time.

Then, we worked on high-level image description, using NLP techniques to extract textual image descriptors from the text accompagning images in a large parallel text-image corpus of news articles. We proposed to annotate images containing logos with named entities from suited categories (names of brands, products, companies, organizations, events and artistic groups). We first developed a fast logo detector relying on the visual word scheme. This detector can achieve 95% detection precision with a 60% recall, or 80% precision with a 80% recall. We used this detector to select images containing one or more logos, then we annotated each of these images with the most frequent suited named entity that appears in the article coming with the image, following a method similar to the one we already used to annotate face images [69]. This annotation method achieves an acceptable annotation accuracy (between 40 and 70 %) on our news corpus. We then explored more complex entity selection criterias, using the document frequency or annotation frequency of named entities. We showed that including annotation frequency to entity scores can provide slightly better results than pure frequency, for logo annotation as well as face annotation. This result means that, on news corpora, images tend to contain similar general information rather than varied precise information.

6.1.2. Advanced Data Analysis Techniques

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

Participant: Annie Morin.

Textual data can be easily transformed in frequency tables and any method working on contingency tables can be used to process them. Besides, with the important amount of available textual data, we need to find convenient ways to process the data and to get invaluable information. It appears that the use of factorial correspondence analysis allows us to get most of the information included in the data. But even after the data processing, we still have a big amount of material and we need visualization tools to display it. We study the relevance of different indicators used to cluster the words on one side and the documents on the other side and we are concerned by the visualization of the outputs of factorial analysis: we need to help the user to go through the huge amount of information we get and to select the most relevant points. Most of the time, we do not pre-process the texts: that means that there is no lemmatization. We also start exploring temporal changes in textual data and the first experiments have been done on newspaper corpus from 1987 to 2003. For the moment, we mainly focus on the visualization of results.

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

Participants: Nguyen Khang Pham, François Poulet.

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 [11]. The GPU version of the algorithm is 130 time 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 datapoints 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 are currently studying the same kind of principles (incremental and parallel) with other classification algorithms in order to deal with very large image datasets.

6.1.2.3. Intensive Use of Data Analysis Methods for Image Mining

Participants: Patrick Gros, Annie Morin, Nguyen Khang Pham, François Poulet.

This work is done with Institut francophone pour l'informatique, Hanoï, Vietnam.

To analyze and retrieve information in image databases, we use the same method as in textual data analysis. This work is part of the Ph.D. thesis of Pham Nguyen Khang. That means that in order to apply Correspondence Analysis (CA) to images, we must define "words" in images. This is usually achieved by 2 stages: (1) vector quantizing automatically extracted local image descriptors (*i.e.*, SIFT) and (2) applying a clustering algorithm (*i.e.*, k-mean clustering) on the set of descriptors to form "visual words". Once the visual words are defined, we then construct a contingency table by crossing visual words and images.

We began experiments by applying CA on a small database (961 images). First, a vocabulary of 1000 visual words was computed from 60 first images using SIFT descriptors (code of D. Lowe) and a k-mean algorithm. CA was then applied on a contingency table of 961 x 1000. We kept only 30 first axes and used those axes for computing image similarity (Euclidean distance was used). Surprisingly, we found some groups of images which belong to the same categories (toys, houses, Eiffel tower...).

Motivated by this promising result we continued our approach on the "caltech4" database (4090 images of 5 categories: faces, airplane, motorbikes, cars (rear) and backgrounds) [67]. About 3000 descriptors sampled from all of descriptors (a third for every category) were clustered to form 2224 visual words. We explored this database on 2 tasks: image categorization and image retrieval. For the first task, we applied CA on the contingency table representing the database and kept only 7 axes (for comparison to PLSA trained with 7 hidden topics). A k-mean algorithm was then invoked to form clusters (categories). The result showed that CA performed slightly better than PLSA.

For image retrieval task, we compared our approach to PLSA and TF*IDF using L1 distance, L2 distance and cosine similarity. In the case of PLSA and CA, the retrieval was performed very fast because the problem dimension was reduced from 2224 to 7. For PLSA and CA, cosine similarity gave better result than L1 and L2 distance. The performances of CA and PLSA were equivalent and much better than that of TF*IDF.

We have also proposed a method for scaling up the problem using inverted files based on image representation quality on axes. Every inverted file was associated with the well represented images (on the axis to which the file belong). Given an image query, the search began by choosing the appropriate inverted files and intersecting those files. The similarity computation is done only on a subset of images resulting from the intersection of inverted files (about 1/5 to 1/8 of entire database). The performance was degraded only about 0.3% with respect to the exhaustive method.

Besides, in order to reduce the learning time and to deal with large databases, we have developed an incremental version of CA algorithm which splits data in blocks and processes block after block [34]. The parallelization of this algorithm on Graphic Process Unit (GPU) showed that the GPU version performs 20 to 30 times faster the CPU one [26]. The retrieval's accuracy is also improved when combining contextual information into our index structure. For that, we have integrated the Contextual Dissimilarity Measure [56] into our retrieval platform using inverted files. This integration was explored in two directions: offline (correction terms are computed before the retrieval task) and online (correction terms are computed only on images in the candidate list). Tests realized on the Nister dataset [60] have shown a significant improvement of the accuracy. Our algorithm have been assessed for large scale datasets: we have merged the Nister dataset

with 1 million other images. With our method, only 0.06% of the dataset was explored (in 1/8 second) and that 0.06% contains 86.7% of relevant images. We also have investigated [30] a new approach for unsupervised classification with random oblique decision trees for very high dimensional data. Nguyen Khang Pham has defended his Ph.D. thesis [10] in November 2009.

6.1.3. Advanced Indexing Algorithms

6.1.3.1. Indexing for Very Large High Dimensional Spaces

Participant: Laurent Amsaleg.

This topic is done in close cooperation with Hervé Jégou from Lear project-team and is a joint work with researchers from Reykjavík University.

It is well known that high-dimensional nearest-neighbor retrieval is very expensive. Dramatic performance gains are obtained using approximate search schemes, such as the popular Locality-Sensitive Hashing (LSH). Several extensions have been proposed to address the limitations of this algorithm, in particular, by choosing more appropriate hash functions to better partition the vector space. All the proposed extensions, however, rely on a *structured* quantizer for hashing, poorly fitting real data sets, limiting its performance in practice.

We have studied the two families of quantization schemes for hashing when used in high-dimensional settings. We conclude that using a k-means *unstructured* quantizer for hashing significantly improves the accuracy of LSH, as it closely fits the data in the feature space. We have also designed two variants of the k-means approach offering different trade-offs in terms of memory usage, efficiency and accuracy.

The joint work with researchers from Reykjavík University has been partially focused on ways to very efficiently compute local image descriptors as video analysis using local descriptors demands high throughput of the local descriptor creation process. The most practical method to achieve the high throughput is to use the GPUs delivered with most recent computers. We have been working on adapting the computation of the Eff2 descriptors, a variant of SIFT, to the GPU. We have compared our GPU-Eff2 descriptors to SiftGPU, another GPU-based variant of SIFT, and showed that while both GPU-based variants yield similar results, the GPU-Eff2 descriptors require less than half of the processing time required by SiftGPU. Furthermore, our analysis shows that the SiftGPU descriptors also use the CPU, making them less scalable than GPU-Eff2.

