IN PARTNERSHIP WITH:
CNRS

Institut national des sciences appliquées de Rennes

Université Rennes 1

Activity Report 2013

Project-Team TEXMEX

Multimedia content-based indexing

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

RESEARCH CENTER
Rennes - Bretagne-Atlantique

THEME
Vision, perception and multimedia interpretation

Table of contents

Project-Team TEXMEX

Keywords: Multimedia, Computer Vision, Natural Language, Data Mining, Machine Learning

Creation of the Project-Team: 2002 November 01.

1. Members

Research Scientists

Patrick Gros [Senior Researchr, Team Leader, HdR]
Laurent Amsaleg [Researcher]
Vincent Claveau [Researcher]
Teddy Furon [Researcher]
Guillaume Gravier [Researcher, HdR]
Hervé Jégou [Researcher]

Faculty Members

Philippe-Henri Gosselin [Professor, HdR]
Ewa Kijak [Associate Professor]
Anne Morin [Associate Professor, until November 2013]
François Poulet [Associate Professor, HdR]
Christian Raymond [Associate Professor]
Pascale Sébillot [Professor, HdR]

Engineers

Sébastien Campion
Rachid Benmokhtar [OSEO Anvar QUAERO project, until April 2013]
Jonathan Delhumeau [OSEO Anvar QUAERO project, until Oct 2013]
Carryn Hayward [OSEO Anvar QUAERO project]
Elodie Lequoc [until Dec 2013]
Arthur Masson [OSEO Innovation, from February 2013]
Diana Moise [OSEO Anvar QUAERO, until May 2013]

PhD Students

Raghavendran Balu [Alcatel-Lucent / Inria Common Labs, from October 2013]
Petra Bosilj [CIFRE]
Mohamed-Haykel Boukadida [CIFRE]
Thanh Nghi Doan [until October 2013]
Khaoula Elagouni [CIFRE Orange]
Julien Fayolle [OSEO Anvar Quaero project, until June 2013]
Gylfi Gudmundsson [OSEO Anvar Quaero project, until September 2013]
Mihir Jain [Inria Grant Cordis]
Ludivine Kuznik [CIFRE INA]
Abir Ncibi
Cédric Penet [CIFRE Technicolor, until October 2013]
Bingqing Qu [CIFRE INA]
Anca-Roxana Simon

Post-Doctoral Fellows

Marie Béatrice Arnulphy [OSEO Anvar Quaero project, from February 2013 until September 2013]
Josip Krapac [ANR FIRE-ID project, until June 2013]
Benjamin Mathon [ANR SECULAR project, until August 2013]
Denis Shestakov [OSEO Anvar Quaero project, until April 2013]
Davy Weissenbacher [OSEO Anvar Quaero project, from February 2013 until December 2013]

Wanlei Zhao [OSEO Anvar Quaero project, until December 2013]

Visiting Scientists

Agni Delvinioti [M. Sc. student, National Technical University of Athens (Grèce), from January 2013 until July 2013]
Bruno Do Nascimento Teixeira [Ph. D. student, Univ. Federal Minas Gerais (Brésil), from April 2013 until October 2013]
Kleber Jacques Ferreira de Souza [Ph. D. student, Univ. Federal Minas Gerais (Brésil), from June 2013 until September 2013]
Emmanuelle Martienne [Associate Professor, Univ. Rennes II, until December 2013]
Fabienne Moreau [Associate Professor, Univ. Rennes II, until December 2013]
Nguyen Khang Pham [Associate Professor, Univ. Rennes I, until June 2013]
Michael Rabbat [Professor, McGill University (Canada), from October 2013 until November 2013]
Giorgos Tolias [postdoc student, National Technical University of Athens (Grèce), January and December 2013]
Laurent Ughetto [Associate Professor, Univ. Rennes II, until December 2013]

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

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

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

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

2.1.1. Advanced algorithms of data analysis, description and indexing

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

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

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

2.1.2. New techniques for linguistic information acquisition and use

Natural languages are a privileged way to carry high level semantic information. Used in speech from an audio track, in textual format or overlaid in images or videos, alone or associated with images, graphics or tables, organized linearly or with hyperlinks, 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

- We have won the FGcomp'2013 challenge, in conjunction with Imagenet, for fine-grain classification of images.
- Best paper award to Cédric Penet at Content-Based Multimedia Indexing .

BEST PAPER AWARD :
[51] **Audio Event Detection in Movies using Multiple Audio Words and Contextual Bayesian Networks in CBMI - 11th International Workshop on Content Based Multimedia Indexing - 2013**. C. PENET, C.-H. DEMARTY, G. GRAVIER, P. GROS.

3. Research Program

3.1. Image description

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

The first description methods associate a unique global descriptor with each image, *e.g.*, a color histogram or correlogram, a texture descriptor. Such descriptors are easy to compute and use, but they usually fail to handle cropping and cannot be used for object recognition. The most successful approach to address a large class of transformations relies on the use of local descriptors, extracted on regions of interest detected by a detector, for instance the Harris detector [87] or the Difference of Gaussian method proposed by David Lowe [89].