6.1.3.2. Challenging the Security of CBIR Systems

Participants: Laurent Amsaleg, Ewa Kijak.

This work is done in collaboration with Teddy Furon from the TEMICS *project-team.*

Content-Based Retrieval (CBR) is the generic term describing the problem of searching for digital material in large multimedia databases. CBR systems are of great diversity: they deal with a plurality of media (text, still images, music, videos) and offer a wide range of querying modes, from simple query-by-example schemes to more complex search processes involving user feedback aiming at best bridging the semantic gap. In a way, CBR systems promote the cultural and historical value of multimedia contents. They make the multimedia databases very useful, their contents reusable, spurring the enrichment of the artistic and cultural patrimony. CBR has proved to be a marvellous technology, recognizing content even when deeply distorted. Overall, CBR systems have so far been used in very cooperative and "friendly" settings where it benefits content providers business, while increasing users digital experience enjoyment.

However, we recently witness another use of this technology. CBR is used to *filter* multimedia contents in order to protect the creation of the few from the piracy of the many. CBR techniques are used to "clean the Internet", stopping the upload of copyrighted material on User Generated Contents sharing platforms such as YouTube, or forbidding downloads from P2P networks. Overall, filtering is an application of CBR techniques that is quite different from its primary goal: the environment is now hostile in the sense that filtering restricts users freedom, controlling and/or forbidding distribution of content.

While the cryptographic community has been investigating the security of systems for years, almost no work address this issue in the computer-vision community. We therefore started to consider how security issues can impact the typical components building a complete content-based information retrieval system (CBIRS). We are exploring three avenues of research that are (i) a threat analysis where we study the crucial elements to consider for assessing the security of CBIRS; (ii) attacking the core technologies of state-of-the-art CBIRS to discover an initial set of potential security flaws; (iii) trying to attack various specific techniques, at the description level and at the database level to show that challenging the security of CBIRS is feasible in practice.

6.1.4. Large scale evaluation of global GIST descriptors

Participant: Laurent Amsaleg.

Our work on this topic is done in close collaboration with Matthijs Douze, Hervé Jégou, Harsmirat Sandhawalia and Cordelia Schmid and the LEAR *project-team.*

To make an image index at the web's scale, each server has to handle 10 to 100 million images. At this scale, it is no longer possible to use local descriptors: the memory usage of the descriptors becomes prohibitive. 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 global descriptors.

We evaluated the search accuracy and complexity of the global GIST descriptor for two applications: same location/object recognition and copy detection. We identified the cases in which a global description can reasonably be used. The comparison is performed against a state-of-the-art bag-of-features representation.

We proposed an indexing strategy for global descriptors that optimizes the trade-off between memory usage and precision. Our scheme provides a reasonable accuracy in some widespread application cases together with very high efficiency: In our experiments, querying an image database of 110 million images takes 0.18 second per image on a single machine.

The system is intended as a rough pre-filter. Images that are preselected (the short-list) can then be reprocessed by a more costly algorithm to produce more accurate results. We evaluated the method on standard datasets used to test image matching methods. The method and the experiments have been published in [20].

6.1.4.1. 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, we can quite well exploit rather large databases of still images and we know how to efficiently query them by content. The next step asks to turn our focus on more complex documents, typically video and audio. There are today several description techniques for audio and video but only very few techniques efficiently perform query-by-content on video or audio databases at large scale. Being able to use such techniques is particularly crucial for professional multimedia archivists.

The state of the art makes such searches possible, but only at a very small scale, i.e., on a very small amount of data. Today, no search technique is efficient enough to allow any practical usage of real-scale audio or video archives. In addition, simply extending existing multidimensional indexing techniques is not possible since they were designed for description schemes in which the concept of sequence is lacking.

We have started investigating this issue in 2007. Overall, deciding whether two sequences of descriptors are similar requires to clarify what elements should be compared, and how the comparison should be enforced. 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. It is in fact an *a posteriori* alignment of the sequences to compare. Here, the similarity of sequences is directly related to the similarity of the descriptions. One of the key points here is that computing optimal alignment is costly in terms of computing time, which is why we investigated ways to approximate the alignment using few computations.

These initial results suggest to push 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 a Support Vector Machine approach used in regression (and not in classification, as usually done). 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 first tried to use cross-similarity estimation based metrics to compare models as direct comparison between models is impossible. Another way we investigated was to build sequence features based on the representability of a sequence with respect to a set of reference models. Such a feature space could then be indexed by any classical indexing techniques.

6.1.4.2. Browsing Personal Image Collections

Participant: Laurent Amsaleg.

Our work on this topic is done in close collaboration with Kári Harðarson from the University of Reykjavík.

The Database Lab at Reykjavík University is currently writing a photo browser. One of the main ideas being tested is to use the location of thumbnails on screen to indicate properties of the underlying photos. Users can select which properties of photographs map to which attribute of the thumbnail, it's location on screen, it's brightness and size etc.

One of the uses for this feature is to let the computer group together pictures on screen that may have the same people on them. The user can see at a glance whether a cluster of photos contains the same person or whether he needs to drag some photos to a different cluster. If he does, the browser notifies the face recognition module that the photos did not portray the same person so that it can learn from it.

This mechanism is general and could be used to classify any property the images may have but using it to represent the results of facial recognition seems promising.

During the preceeding years, we have been focusing on the underlying database and on the on-screen presentation. We did not had, however, a module that would recognize faces and return the amount of likeness between faces as a numeric value that could be displayed.

We used the financial support of the Eff^2 Associate Team to send to Reykjavík University an intern for developping such a tool and integrate the resulting software modules in the current browser prototype. The intern had to find and evaluate an already available face recognition library and then find a way to package it in such a way that our browser could use it to decide where to place thumbnails on the screen.

He finished the project successfully and handed in a module before he left Iceland. Although the browser has not been connected to the library yet, we have seen a demonstration where the library is informed of pictures to scan and returns information about the location of faces in the pictures and their respective likenesses. When fall semester ends, the browser work will continue and the library will be connected to the browser sometime in the spring.

In addition, it is worth noting that another French intern did some work for that image browser, payed by Reykjavík University, however. Both works make a consistent story: in addition to face recognition, the second project was to design a "slideshow" presentation module which takes a set of photos and analyzes them to look for similar photos. Photos that have something in common are then displayed in sequence during the presentation. If many photos have similar subjects, they are reduced in size and shown side by side during the show. Even though the pattern recognition is primitive, it nevertheless serves to make the slide presentations more pleasing to the eye and it looks as if someone planned it carefully.

6.2. New Techniques for Linguistic Information Acquisition and Use

6.2.1. NLP for Document Description

6.2.1.1. 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) and started in October with the thesis of Ali Reza Ebadat.

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 of terminologies. In this context, we focused our work on the term normalisation problem: we are developing various methods to generate the variant forms of an initial list of terms (acronyms, graphical, morphological, morpho-syntactic variants), based on previous work on analogical learning.

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. These relations can be very different from a domain to another and from a final application to another. Thus, it is important to develop generic methods to detect them. This year we began to study the use of machine learning techniques to extract relations. In particular, we began to develop an system inferring extraction patterns from examples, based on previous work using Inductive Logic Programming. We also studied the use of probabilistic approaches like Conditional Random Field to detect relations.

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 ans standard program guide information, closed captions, associated websites... All these sources offers a way to exploit complementary information that can be used to semantically annotate multimedia document. This year, TEXMEX and other Quaero partners were specially involved in the definition, the building and the annotation of such a multimedia corpus composed of football matches.

6.2.2. Oral and Textual Information Retrieval

6.2.2.1. Phonetization

Participant: Vincent Claveau.

Phonetization is a crucial step for oral document processing. In 2009, we have proposed a new letter-to-phoneme conversion approach; it is automatic, simple, portable and efficient. It relies on a machine learning technique initially developed for transliteration and translation; the system infers rewriting rules from examples of words with their phonetic representations [29], [19]. This approach was evaluated in the framework of the Pronalsyl Pascal challenge, which includes several datasets on different languages. The obtained results equal or outperform those of the best known systems. Moreover, thanks to the simplicity of our technique, the inference time of our approach is much lower than those of the best performing state-of-the-art systems.

6.2.2.2. Information Retrieval in the TV context

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

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

The main focus of our research is to conceive new generation of information retrieval (IR) systems capable of retrieving information from TV data. Directly indexing automatic transcripts is errorprone because of transcription errors, in particular in the TV context where error rates can be high for some programs. The main challenge is therefore to develop IR approaches able to retrieve information in degraded text.

In October 2009, Julien Fayolle (supervised by F. Moreau, C. Raymond and P. Gros) started a Ph.D. whose aim is to investigate IR approaches robust to transcription errors. To this end, we focus on three main points: (i) detecting portions of transcripts most likely to contain errors, (ii) designing innovative representations to ensure more flexibility than purely text or phonetic ones, and (iii) adapting IR mechanisms to the new representations and to the TV context where the notion of document is not clearly defined. As an initial step, we are studying a hybrid lexico-phonetic representation, emphasizing named entities which are highly problematic for automatic speech recognition systems.

6.2.2.3. 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 Information Retrieval Systems (IRSs). This model has rapidly shown its limitations and is no more used in Information retrieval (IR). Among the reasons, the Boolean approach 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 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 [17]. These positive results have validated the proposed approach, while negative ones have given some insights on the properties needed by such a model [18].

From these encouraging results, perspectives have been derived, and will be addressed. The link between our fuzzy model and logical models in IR, and with language models in IR, are currently studied. Among other perspectives, this graded inclusion-based model gives new and theoretically grounded ways for a user to easily weight his query terms, to include negative information in his queries, or to expand them with related terms...

6.3. New processing tools for audiovisual documents

6.3.1. TV Stream Structuring

6.3.1.1. TV stream macro-segmentation

Participants: Patrick Gros, Gaël Manson.

This work is done in the framework of the thesis of Gael Manson at Orange Labs (former France-Telecom R&D).

TV stream macro-segmentation basically aims at precisely determining the start and the end of each of the broadcasted programs and inter-programs.

A first step in automatic TV macro-segmentation is segmenting TV stream in programs and inter-programs segments. The most promising automatic approach for that consists in detecting inter-programs as near-identical repeated sequences. However, the resulting repeated sequences can only detect sequences that repeat in the stream. Some of these sequences are actually inter-programs and others belong to long programs. Moreover, some inter-programs do not repeat.

It is therefore necessary to classify the segments resulting from the repeated sequences detection phase (the occurrences of repeated sequences and the rest of the stream) into program and inter-program segments. Our solution for that is based on Inductive Logic Programming. In addition to intrinsic features of each segment (e.g. duration, number of repetitions), our technique makes use of the relational and contextual information of the segments in the stream (e.g. the class of following segments, the class of other occurrences of repetitions).

The last step in automatic TV macro-segmentation is TV program extraction. One TV program can be split in several parts over a set of consecutive program segments. Consecutive program segments of the same TV program have thus to be reunified or fused in order to retrieve the entire TV programs. When the EPG or EIT metadata is available, we align the program segments with it to perform the reunification. When no metadata is available, our approach relies on analyzing the visual content and characteristics of each pair of consecutive segments in order to decide if they have to be reunified or not. It uses, amongst others, content-based descriptors like the color distribution, the number of faces in each segment and also the number of near-identical shots between the two segments. These descriptors are then used within an SVM classifier that makes the final decision.

6.3.1.2. Repetition Detection-based TV Structuring

Participants: Patrick Gros, Zein Al-Abidin Ibrahim, Emmanuelle Martienne, Sébastien Campion.

This work is done under the Semim@ges project in collaboration with the CAIRN team.

In the frame of the ANR Semim@ges project, we have tried to develop a slightly different approach than that we developed with Gaël Manson. This approach is based on the same fundamentals: detection of repeated sequences and classification of these segments followed by an alignment with an Electronic Program Guide. The difference comes from the fact that we base the classification step of the repetition patterns themselves, and not only on the local environment of each segment which is repeated somewhere else. As a matter of fact, the classification tries to differentiate the repetition schemes of advertisements, programs, and other segments based on their number of apparitions and on various descriptors on each of the occurrences. The method provided first promising results, but have strong experimental constraints. The repetition patterns must be learned on long video sequences (several weeks). The repetition patterns have variable sizes when many learning techniques requires fixed size data. On the other hand, this method allow to take into account long-term information.

6.3.1.3. Semantic TV Program Verification

Participants: Camille Guinaudeau, Pascale Sébillot.

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

Over previous years, we have developed an approach for the automatic labeling of program segments from the electronic program guide (EPG) [59]. In 2008, we developed a method to validate this labeling from a semantic point of view. In this method, program descriptions from a TV guide are associated with each segment in the TV stream, using IR methods on the phonetic and lexical transcripts of the speech material. As a result, each segment is associated with a program name which is used for the validation of the EPG alignment. In 2009, we improved this technique relying on an anchorage system that limits the association of a segment and a description only if they are contained in the same 2 h time-slot [32], [21]. The new method was evaluated on a larger corpus (650 segments, transcripts lengths vary from 7 to 27,150 words, with an average of 2,643.). Time anchorage improved recall and precision resp. from 64 % to 86 % and from 47 % to 63 %.

6.3.2. Program Structuring

6.3.2.1. Stochastic Models for Video Description

Participants: Siwar Baghdadi, Patrick Gros.

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

Bayesian Networks are an elegant and powerful semantic analysis tool. They combine an intuitive graphical representation with efficient algorithms for inference and learning. They also allow the representation in a comprehensive manner of the interaction between a system variable and the integrating of external knowledge. Unlike HMM and segment models, structures of Bayesian Networks are very flexible and can be learned from data. We explored the idea of using Bayesian Networks and their temporal extension Dynamic Bayesian Networks to do event detection in video streams. As a first application we have chosen commercial detection.