The detectors select a square, circular or elliptic region that is described in turn by a patch descriptor, usually referred to as a local descriptor. The most established description method, namely the SIFT descriptor [89], was shown robust to geometric and photometric transforms. Each local SIFT descriptor captures the information provided by the gradient directions and intensities in the region of interest in each region of a 4×4 grid, thereby taking into account the spatial organization of the gradient in a region. As a matter of fact, the SIFT descriptor has became a standard for image and video description.

Local descriptors can be used in many applications: image comparison for object recognition, image copy detection, detection of repeats in television streams, etc. While they are very reliable, local descriptors are not without problems. As many descriptors can be computed for a single image, a collection of one million images generates in the order of a billion descriptors. That is why specific indexing techniques are required. The problem of taking full advantage of these strong descriptors on a large scale is still an open and active problem. Most of the recent techniques consists in computing a global descriptor from local ones, such as proposed in the so-called bag-of-visual-word approach [96]. Recently, global description computed from local descriptors has been shown successful in breaking the complexity problem. We are active in designing methods that aggregate local descriptors into a single vector representation without loosing too much of the discriminative power of the descriptors.

3.2. Corpus-based text description and machine learning

Our work on textual material (textual documents, transcriptions of speech documents, captions in images or videos, etc.) is characterized by a chiefly corpus-based approach, as opposed to an introspective one. A corpus is for us a huge collection of textual documents, gathered or used for a precise objective. We thus exploit specialized (abstracts of biomedical articles, computer science texts, etc.) or non specialized (newspapers, broadcast news, etc.) collections for our various studies. In TEXMEX, according to our applications, different kinds of knowledge can be extracted from the textual material. For example, we automatically extract terms characteristic of each successive topic in a corpus with no a priori knowledge; we produce representations for documents in an indexing perspective [95]; 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 [82]. For our various problems involving language processing, we use both approaches, making the most of existing machine learning techniques or proposing new ones. Relying on advantages of both methods, we aim at developing machine learning solutions that are automatic and generic enough to make it possible to extract, from a corpus, the kind of elements required by a given task.

3.3. Stochastic models for multimodal analysis

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

Markov models are composed of a set of states, of transition probabilities between these states and of emission probabilities that provide the probability to emit a given symbol at a given state. Such models allow generating sequences. Starting from an initial state, they iteratively emit a symbol and then switch in a subsequent state according to the respective probability distributions. These models can be used in an indirect way. Given a sequence of symbols (called observations), hidden Markov models (HMMs) [93]) aim at finding the best sequence of states that can explain this sequence. The Viterbi algorithm provides an optimal solution to this problem.

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

Two of the main drawbacks of the HMMs is that they can only emit a unique symbol per state, and that they imply that the duration in a given state follows an exponential distribution. Such drawbacks can be circumvented by segment models [91]. These models are an extension of HMMs were each state can emit several symbols and contains a duration model that governs the number of symbols emitted (or observed) for this state. Such a scheme allows us to process features at different rates.

Bayesian networks are an even more general model family. Static Bayesian networks [85] 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 networks [90] are a generalization of the previous models. Such networks are composed of an elementary network that is replicated at each time stamp. Duration variable can be added in order to provide some flexibility on the time processing, like it was the case with segment models. While HMMs and segment models are well suited for dense segmentation of video streams, Bayesian networks offer better capabilities for sparse event detection. Defining a trash state that corresponds to non event segments is a well known problem is speech recognition: computing the observation probabilities in such a state is very difficult.

3.4. Multidimensional indexing techniques

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

Traditional multidimensional indexing techniques typically group high dimensional features vectors into cells. At querying time, few such cells are selected for searching, which, in turn, provides performance as each cell contains a limited number of vectors [84]. Cell construction strategies can be classified in two broad categories: *data partitioning* indexing methods that divide the data space according to the distribution of data, and *space partitioning* indexing methods that divide the data space along predefined lines and store each descriptor in the appropriate cell.

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

Recently, several approaches make use of quantization operations. This, somehow, transforms costly nearest neighbor searches in multidimensional space into efficient uni-dimensional accesses. One seminal approach, the LSH technique [86], uses a structured scalar quantizer made of projections on segmented random lines, acting as spatial locality sensitive hash-functions. In this approach, several hash functions are used such that co-located vectors are likely to collide in buckets. Other approaches use unstructured quantization schemes, sometimes together with a vector aggregation mechanism [96] to boost performance.

3.5. Data mining methods

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

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

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

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

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

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

4. Application Domains

4.1. Copyright protection of images and videos

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

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

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

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

4.2. Video database management

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

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

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

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

4.3. Textual database management

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

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

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

5. Software and Platforms

5.1. Software

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

5.1.1. New Software

5.1.1.1. DeCP-Index

Participants: Laurent Amsaleg [Correspondent], Gylfi Gudmundsson, Diana Moise, Denis Shestakov.