We modeled the video stream by a Dynamic Bayesian Network. According to this model, the video stream is a sequence of observations (a set of multimodal features). Each observation is generated according to the state of the system (program or commercial). The model is fed with knowledge about the commercial segment duration. Detecting commercial segment is then a problem of inferring the optimal sequence of hidden nodes with the convenient duration. Structure learning allowed us to learn the optimal interaction between variables. Future work involves the extension of our model to do event detection in soccer games. The challenge of this part is to take into consideration all kinds of feature interactions (spatial, temporal at short or long term).

6.3.2.2. Automatic discovery of audiovisual structuring events in videos

Participants: Mathieu Ben, Sébastien Campion.

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

Extraction of structuring events in video programs is an important pre-processing step for higher level video analysis tasks. However most current techniques rely on supervised approaches specifically dedicated to a given target event, for example detection of anchorperson shots in TV news programs.

To overcome this genericity issue, we have developed a cross-modal technique for the automatic discovery of audiovisual structuring events in TV programs, using only little prior knowledge for the definition of the targeted events. The algorithm is based on two separate hierarchical clustering processes, one for audio segments and one for video shots. The two resulting clustering trees are then crossed by measuring the mutual information between each pair of audio/video (A/V) clusters. The A/V cluster pairs are then ordered in an N-best list from which we are able to obtain a confidence score for each segment in the video. Applying a varying threshold on these scores, one can select more or less segments thus fixing a tradeoff between missed structuring events and falsely detected ones.

Experiments on several kinds of TV programs have shown that the technique is able to extract the most relevant parts of the video, from a structuring point of view: anchorperson shots for TV news and report programs, audio/video jingles separating the reports for flash news programs. On a manually labeled database of 59 TV news programs, we were able to obtain more than 80% precision with 80% recall for the detection of anchorperson shots.

We are currently investigating how to efficiently use the results from this structuring event extraction module to guide the "best path" decoding process of our topic segmentation system.

6.3.3. 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 and Pierre Cauchy 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 specificities 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. Finally, we study the effective use of semantic analysis for video structuring.

In 2009, we mostly focused on transcript-based topic segmentation and on unsupervised topic adaptation of the ASR system.

In order to achieve robust topic segmentation of TV streams, we worked on improving our transcript-based topic segmentation method based on an extention of the initial work of Utiyama and Isahara [70] that accounts for additional knowledge sources such as acoustic cues or semantic relations between words [55]. Firstly, we further investigated the use of semantic relations, implementing a mathematically rigorous framework to account for such relations and comparing several methods for their automatic corpus-based acquisition. We demonstrated that directly using automatically generated semantic relations increases precision on topic

boundaries to the expanse of a lower recall. Secondly, we combined transcript-based topic segmentation with automatic discovery of structural elements in videos (see 7.2.1), using the latter as a prior information on the location of topic boundaries. Preliminary results showed that the detection of structural elements is not accurate enough, in particular regarding the boundaries of such elements, to improve topic segmentation. Future work includes the selection of relevant semantic relations and better consideration of structural elements.

Since the quality of transcriptions is a central factor for many NLP tasks in the semantic analysis step, we aimed at thematically adapting the language model of the ASR system to thematically coherent segments using related documents retrieved from the Internet [22]. More specifically, we sought to better understand which factors of the MDI-based language model adaptation step most affect the recognition rate. Experiments reported in [22] have shown that considering only a small number of terms in MDI contraints, *i.e.*, topic-specific words, is sufficient to perform an efficient adaptation. In addition to this result, it has also been shown that these terms can be automatically extracted from a small topic-specific corpus without any prior knowledge.

With the arrival of Christian Raymond in September 2009, we initiated work on entity detection in spoken document as named entities are key elements for speech-based segmentation, structuring and indexing of videos. We developed and compared two baseline methods for named entity detection in speech transcripts, respectively based on support vector machines and conditional random fields.

As mentioned in introduction, using speech as a description of multimedia segments makes it possible to structure and organize collections of documents from a semantic viewpoint. Our work on topic segmentation and keyword extraction over the last few years is at the core of a demonstration of multimedia navigation in a collection of TV news documents that we implemented in 2009 (see 5.2.1).

7. Contracts and Grants with Industry

7.1. Contracts, Initiatives and Participation to Networks of Technological Research

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 region 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 Thomson 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.

7.1.2. Contract with Thomson

Participants: Patrick Gros, Siwar Baghdadi.

Duration: 36 months, finishing March 31st, 2009.

S. Baghdadi's Ph.D. thesis is supported by a CIFRE grant in the framework of a contract between Thomson and TEXMEX.

7.1.3. Contract with Orange Labs

Participants: Patrick Gros, Gaël Manson.

Duration: 36 months, starting in November 2006.

G. Manson's Ph.D. thesis is supported by a CIFRE grant in the framework of a contract between Orange Labs and TEXMEX.

7.1.4. Contract with Orange Labs

Participants: Pascale Sébillot, Khaoula Elagouni.

Duration: 36 months, starting in October 2009.

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

7.2. European Initiatives

7.2.1. Quaero

Participants: Laurent Amsaleg, Mathieu Ben, Vincent Claveau, Patrick Gros, Camille Guinaudeau, Ewa Kijak, Stacy Payne, Pascale Sébillot, Fabienne Moreau, Julien Fayolle, Ali Reza Ebadat.

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

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, and Multi-modal Video Structuring.

In 2009, TEXMEX's participation in QUAERO increased steadily over the course of the year. Starting with the 3 people already hired on the contract (Camille Guinaudeau, PhD student, and Stacy Payne and Mathieu Ben, engineers), they were joined by 3 interns who did annotation work on the video segmentation corpus over the summer, followed by the arrival of 2 PhD students in Sept-Oct (Julien Fayolle, Ali Reza Ebadat). Finally, in December an engineer, initially foreseen for VistaS team but reassigned to TEXMEX, started working on one of the tasks.

8. Other Grants and Activities

8.1. National Initiatives

8.1.1. ANR project Semim@ges

Participants: Zein Al-Abidin Ibrahim, Patrick Gros, Emmanuelle Martienne, Sébastien Campion.

Duration : 27 months, starting in January 2007. Partners: Orange Labs, TDF, Kersonic, Telisma, CAIRN team.

The project is devoted to TV data exploitation and repurposing. Two main applications were considered: TV news analysis, and TV streams structuring. TEXMEX project-team was mainly involved in the second one. The aim of our work was to structure automatically long TV streams in more usable units like programs or non-program sequences, exactly like it was done in Xavier Naturel's thesis, but with no a priori manual annotation. A method was developed based of the detection of repeated segments and on an automatic unsupervised classification of the repeated sequences. The Semim@ges project was demonstrated during the NEM Summit in Saint Malo, France, in September 2009.

8.1.2. ANR project ICOS-HD

Participants: 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 resulting 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 on the study of new signal representation amenable to both compression and image description, as well as descriptor adaptation for image retrieval in large databases.

8.1.3. ARC INRIA RAPSODIS: Syntactic and Semantic Information-Based Automated Speech Recognition

Participants: Camille Guinaudeau, Gwénolé Lecorvé, Pascale Sébillot.

Duration: 2 years, starting in February 2008. Partners: METISS, PAROLE, TALARIS *project-teams, CEA-LIST/LIC2M.*

This project aims at improving automatic speech recognition (ASR) by integrating linguistic information. Based on former work by S. Huet concerning the incorporation of morpho-syntactic knowledge in a post-processing stage of the transcription, we experiment, together with our partners, the deep insertion of automatically obtained semantic relations (especially paradigmatic ones) and syntactic knowledge within an ASR system.

In 2009, the objectives of the project were extended to include semantic knowledge acquisition and the use of such knowledge for spoken document processing in addition to speech transcription. In this extended framework, we have worked on corpus-based acquisition of semantic relations for topic segmentation of spoken documents. We compared various classical methods for relation acquisition and measured their impact on out topic segmentation system.

8.2. International Collaborations

8.2.1. 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.

Image databases, and content-based image retrieval systems in particular, have become increasingly important in many applications areas. While extremely effective (they return high quality results), these systems are very inefficient (they answer very slowly) due to their complexity, to the curse of dimensionality problems and to the scale at which they have to run when dealing with collections of realistic sizes.

Originally, 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. We found, however, that the inefficiency observation applies not only to images, but to most multimedia documents as soon as the retrieval is based on the similarity of low-level descriptions.

Therefore, the Eff2 associate team has widen its scope and tackles efficient database support for more general large-scale multimedia applications and problems that mostly deal with high-dimensional low-level features. We now investigate additional issues such as efficient query execution on large collections of sequences of low-level descriptions (e.g., audio or video collections). We have also initiated another thread of research by investigating the browsing of personal image collections. Going to sequences and addressing browsing issues (and not solely searching) is a new development in our cooperation.

8.2.2. 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

A. Morin got two Egide contracts with Slovenia (Proteus) and Croatia (Cogito) for 2007 and 2008 on knowledge discovery and visualization for textual data. In Slovenia, we work with Blaz Zupan and Janez Demsar from faculty of Computer and Information Science, University of Ljubljana and in Croatia with Bojana Dalbelo Bašić from faculty of Electrical Engineering and Computing, university of Zagreb. The French laboratory ERIC (university of Lyon 2) is the other French partner.

The concerned research teams have different expertise on the same subject: machine learning for the Slovenian and Croatian teams, statistics for the French teams 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 plan to implement a new visualization system for textual data. Proposed collaboration includes sharing of a number of Ph.D. students.

In spite of the end of our Egide collaboration, two papers were presented this year, the first one with Slovenian team, dealing with subgroups discovery when there exists dependencies between explanatory variables [27]. The second one deals with textual data analysis and correspondence analysis. It investigates how the document representations (words, word n-grams, letter n-grams) influences the quality of visualization [13]. Additionally, Artur Šilić from the university of Zagreb continued his work on temporal changes in text streams using the New York Times corpus.

8.2.3. Collaboration with Nagoya University and National Institute of Informatics, Japan

Participants: Laurent Amsaleg, Patrick Gros, Fabienne Moreau, Annie Morin, Pascale Sébillot, Julien Fayolle.

In the framework of our collaboration with both the University of Nagoya (Ichiro Ide, Associate Professor) and the National Institute of Informatics in Tokyo (Shin'ichi Satoh, Professor), Julien Fayolle spent 4.5 months in Japan for a joint Master's degree internship dedicated to topic segmentation of TV broadcast news, integrating video clues, and to topic-threading.

8.3. Visits of foreign researchers, Invitations to foreign labs

8.3.1. Visit to the University of Reykjavík

Participants: Laurent Amsaleg, Ewa Kijak.

Laurent Amsaleg and Ewa Kijak were invited to visit the University of Reykjavík for a week. They gave classes on multidimensional indexing and clustering and on still images and video descriptions. They also met several students and discussed on-going projects. Future work on audio indexing and classification can be foreseen.

8.3.2. Visit of members of University of Reykjavík

Participants: Laurent Amsaleg, Ewa Kijak.

Björn Ãór Jónsson, Haukur Palmason et Gylfi Ãór GuÃ°mundsson spent one week at Irisa. They came to push the work initiated in Iceland on audio indexing and classification of musical tunes. In addition, discussions on a new high-dimensional clustering algorithm took place. We also discussed its integration into our demo environment.

8.3.3. Visit from members of University of Nagoya and NII Tokyo

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

Partner: Ichiro Ide, Associate professor at the University of Nagoya, and Shin'ichi Satoh, Professor at NII (National institute of informatics) Tokyo, Japan.

In the framework of TEXMEX's collaboration with the University of Nagoya and NII Tokyo, Ichiro Ide spent 5 days in the team in March-April 2009. These days were dedicated to discussions and common work about TV structuring, topic segmentation, and to the definition of the subject of J. Fayolle's internship.

9. Dissemination

9.1. Conference, Workshop and Seminar Organization

- P. Gros was co-chair of the First International Workshop on Content-Based Audio/Video Analysis for Novel TV Services which was organized in conjonction with the International IEEE Symposium on Multimedia, San Diego, California, USA in December 2009.
- F. Poulet, B. Le Grand, T-N. Do and M-A. Aufaure organized and edited the proceedings of the 7^{th} Workshop Visualisation et Extraction de Connaissances co-located with Extraction et Gestion de Connaissances, (EGC'09), Strasbourg, France, Jan. 2009.
- A. Morin has coorganized an invited paper meeting on Random Projection for multimedia retrieval during the 57th Session of the International Statistical Institute hold on 16-22 August 2009 in Durban, South-Africa.

9.2. Involvement with the Scientific Community

- L. Amsaleg:
  - was a program committee member of 2008 ACM SIGMOD International Conference on Management of Data, Providence, Rhode Island, USA;
  - was a program committee member of Bases de données avancées, BDA 2009, Namur, Belgium;
  - was a program committee member of Conférence en Recherche d'Information et Applications, CORIA 2009, Giens, France, May 2009;
  - was a program committee member of the 15th International MultiMedia Modeling Conference, Sophia Antipolis, France;
  - was in the reading committee of the ACM Transactions on Multimedia Computing, Communications and Applications (ACM TOMCCAP), 2009;
  - was the co-organizer of a GRD Isis special day "Passage à l'échelle dans la recherche d'information multimédia", Paris;
  - was an evaluator for the Swiss National Science Foundation, 2009.
  - was an evaluator for the French ANR, 2009.
- V. Claveau:
  - was a program committee member of TALN'09 (16^{e} conférence nationale Traitement automatique des langues naturelles), Senlis, France, June 2009;
  - was a program committee member of ECML-PKDD 09, Bled, Slovenia, Sept. 2009;
  - was a reviewing committee member of ESSLLI, 21st European Summer School in Language, Logic and Information, Bordeaux, France, July 2009;
  - was a program committee member of Conférence en Recherche d'Information et Applications, CORIA 2009, Giens, France, May 2009;
  - was a reviewer for the journal Multimedia Tools and Applications;

  - was an editorial committee member of the special issue on "Opinion Mining" of the RNTI journal, 2009.
  - was a reviewing committee for the journal TAL, Traitement Automatique des Langues.