DeCP-Index is a Map-Reduce oriented implementation of the vectorial quantization scheme developped during the PhD of Gylfi Gudmundsson. It is in Java.

First APP deposit: IDDN.FR.001.500011.000.S.P.2013.000.40000

5.1.1.2. DeCP-Scripts

Participants: Laurent Amsaleg [Correspondent], Gylfi Gudmundsson, Diana Moise, Denis Shestakov.

DeCP-Scripts is a series of script for installing, configuring and deploying the Map Reduce framework over the grid infrastructure.

First APP deposit: IDDN.FR.001.500012.000.S.P.2013.000.40000

*5.1.1.3. *SVM*

Participants: François Poulet [correspondent], Thanh Nghi Doan.

Soon available from the home page of the Inria Parallel Benchmark Suite http://www.irisa.fr/alf/index.php?option=com_content&view=article&id=82&Itemid=&lang=fr.

*SVM include a set of parallel and incremental SVM classifiers for large scale classification tasks on GPU, CPU or cluster / Grid.

5.1.2. Main software started before 2012

5.1.2.1. Peyote

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

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

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

5.1.2.2. Aabot

Participant: Jonathan Delhumeau.

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

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

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

5.1.2.3. Pqcodes

Participant: Hervé Jégou [correspondent].

Jointly maintained with Matthijs Douze, Inria/LEAR.

Pqcodes is a library which implements the approximate k nearest neighbor search method of [88] based on product quantization. This software has been transferred to two companies (in August 2011 and May 2012, respectively).

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

5.1.2.4. Yael

Participant: Hervé Jégou [correspondent].

Jointly maintained with Matthijs Douze, from Inria/LEAR.

Yael is a C/python/Matlab library providing (multi-threaded, Blas/Lapack, low level optimization) implementations of computationally demanding functions. In particular, it provides very optimized functions for k-means clustering and exact nearest neighbor search. The library has been downloaded about 2,000 times in 2013.

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

5.1.2.5. BonzaiBoost

Participant: Christian Raymond [correspondent].

Available at http://bonzaiboost.gforge.inria.fr/.

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

5.1.2.6. Irisa_Ne

Participant: Christian Raymond [correspondent].

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

5.1.2.7. IRISA News Topic Segmenter (irints)

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

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

5.1.3. Other softwares

- BAG OF COLORS: describe images based on color
- I-DESCRIPTION: IDDN.FR.001.270047.000.S.P.2003.000.21000
- ASARES: symbolic machine learning system to infer corpus-specific morpho-syntactic and semantic patterns from descriptions of pairs of linguistic elements found in a corpus in which the components are linked by a given semantic relation IDDN.FR.001.0032.000.S.C.2005.000.20900
- ANAMORPHO: detects morphological relations between words in many languages IDDN.FR.001.050022.000.S.P.2008.000.20900
- DIVATEX: audio/video frame server IDDN.FR.001.320006.000.S.P.2006.000.40000
- NAVITEX: video annotation tool IDDN.FR.001.190034.000.S.P.2007.000.40000
- TELEMEX: web service that enables TV and radio stream recording
- VIDSIG: small and robust video signature (64 bits per image)
- VIDSEG: multimodal video segmentation IDDN.FR.001.250009.000.S.P.2009.000.40000
- ISEC: web application used as graphical interface for content-based image search engines
- GPU-KMEANS: k-means algorithm on GPU
- CORRESPONDENCE ANALYSIS: factorial correspondence analysis (FCA) for image retrieval.
- GPU CORRESPONDENCE ANALYSIS: GPU implementation of CORRESPONDENCE ANALYSIS
- CAVIZ: interactive graphical tool to display and extract knowledge from the results of a CORRESPONDENCE ANALYSIS on images
- KIWI: keyword extraction from texts and ASR transcripts
- TOPIC SEGMENTER: topic segmentation of texts and ASR transcripts.
- S2E: automatic discovery of audiovisual structuring events in videos.
- 2PAC: builds classes of words of similar meanings (“semantic classes”) IDDN.FR.001.470028.000.S.P.2006.000.40000
- FAESTOS: Fully Automatic Extraction of Sets of keywords for TOpic characterization and Spotting IDDN.FR.001.470029.000.S.P.2006.000.40000
- FISHNET: automatic Web pages grabber associated with a specific theme
- MATCH MAKER: semantic relation extraction by statistical methods.
- IRISAPHON: grapheme to phoneme conversion

- PYTHON-GEOHASH: implementation of the geometric hashing algorithm [99]
- AVSST: automatic video stream structuring tool (detection of repetitions, classification program/inter-program, EPG alignment) with GUI
- TVSEARCH: content-based retrieval search engine to search and propagate manual annotation such as advertisement in a TV corpora.
- SAMUSA: multimedia content speech/music segmentation
- KERTRACK: visual graphical interface for tracking visual targets based on particle filter tracking or mean-shift.
- MOZAIC2D: spatio-temporal mosaic based on dominant motion compensation.
- BABAZ: audio database management system with an audio-based search function IDDN.FR.001.010006.000.S.P.2012.000.10000
- PIMPY: Python module and binders for multimedia content indexing

5.2. Demonstration: Texmix

Participants: Sébastien Campion [correspondent], Guillaume Gravier.