- P. Gros:
  - was program co-chair of the seventh International Workshop on Content Based Multimedia Indexing (CBMI) which was held in Chania, Greece in July 2009;
  - was program co-chair of the NEM summit "Towards Future Media Internet", Saint Malo, France, September 2009;
  - was program committee member of the Conference on Information Retrieval and Application CORIA'09, Giens, France, May 2009;
  - was a program committee member of the Conference "Compression et représentation des signaux audiovisuels" CORESA'09, Toulouse, March 2009;
  - was a program committee member of the GRETSI conference, Dijon, September 2009;
  - was a program committee member of the International conference on signal processing and multimedia applications SIGMAP'09, Milano, Italy, July 2009;
  - was a program committee member of the First International Conference on Creative Content Technologies CONTENT, Cancun, Mexico, November 2009;
  - was a program committee member of the Conférence en Recherche d'Information et Applications CORIA, Presqu'ile de Giens, France, May 2009;
  - was a program committee member of the 3rd International Conference on Imaging for Crime Detection and Prevention ICDP, London, UK, December 2009;
  - was a program committee member of the 2009 IEEE Pacific-Rim Conference on Multimedia PCM, Bangkok, Thailand, December 2009;
  - was a program committee member of the conference "Traitement et analyse de l'information : Méthodes et Applications" TAIMA, Hammamet, Tunisia, May 2009.

- E. Kijak:
  - was co-organizer and co-chair of the special session "Scalable indexing and retrieval of video content" for CBMI 2009 (7th International Workshop on Content-Based Multimedia Indexing), Chania, Crete, June 2009.

- F. Moreau:
  - was a reviewing committee member for Language Resources and Evaluation (LRE) journal.

- A. Morin:
  - was a program committee member of ITI 2009 (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'09, Visual INformation Communications International, Sydney, Australia, September 2009;
  - was a program committee member of AusDM'09, Australasian Data Mining Conference, Melbourne, Australia, December 2009;
  - was a program committee member of EGC'09, Extraction et Gestion de Connaissances, Strasbourg, France, January 2009;

  - was co-organizer of the 7th workshop Visualisation et Extraction de Connaissances, (AVEC-EGC'09), Strasbourg, January 2009;
  - was a reviewing committee member for IEEE Transactions on Image Processing;
  - was a reviewing committee member for CIVR'09, ACM International Conference on Image and Video Retrieval;
  - was a reviewing committee member for EJOR, European Journal of Operational Research;
  - was a reviewing committee member for CIE 39, International Conference on Computer and Industrial Engineering;
  - was a reviewing committee member for JESA, Journal Européen des Systèmes Automatisés;
  - was a reviewing committee member RIA, Revue d'Intelligence Artificielle.

- P. Sébillot:
  - was a reading committee member of TALN 2009 (16e conférence nationale Traitement automatique des langues naturelles), Senlis, France, June 2009;
  - was a program committee member of DEFT'09 (5e défi fouille de textes), Senlis, France, June 2009;
  - was a reviewer committee member of NACAP 2009 (Networks and Their Philosophical Implications), Indiana University, Bloomington, Indiana, USA, June 2009;
  - was a reading committee member of the special issue Machine learning and NLP of TAL Journal (Traitement automatique des langues);
  - was a program committee member of TIA-2009 (Terminologie et intelligence artificielle), Toulouse, France, November 2009.
  - is an editorial committee member of the Journal TAL (Traitement automatique des langues; since July 2009)

- P. Tirilly:
  - was a program committee member of RJCRI'09 (Rencontres Jeunes Chercheurs en Recherche d'Informations), Giens, France, May 2009.
  - was a reviewing committee member for the journal Multimedia Tools and Applications.

- L. Ughetto:
  - was a program committee member of the Rencontres Francophones sur la Logique Floue et ses Applications (LFA'09), Annecy, France, November 2009.

9.3. Teaching Activities

- L. Amsaleg 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.
- L. Amsaleg: Multidimensional Indexing Techniques, University of Reykjavík, Master.
- L. Amsaleg: High-Dimensional Clustering Techniques, University of Reykjavík, Master.
- 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: Analysis of audiovisual documents and flows for indexing, Master by research in computer science (2^{nd} year), University of Rennes 1.

- E. Kijak and P. Tirilly: Digital Documents Indexing and Retrieval, Professional Master in Computer Science, 2nd year, IFSIC, 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.
- V. Claveau: Symbolic Sequential Data, Master by research in computer science (2nd year), University of Rennes 1.
- F. Poulet is in charge of the Professional Master in computer science (2nd year), MITIC, Computer Science Methods and Information and Communication Technologies, University of Rennes 1.
- F. Poulet: Supervised Learning. Master by research in computer science (2nd year), University of Rennes 1.
- F. Poulet: Introduction to Data Mining. Professionnal Master in Computer Science, 2nd year, IFSIC, University of Rennes 1.
- F. Poulet: Mining Symbolic Data. Professionnal Master in Computer Science, 2nd year, IFSIC, University of Rennes 1.
- F. Poulet: Data Warehouses. Professionnal Master in Computer Science, 2nd year, IFSIC, University of Rennes 1.
- F. Poulet: Applications and Problem Solving. Professionnal Master in Computer Science, 2nd year, IFSIC, University of Rennes 1.
- F. Poulet: Advanced Databases, Graduate Engineering Degree, 3rd year, IFSIC, University of Rennes 1.

9.4. Participation to Seminars, Workshops, Invited Conferences

- P. Sébillot gave an invited talk at INRIA Nancy - Grand Est research center about topic analysis of written and oral corpora, April 2009.

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", vol. 2001, n° 4, March 2001, p. 108-124.

[2] J. ANDRÉ, A. MORIN, H. RICHY. *Comparison of Literary Texts using Biological Sequence Comparison and Structured Documents Capabilities*, in "Proceedings of the International Conference on Computational Linguistics, Speech and Document Processing, ICCLSDP'98, Calcutta, India", February 1998.

[3] 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", vol. 4, August 2003, p. 493–525.

[4] M. DELAKIS, G. GRAVIER, P. GROS. *Multimodal Segmental-Based Modeling of Tennis Video Broadcasts*, in "Proceedings of IEEE International Conference on Multimedia and Expo, Amsterdam, Netherlands", July 2005, http://www.irisa.fr/texmex/publications/versionElect/2005/icme2005_manolis.pdf.

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

[6] H. LEJSEK, F.-H. ÁSMUNDSSON, B.-Þ. JÓNSSON, L. AMSALEG. *Scalability of Local Image Descriptors: A Comparative Study*, in "Proceedings of the 14th ACM International Conference on Multimedia, Santa Barbara, CA, USA", October 2006, http://www.irisa.fr/texmex/publications/versionElect/2006/fp33a-lejsek.pdf.

[7] R. PRIAM, A. MORIN. *Visualisation des corpus textuels par treillis de multinomiales auto-organisées - Généralisation de l'analyse factorielle des correspondances*, in "Revue Extraction des Connaissances et Apprentissage (Actes EGC'02)", vol. 1, n^{o} 4, 2002, p. 407-412.

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

[9] F. TONNIN, P. GROS, C. GUILLEMOT. *Analysis of multiresolution representations for compression and local description of images*, in "Proceedings of the 8th International Conference on Visual Information and Information Systems, VISUAL'05, Amsterdam, The Netherlands", 2005.