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

In 2013, we extended the Texmix demonstration to include transcript-free summarization using word discovery.

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

5.3. Experimental platform

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

Until 2005, we used various computers to store data and to carry out experiments. In 2005, we began work to specify and set-up dedicated equipment to experiment on very large collections of data. During 2006 and 2007, we specified, bought and installed our first complete platform. It is organized around a very large storage capacity (155TB), and contains 4 acquisition devices (for Digital Terrestrial TV), 3 video servers, and 15 computing servers partially included in the local cluster architecture (IGRIDA). A dedicated website has been developed in 2009 to provide a user support. It contains useful information such as references of available and ready to use software on the cluster, list of corpus stored on the platform, pages for monitoring disk space consumption and cluster loading, tutorials for best practices and cookbooks for treatments of large datasets. In 2010, we have acquired a new large memory server with 144GB of RAM which is used for memory demanding tasks. The previous server dedicated to this kind of jobs (acquired in 2008) has been upgraded to 96GB of RAM. In 2012, we extended our storage capacity to 215TB and expanded our computing resources with two new large memory servers with 256GB of RAM for each of them. Both have their own HPC storage of 12TB. This year our backbone network was fully upgraded in order to connect each element of the platform with a 10GB/s bandwith.

A new distributed file system architecture was design and will be implement in 2014.

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

5.4. Web services

Participant: Sébastien Campion [correspondent].

This year after a first prototyping of web service where each one of our algorithm was deployed on it's own server, we decided to develop a second version more centralized and named AllGo. AllGo was designed, developed and deployed in order to save resources unnecessarily locked and painful maintenance tasks.

Available at http://allgo.irisa.fr, AllGo currently host five TexMex web services (Samusa, Otis, Termex, Nero, VidSeg).

AllGo infrastructure is based on the Ruby On Rails (ROR) framework for the web "frontoffice" part. ROR enable to create and run task with an HTML or XML, JSON API. SideKiq schedule each job on several nodes. Finally, thanks to the new linux container technology named Docker, applications are configured and deployed on agnostic nodes, inside their container. Container must be seen as very light virtual machine. All our application are stored in a private registry. Data are shared with the NFS protocol. A automation software named Puppet manage infrastructure throughout its lifecycle, from provisioning and configuration to orchestration and reporting.

6. New Results

6.1. Description of multimedia content

6.1.1. Multiscale image representations with component trees

Participants: Petra Bosilj, Ewa Kijak.

Joint work with Sébastien Lefevre, IRISA/SEASIDE, France.

The goal of this work is to study deeply the use of component trees, which aim at representing an image by the regions it contains at various scales through a tree-based structure, and their ability in the context of content-based image indexing and retrieval. Their invariance properties and their robustness to noise have motivated recent work in image indexing [83], [97], [98], but their usage in this field stays limited. The first part of this work was mainly dedicated to the study of various existing hierarchical representations. This leads to the presentation of a technique that arranges the elements of hierarchical representations of images according to a coarseness attribute [24]. The transformation is similar to filtering a hierarchy with a non-increasing attribute, and includes the results of multiple simple filterings with an increasing attribute. The transformed hierarchy can be then used for search space reduction prior to the image analysis process because it allows for direct access to the hierarchy elements at the same scale or a narrow range of scales.

6.1.2. Image representation

Participants: Rachid Benmokhtar, Jonathan Delhumeau, Guillaume Gravier, Philippe-Henri Gosselin, Hervé Jégou, Wanlei Zhao.

Partially in collaboration with Patrick Pérez, Technicolor, France.

Recent work on image retrieval have proposed to index images by compact representations encoding powerful local descriptors, such as the closely related vector of aggregated local descriptors (VLAD) and Fisher vector (FV). By combining them with a suitable coding technique, it is possible to encode an image in a few dozen bytes while achieving excellent retrieval results. We have pursed the research on this line of research by proposing two complementary contributions.

In [30], we revisited some assumptions proposed in this context regarding the handling of "visual burstiness", and shows that ad-hoc choices are implicitly done which are not desirable. Focusing on VLAD without loss of generality, we propose to modify several steps of the original design. Albeit simple, these modifications significantly improve VLAD and make it compare favorably against the state of the art.

In [65], we proposed a pooling strategy for local descriptors to produce a vector representation that is orientation-invariant yet implicitly incorporates the relative angles between features measured by their dominant orientation. This pooling is associated with a similarity metric that ensures that all the features have undergone a comparable rotation. This approach is especially effective when combined with dense oriented features, in contrast to existing methods that either rely on oriented features extracted on key points or on non-oriented dense features. The interest of our approach in a retrieval scenario is demonstrated on popular benchmarks comprising up to 1 million database images.

In [22], we propose to reduce the dimensionality of visual features for image categorization. We iteratively select sets of projections from an external dataset, using Bagging and feature selection thanks to SVM normals. Features are selected using weights of SVM normals in orthogonalized sets of projections. The bagging strategy is employed to improve the results and provide more stable selection. The overall algorithm linearly scales with the size of features, and is thus able to process large state-of-the-art image representations. Given Spatial Fisher Vectors as input, our method consistently improves the classification accuracy for smaller vector dimensionality, as demonstrated by our results on the popular and challenging PASCAL VOC 2007 benchmark.

6.1.3. Video classification

Participants: Kleber Jacques Ferreira de Souza, Guillaume Gravier, Philippe-Henri Gosselin.

In collaboration with Silvio Jamil F. Guimarães, PUC Minas, Brazil.

Most current motion descriptors for video classification are based on simple video segments, such as rectangular space-time blocks, or more recently rectangular space blocks that follow local trajectories. The aim of this study is to consider more complex video segments that better fit space-time elements of videos, thanks to recent methods for video segmentation proposed by S. Guimarães et al. These methods combine at the same time a fast extraction and stable regions, two essential properties for video indexing. The computation of local motion descriptors on these video segments lead to better video classification for human action recognition, when compared to current video indexing techniques.

6.1.4. Geo-localization of videos with multi-modality

Participants: Jonathan Delhumeau, Guillaume Gravier, Hervé Jégou.

Joint work with Michele Trevisiol, Yahoo! Labs, Spain, who visited the team in 2012.

Geotagging is the process of automatically adding geographical identification metadata to media objects, in particular to images and videos. In [63], we present a strategy to identify the geographic location of videos. First, it relies on a multi-modal cascade pipeline that exploits the available sources of information, namely the user upload history, his social network and a visual-based matching technique. Second, we present a novel divide & conquer strategy to better exploit the tags associated with the input video. It pre-selects one or several geographic area of interest of higher expected relevance and performs a deeper analysis inside the selected area(s) to return the coordinates most likely to be related to the input tags. The experiments were conducted as part of the MediaEval 2012 Placing Task, where we obtained the best results among the competitors when using no external information, i.e. not using any gazetteers nor any other kind of external information.

6.1.5. Violent keysound detection with audio words and Bayesian networks

Participants: Guillaume Gravier, Patrick Gros, Cédric Penet.

Joint work with Claire-Hélène Demarty, Technicolor, France.

We investigated a novel use of the well known audio words representations to detect specific audio events, namely gunshots and explosions, in order to get more robustness towards soundtrack variability in Hollywood movies [51]. An audio stream is processed as a sequence of stationary segments. Each segment is described by one or several audio words obtained by applying product quantization to standard features. Such a representation using multiple audio words constructed via product quantisation is one of the novelties described in this work. Based on this representation, Bayesian networks are used to exploit the contextual

information in order to detect audio events. Experiments are performed on a comprehensive set of 15 movies, made publicly available. Results are comparable to the state of the art results obtained on the same dataset but show increased robustness to decision thresholds, however limiting the range of possible operating points in some conditions. Late fusion provides a solution to this issue.

6.2. Large scale indexing and classification

6.2.1. Parallelism and distribution for very large scale content-based image retrieval

Participants: Gylfi Gudmundsson, Diana Moise, Denis Shestakov, Laurent Amsaleg.

Two observations drove the design of the high-dimensional indexing technique developed in the framework of the Ph. D. thesis of Gylfli Gudmundson. Firstly, the collections are so huge, typically several terabytes, that they must be kept on secondary storage. Addressing disk related issues is thus central to our work. Secondly, all CPUs are now multi-core and clusters of machines are a commonplace. Parallelism and distribution are both key for fast indexing and high-throughput batch-oriented searching.

We developed a high-dimensional indexing technique called eCP. Its design includes the constraints associated to using disks, parallelism and distribution. At its core is an non-iterative unstructured vectorial quantization scheme. eCP builds on an existing indexing scheme that is main memory oriented. The first contribution in eCP is a set of extensions for processing very large data collections, reducing indexing costs and best using disks. The second contribution proposes multi-threaded algorithms for both building and searching, harnessing the power of multi-core processors. Datasets for evaluation contain about 25 million images or over 8 billion SIFT descriptors. The third contribution addresses distributed computing. We adapt eCP to the MapReduce programming model and use the Hadoop framework and HDFS for our experiments. This time we evaluate eCP's ability to scale-up with a collection of 100 million images, more than 30 billion SIFT descriptors, and its ability to scale-out by running experiments on more than 100 machines.

6.2.2. Contributions in image indexing

Participants: Hervé Jégou, Giorgos Tolias.