Year Publications

Doctoral Dissertations and Habilitation Theses

[10] N.-K. PHAM. *Analyse factorielle des corrspondances pour l'indexation et la recherche d'information dans une grande base de données d'images*, Université de Rennes 1, November 2009, http://www.irisa.fr/centredoc/publis/theses/, Thèse de doctorat.

Articles in International Peer-Reviewed Journal

[11] T.-N. DO, V.-H. NGUYEN, F. POULET. *GPU-based Parallel SVM Algorithm*, in "JFCST, Journal of Frontiers of Computer Science and Technology", vol. 3, n^{o} 4, 2009, p. 368-377 VN .

[12] 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 2009, http://www.irisa.fr/texmex/publications/versionElect/2009/huet-csl-09.pdf.

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

[14] R. TAVENARD, L. AMSALEG, G. GRAVIER. *Model-based similarity estimation of multidimensional temporal sequences*, in "Annals of Telecommunications", vol. 64, Issue 5 (2009), 2009, p. 381-390, http://dx.doi.org/10.1007/s12243-009-0091-4.

Articles in National Peer-Reviewed Journal

[15] F. POULET, T.-N. DO, V.-H. NGUYEN. *SVM incrémental et parallèle sur GPU*, in "RNTI-E-15, Revue des Nouvelles Technologies de l'Information", 2009, p. 103-114 VN .

International Peer-Reviewed Conference/Proceedings

[16] S. BAGHDADI, C.-H. DEMARTY, G. GRAVIER, P. GROS. *Apprentissage de structure dans les réseaux bayésiens pour la détection d'évènement vidéo*, in "Actes de la conférence Traitement et Analyse de l'Information : Méthodes et Applications (TAIMA), Hammamet, Tunisie", may 2009.

[17] P. BOSC, V. CLAVEAU, O. PIVERT, L. UGHETTO. *Graded-Inclusion-Based Information retrieval Systems*, in "Proceedings of European Conference on Information Retrieval (ECIR'09), Toulouse, France", M. BOUGHANEM, C. BERRUT, J. MOTHE, C. SOULE-DUPUY (editors), LNCS 5478, Springer, April 2009, p. 321–336.

[18] P. BOSC, L. UGHETTO, O. PIVERT, V. CLAVEAU. *Implication-Based and Cardinality-Based Inclusions in Information Retrieval*, in "Proceedings of the IEEE International Conference on Fuzzy Systems (Fuzz-IEEE'09), Jeju Island, South Korea", August 2009, p. 2088–2093, http://dx.doi.org/10.1109/FUZZY.2009.5277249.

[19] V. CLAVEAU. *Letter-to-phoneme conversion by inference of rewriting rules*, in "Proceedings of the 10th Annual Conference of the International Speech Communication Association, InterSpeech, Brighton UK", September 2009.

[20] M. DOUZE, H. JÉGOU, H. SINGH, L. AMSALEG, C. SCHMID. *Evaluation of GIST descriptors for web-scale image search*, in "International Conference on Image and Video Retrieval, Island of Santorini, Greece", July 2009.

[21] C. GUINAUDEAU, G. GRAVIER, P. SÉBILLOT. *Can Automatic Speech Transcripts be Used for Large Scale TV Stream Description and Structuring ?*, in "First International Workshop on Content-Based Audio/Video Analysis for Novel TV Services, San Diego, CA, USA", December 2009, http://www.irisa.fr/texmex/publications/versionElect/2009/pascale09.pdf, In conjunction with the International IEEE Symposium on Multimedia.

[22] G. LECORVÉ, G. GRAVIER, P. SÉBILLOT. *Constraint selection for topic-based MDI adaptation of language models*, in "Proceedings of the International Conference on Speech and Language Technology (Interspeech'09), Brighton, UK", September 2009, p. 368–371, http://www.irisa.fr/texmex/publications/versionElect/2009/LeGS2009.pdf.

[23] H. LEJSEK, F. H. ÁSMUNDSSON, K. DAÐASON, B. JÓHANNESSON, Æ. KVARAN, B. Þ. JÓNSSON, L. AMSALEG. *Videntifier Forensic: A New Law Enforcement Service for Automatic Identification of Illegal Video Material*, in "ACM Multimedia 2009 Workshop: Multimedia in Forensics, Beijing, P. R. China", October 2009 IS .

[24] H. LEJSEK, F. H. ÁSMUNDSSON, K. DAÐASON, B. Þ. JÓNSSON, L. AMSALEG. *Videntifier Forensic: Robust and Efficient Detection of Illegal Multimedia, a demonstration*, in "ACM Multimedia, Beijing, P. R. China", October 2009 IS .

[25] N.-K. PHAM, A. MORIN, P. GROS, Q.-T. LE. *Accelerating image retrieval using factorial correspondence analysis on GPU*, in "Proceedings of the 13th International Conference on Computer Analysis of Images and Patterns (CAIP'09), Münster, Germany", vol. 5702, Springer LNCS, 2009, p. 565-572.

[26] N.-K. PHAM, A. MORIN, P. GROS. *Accelerating Image Retrieval Using Factorial Correspondence Analysis on GPU*, in "Proceedings of CAIP", 2009, p. 565-572 VN .

[27] L. UMEK, B. ZUPAN, M. TOPLAK, A. MORIN, J.-H. CHAUCHAT, G. MAKOVEC, D. SMRKE. *Subgroup Discovery in Data Sets with Multi-dimensional Responses: A Method and a Case Study in Traumatology*, in "Proceedings of AIME", 2009, p. 265-274 HR SI .

[28] J. ZEPEDA, E. KIJAK, C. GUILLEMOT. *SIFT-Based Local Image Description using Sparse Representations*, in "Proceedings of the IEEE International Workshop on Multimedia Signal Processing (MMSP09), Rio de Janeiro, Brazil", October 2009.

National Peer-Reviewed Conference/Proceedings

[29] V. CLAVEAU. *La phonétisation comme un problème de translittération*, in "Actes de la conférence Traitement automatique des langues naturelles, TALN'09, Senlis, France", June 2009, short paper.

[30] T.-N. DO, S. LALLICH, N.-K. PHAM, P. LENCA. *Un nouvel algorithme de forêts aléatoires d'arbres obliques particulièrement adapté à la classification de données en grandes dimensions*, in "Actes de la conférence Extraction et Gestion de Connaissances (EGC'09)", 2009, p. 79-90.

[31] T.-N. DO, N.-K. PHAM, F. POULET. *Une méthode anthropocentrée pour la construction d'arbres de décision*, in "Proceedings of 7ème Atelier Visualisation et Extraction de Connaissances, Extraction et Gestion de Connaissances (EGC'09), Strasbourg, Fance", F. POULET, B. LE GRAND, T.-N. DO, M.-A. AUFAURE (editors), January 2009, p. 33-43 VN .

[32] C. GUINAUDEAU. *Recherche d'information textuelle et phonétique pour le contrôle de l'étiquetage automatique d'émissions dans un flux télévisuel*, in "Actes des Rencontres Jeunes Chercheurs en Recherche d'Information (RJCRI'09), Presqu'île de Giens, France", May 2009, http://www.irisa.fr/texmex/publications/versionElect/2009/camille09.pdf.

[33] T.-B. NGUYEN, P. LENCA, T.-N. DO, F. POULET. *Visualisation de réseaux d'experts*, in "Proceedings of 7ème Atelier Visualisation et Extraction de Connaissances, Extraction et Gestion de Connaissances (EGC'09), Strasbourg, France", F. POULET, B. LE GRAND, T.-N. DO, M.-A. AUFAURE (editors), January 2009, p. 1-6 VN .

[34] N.-K. PHAM, A. MORIN, P. GROS, Q.-T. LE. *Utilisation de l'analyse factorielle des correspondances pour la recherche d'images à grande échelle*, in "Actes de la conférence Extraction et Gestion de Connaissances (EGC'09)", 2009, p. 283-294 VN .

[35] L. UGHETTO, O. PIVERT, V. CLAVEAU, P. BOSC. *Recherche d'information et inclusions graduelles*, in "Actes des Journées Francophones sur la Logique Floue et ses Applications (LFA'09), Annecy, France", Cépaduès éditions, November 2009, p. 125–132.

[36] L. UGHETTO, O. PIVERT, V. CLAVEAU, P. BOSC. *SRI à base d'inclusion graduelle*, in "Actes de la Conférence en Recherche d'Infomations et Applications (CORIA'09), Presqu'île de Giens, France", May 2009, p. 235–250.

Scientific Books (or Scientific Book chapters)

[37] V. CLAVEAU. *Translation of Biomedical Terms by Inferring Rewriting Rules*, in "Information Retrieval in Biomedicine: Natural Language Processing for Knowledge Integration", V. PRINCE, M. ROCHE (editors), chap. 6, IGI - Global, 2009.

[38] T.-N. DO, F. POULET. *Kernel-Based Algorithms and Visualization for Interval Data Mining (invited chapter)*, in "Mining Complex Data", D. A. ZIGHED, S. TSUMOTO, Z. W. RAS, H. HACID (editors), Studies in Computational Intelligence, vol. 165, Springer-Verlag, 2009, p. 75-91 VN .

[39] N.-K. PHAM, A. MORIN, P. GROS, Q.-T. LE. *Intensive use of factorial correspondence analysis for large scale content-based image retrieval*, in "Advances in Knowledge Discovery and Management (AKDM'09)", Springer-Verlag, 2009.

[40] D. THAN-NGHI, S. LALLICH, N.-K. PHAM, P. LENCA. *Classifying very-high-dimensional data with random forests of oblique decision trees*, in "Advances in Knowledge Discovery and Management (AKDM'09)", Springer-Verlag, 2009.

Books or Proceedings Editing

[41] F. POULET, B. LE GRAND, T.-N. DO, M.-A. AUFAURE (editors). *Actes du 7ème Atelier Visualisation et Extraction de Connaissances, Extraction et Gestion de Connaissances (EGC'09)*, January 2009.

Research Reports

[42] P. TIRILLY, V. CLAVEAU, P. GROS. *A review of weighting schemes for bag of visual words image retrieval*, n^{o} 1927, IRISA - INRIA Rennes Bretagne Atlantique, April 2009, Technical report.

References in notes

[43] 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.

[44] C. C. AGGARWAL, A. HINNEBURG, D. A. KEIM. *On the surprising behavior of distance metrics in high dimensional space*, in "Lecture Notes in Computer Science", Springer, 2001, p. 420–434.

[45] G. AMATI, C. J. VAN RIJSBERGEN. *Probabilistic models of information retrieval based on measuring the divergence from randomness*, in "ACM Transactions on Information Systems", vol. 20, n^{o} 4, 2002, p. 357–389, http://doi.acm.org/10.1145/582415.582416.

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

[47] 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", vol. 7, n^{o} 5/6, 2002.

[48] C. BÖHM, S. BERCHTOLD, D. KEIM. *Searching in High-dimensional Spaces: Index Structures for Improving the Performance of Multimedia Databases*, in "ACM Computing Surveys", vol. 33, n^{o} 3, September 2001.

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

[50] R. FAGIN, R. KUMAR, D. SIVAKUMAR. *Efficient similarity search and classification via rank aggregation*, in "Proceedings of the ACM SIGMOD Conference, San Diego, CA", 2003.

[51] L. FLORACK, B. TER HAAR ROMENY, J. KOENDERINK, M. VIERGEVER. *General Intensity Transformation and Differential Invariants*, in "Journal of Mathematical Imaging and Vision", vol. 4, n^{o} 2, 1994, p. 171-187.

[52] 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.

[53] A. GUTTMAN. *R-Trees: A Dynamic Index Structure for Spatial Searching*, in "ACM SIGMOD", 1984.

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

[55] S. HUET, G. GRAVIER, P. SÉBILLOT. *Un modèle multi-sources pour la segmentation en sujets de journaux radiophoniques*, in "Actes de la 15ème conférence sur le Traitement Automatique des Langues Naturelles (TALN), Avignon, France", june 2008, p. 49-58, http://www.irisa.fr/texmex/publications/versionElect/2008/HGS08b.pdf.

[56] H. JÉGOU, H. HARZALLAH, C. SCHMID. *A contextual dissimilarity measure for accurate and efficient image search*, in "Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR'07, Minneapolis, Minnesota, USA", June 2007, p. 1–8.

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

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

[59] X. NATUREL. *Structuration automatique de flux vidéos de télévision*, University of Rennes 1, April 2007, http://www.irisa.fr/centredoc/publis/theses/irisafolder.2007-03-01.1633154603/irisapublication.2007-04-26.0496201623, Ph. D. Thesis.

[60] D. NISTÉR, H. STEWÉNIUS. *Scalable recognition with a vocabulary tree*, in "Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR'06, New-York, New-York, USA", vol. 2, June 2006, p. 2161–2168.

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

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

[63] J. ROBINSON. *The K-D-B-Tree: A Search Structure For Large Multidimensional Dynamic Indexes*, in "ACM SIGMOD", 1981.

[64] G. SALTON, C. BUCKLEY. *Term-weighting approaches in automatic text retrieval*, in "Information Processing and Management", vol. 24, n^{o} 5, 1988, p. 513–523.

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

[66] C. SCHMID, R. MOHR. *Local Grayvalue Invariants for Image Retrieval*, in "ieee Transactions on Pattern Analysis and Machine Intelligence", vol. 19, n^{o} 5, May 1997, p. 530–534.

[67] J. SIVIC, B. C. RUSSELL, A. A. EFROS, A. ZISSERMAN, W. T. FREEMAN. *Discovering Object Categories in Image Collections*, in "Proceedings of the International Conference on Computer Vision, Beijing, China", 2005.

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

[69] P. TIRILLY, E. MARTIENNE, V. CLAVEAU, P. GROS. *Association d'un détecteur de visages et d'un détecteur d'entités nommées pour l'annotation automatique d'images*, in "Rencontres Jeunes Chercheurs en Recherche d'Information (RJCRI), Saint-Etienne, France", 2007.

[70] 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.