Partially in collaboration with Yannis Avrithis, National Technical University of Athens, Greece, Cai-Zhi Zhu and Shin'ichi Satoh, Japanese National Institute of Informatics, Japan.

In [62], we have considered a framework and its associated family of metrics to compare images based on their local descriptors. It encompasses the VLAD descriptor and matching techniques such as Hamming embedding. Making the bridge between these approaches leads us to propose a match kernel that takes the best of existing techniques by combining an aggregation procedure with a selective match kernel. Finally, the representation underpinning this kernel is approximated, providing a large scale image search both precise and scalable, as shown by our experiments on several benchmarks. We give a Matlab package associated with the paper that allows to reproduce the results of the most interesting variant.

On the same topic, we propose in [78] a query expansion technique for image search that is faster and more precise than the existing ones. An enriched representation of the query is obtained by exploiting the binary representation offered by the Hamming embedding image matching approach: The initial local descriptors are refined by aggregating those of the database, while new descriptors are produced from the images that are deemed relevant. This approach has two computational advantages over other query expansion techniques. First, the size of the enriched representation is comparable to that of the initial query. Second, the technique is effective even without using any geometry, in which case searching a database comprising 105k images typically takes 79 ms on a desktop machine. Overall, our technique significantly outperforms the visual query expansion state of the art on popular benchmarks. It is also the first query expansion technique shown effective on the UKB benchmark, which has few relevant images per query.

Finally, in [67] we considered a problem related to object retrieval, where we aim at retrieving, from a collection of images, all those in which a given query object appears. This problem is inherently asymmetric: the query object is mostly included in the database image, while the converse is not necessarily true. However, existing approaches mostly compare the images with symmetrical measures, without considering the different roles of query and database. This paper first measures the extent of asymmetry on large-scale public datasets reflecting this task. Considering the standard bag-of-words representation, we then propose new asymmetrical dissimilarities accounting for the different inlier ratios associated with query and database images. These asymmetrical measures depend on the query, yet they are compatible with an inverted file structure, without noticeably impacting search efficiency. Our experiments show the benefit of our approach, and show that the visual object retrieval task is better treated asymmetrically, in the spirit of state-of-the-art text retrieval.

6.2.3. Outlier detection applied to content-based image retrieval

Participants: Teddy Furon, Hervé Jégou.

The primary target of content based image retrieval is to return a list of images that are the most similar to a query image, which is usually done by ordering the images based on a similarity score. In most state-of-the-art systems, the magnitude of this score is very different from one query to another. This prevents us from making a proper decision about the correctness of the returned images. Our work [74] considers the applications where a confidence measurement is required, such as in copy detection or when a re-ranking stage is applied on a short-list such as geometrical verification. For this purpose, we formulate image search as an outlier detection problem, and propose a framework derived from extreme values theory. We translate the raw similarity score returned by the system into a relevance score related to the probability that a raw score deviates from the estimated model of scores of random images. The method produces a relevance score which is normalized in the sense that it is more consistent across queries. Experiments performed on several popular image retrieval benchmarks and state-of-the-art image representations show the interest of our approach.

6.2.4. Exploiting motion characteristics for action classification in videos

Participants: Mihir Jain, Hervé Jégou.

In collaboration with Patrick Bouthemy, Inria/Serpico, France.

Several recent studies on action recognition have attested the importance of explicitly integrating motion characteristics in video description. In this work [43], we have re-visited the use of motion in videos, in order to better exploit it and improve action recognition systems. First, we established that adequately decomposing visual motion into dominant and residual motions, both in the extraction of the space-time trajectories and for the computation of descriptors, significantly improves action recognition algorithms. Then, we designed a new motion descriptor, the DCS descriptor, based on differential motion scalar quantities, divergence, curl and shear features. It captures additional information on the local motion patterns enhancing results. Finally, applying the recent VLAD coding technique proposed in image retrieval provides a substantial improvement for action recognition. Our three contributions are complementary and lead to significantly outperform all reported results on three challenging datasets, namely Hollywood 2, HMDB51 and Olympic Sports.

6.2.5. Recognizing events in videos

Participant: Hervé Jégou.

In collaboration with Matthijs Douze, Jérôme Revaud and Cordelia Schmid, Inria/LEAR, France.

We have addressed the problem of event retrieval for large-scale video collection. Given a video clip of a specific event, e.g., the wedding of Prince William and Kate Middleton, the goal is to retrieve other videos representing the same event from a dataset of over 100k videos.

Our first approach [55] encodes the frame descriptors of a video to jointly represent their appearance and temporal order. It exploits the properties of circulant matrices to compare the videos in the frequency domain. This offers a significant gain in complexity and accurately localizes the matching parts of videos. Furthermore, we extend product quantization to complex vectors in order to compress our descriptors, and to compare them in the compressed domain. Our method outperforms the state of the art both in search quality and query time on two large-scale video benchmarks for copy detection, Trecvid and CCweb. The evaluation has also been done on a new challenging dataset for event retrieval that we introduce: EVVE.

In a subsequent paper [39], we have made two other contributions to event retrieval in large collections of videos. First, we propose hyper-pooling strategies that encode the frame descriptors into a representation of the video sequence in a stable manner. Our best choices compare favorably with regular pooling techniques based on k-means quantization. Second, we introduce a technique to improve the ranking. It can be interpreted either as a query expansion method or as a similarity adaptation based on the local context of the query video descriptor. Experiments on public benchmarks show that our methods are complementary and improve event retrieval results, without sacrificing efficiency.

6.2.6. *Large-scale SVM image classification*

Participants: Thanh Nghi Doan, François Poulet.

Visual recognition remains an extremely challenging problem in computer vision research. Large datasets with millions images for thousands categories poses more challenges. We extend the state-of-the-art large scale linear classifier LIBLINEAR SVM and nonlinear classifier Power Mean SVM in two ways. The first one is to build a balanced bagging classifier with sampling strategy. The second one is to parallelize the training process of all binary classifiers with several multi-core computers [35]. We also applied the same approach to the stochastic gradient descent support vector machines (SVM-SGD) and to both state-of-the-art large linear classifier LIBLINEAR-CDBLOCK and nonlinear classifier Power Mean SVM in an incremental and parallel way [36].

6.2.7. *Video copy detection with SNAP, a DNA indexing algorithm*

Participants: Laurent Amsaleg, Guillaume Gravier.

In collaboration with Leonardo S. De Oliveira, Zenilton Kleber G. Do Patrocínio Jr. and Silvio Jamil F. Guimarães, PUC Minas, Brazil.

Near-duplicate video sequence identification consists in identifying real positions of a specific video clip in a video stream stored in a database. To address this problem, we proposed a new approach based on a scalable sequence aligner borrowed from proteomics [79]. Sequence alignment is performed on symbolic representations of features extracted from the input videos, based on an algorithm originally applied to bioinformatics. Experimental results demonstrate that our method performance achieved 94 % recall with 100 % precision, with an average searching time of about 1 second.

6.3. Security of multimedia contents and applications

6.3.1. *Approximate nearest neighbors search with security and privacy requirements*

Participants: Benjamin Mathon, Laurent Amsaleg, Teddy Furon.

In collaboration with Julien Bringer, Morpho, France.

This work presents a moderately secure but highly scalable and fast approximate nearest neighbors search. Our philosophy is to start from a state-of-the-art technique in this field based on approximate metrics: Euclidean distance based search in [47], [70], and cosine similarity based search in [42]. We then analyze the threats, and patch them avoiding as much as possible bricks penalizing too much the scalability and the speed. On the other hand, we do not completely prevent the players to infer some knowledge, but these limitations are well explained and experimentally assessed. The experimental body uses database of size much bigger than what the past secure solutions can handle.

6.3.2. A privacy-preserving framework for large-scale content-based information retrieval

Participants: Ewa Kijak, Laurent Amsaleg, Teddy Furon.

In close cooperation with Stéphane Marchand-Maillet, Li Weng and April Morton, University of Geneva, Switzerland.

We propose a privacy protection framework for large-scale content-based information retrieval. It offers two layers of protection. First, robust hash values are used as queries instead of original content or features. Second, the client can choose to omit certain bits in a hash value to further increase the ambiguity for the server. Due to the reduced information, it is computationally difficult for the server to know the client's interest. The server has to return the hash values of all possible candidates to the client. The client performs a search within the candidate list to find the best match. Since only hash values are exchanged between the client and the server, the privacy of both parties is protected.

We introduce the concept of *tunable privacy*, where the privacy protection level can be adjusted according to a policy. It is realized through hash-based piece-wise inverted indexing. The idea is to divide a feature vector into pieces and index each piece with a sub-hash value. Each sub-hash value is associated with an inverted index list.

The framework has been extensively tested using a large image database. We have evaluated both retrieval performance and privacy-preserving performance for a particular content identification application. Two different constructions of robust hash algorithms are used. One is based on random projections; the other is based on the discrete wavelet transform. Both algorithms exhibit satisfactory performance in comparison with state-of-the-art reference schemes. The results show that the privacy enhancement slightly improves the retrieval performance.

We consider the *majority voting attack* for estimating the query category and ID. Experiment results show that this attack is a threat when there are near-duplicates, but the success rate decreases with the number of omitted bits and the number of distinct items.

6.3.3. Privacy preserving data aggregation and service personalization using highly-scalable indexing techniques

Participants: Raghavendran Balu, Laurent Amsaleg, Hervé Jégou, Teddy Furon.

In collaboration with Armen Aghasaryan, Dimitre Davidov and Makram Bouzid, Alcatel-Lucent, and Sébastien Gambs, Inria/CIDRE, in the framework of the Alcaltel-Lucent / Inria common Lab.

A challenging approach to the problem of privacy preserving data aggregation and service personalization has recently been proposed in Bell Labs, which introduces a privacy-preserving intermediation layer between end-users and service providers. It uses a distributed variant of a Locality Sensitive Hashing (LSH) techniques of doing scalable nearest-neighbor search, adapted in a novel way, to discover similar users while preserving their privacy. This approach faces however several important challenges that will be targeted in the scope of this collaboration. The challenges are:

- *LSH optimization:* Definitions of hash functions as well as various LSH parameters need to be automatically tuned in order to achieve a good quality of generated recommendations with an expected level of the procured user anonymity. An interesting issue is the possibility of supervised machine learning. If some public profiles are available, more efficient clustering methods boost the quality of the recommendation service but their levels of anonymity have never been assessed so far.
- *Irreversibility of anonymization:* This needs to be evaluated for different attack models, e.g. exploiting the knowledge of LSH hashing functions or any other publically available information on users. It is equivalent as being able to define the region of the super high-dimensional space mapped into the same hashing results. This attack is bound to fail as this region is too large to leak information. However, the prior knowledge about the sparseness of the profiles might drastically reduce this region, and hence weaken the privacy.

- *System dynamics:* Dealing with the cold-start problem or controlling the dynamics of a running system when the profiles and the cluster assignments evolve over the time is yet another challenge this approach is confronted with. If these temporal issues are well studied in conventional relational databases, no clear solution is efficient in the recommendation area, and a fortiori in privacy enhancing recommendation systems.

6.4. Structuring multimedia content and summarization

6.4.1. Stream labeling for TV Structuring

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

In this application, we focus on the problem of labeling the segments of TV streams according to their types (eg. programs, commercial breaks, sponsoring...). During this year, following the work initiated in 2012, we have proposed an in-depth analysis of the use of conditional random fields (CRF) for our task [50]. Through several experiments conducted on real TV streams, we have shown that the CRF yields high results compared with state-of-the-art approaches. In particular, CRF offers several ways to efficiently take the sequenciality of our stream labeling problem into account. We also showed that it is robust when dealing with few training data or few features.

6.4.2. Statistical tests for repetition detection in TV streams

Participant: Patrick Gros.

Detecting all repeated sequences in a TV stream is the first step of all techniques of TV stream structuring. We have improved our technique in several ways. First, a statistical hypothesis test with a corrected risk of Bonferroni was used to clean the repetitions of small sequences. Second, a content-based test is used to clean the remaining sequences, but also to complete the repeated sequences to their maximal length. One of our objective is to reduce the number of descriptor needed to achieve this task, given that this computation is the most expensive of the method. As a matter of fact, the method required computing the descriptors of 15.4 % of the images only.

6.4.3. Video summarization with constraint programming

Participants: Mohamed-Haykel Boukadida, Patrick Gros.

Joint work with Sid-Ahmed Berrani, Orange labs.

Up to now, most video summarization methods are based on concepts like saliency and often use a single modality. In order to develop a more general framework, we propose to use a constraint programming approach, where summarizing a video is seen as a constraint resolution problem, which consists in choosing certain excerpts with respect to various criteria. This year we studied several ways to model the problem in order to gain a maximum flexibility in the summary. A first model was based on the selection of shots, the second one on the selection of parts of shots; The third one does not relies on shots and select image sequences directly. The challenge is to express the useful constraints with these models and the limited possibilities of the solver.

6.4.4. Transcript-free spoken content summarization using motif discovery

Participants: Sébastien Campion, Guillaume Gravier.

Joint work with Frédéric Bimbot and Nathan Souviráa-Labastié, Inria/PANAMA, France.

Exploiting previous results on the unsupervised discovery of repeating words in speech signals, we proposed a method dedicated to transcript-free spoken content summarization. Extractive summarization is performed by selecting a small number of segments, typically one or two, which contains most of the repeated fragments [77]. Audio summaries were included in the Texmix demonstration and are currently being evaluated.

6.4.5. TV program structure discovery using grammatical inference

Participants: Guillaume Gravier, Bingqing Qu.

Joint work with Félicien Vallet and Jean Carrive, Institut National de l'Audiovisuel.

Video structuring, in particular applied to TV programs which have strong editing structures, mostly relies on supervised approaches either to retrieve a known structure for which a model has been obtained or to detect key elements from which a known structure is inferred. We investigated an unsupervised approach to recurrent TV program structuring, exploiting the repetitiveness of key structural elements across episodes of the same show. We cast the problem of structure discovery as a grammatical inference problem and show that a suited symbolic representation can be obtained by filtering generic events based on their reoccurring property [92]. The method follows three steps: *i*) generic event detection, *ii*) selection of events relevant to the structure and *iii*) grammatical inference from a symbolic representation. Experimental evaluation is performed on three types of shows, viz., game shows, news and magazines, demonstrating that grammatical inference can be used to discover the structure of recurrent programs with very limited supervision.

6.4.6. Discovering and linking related images in large collections

Participants: Guillaume Gravier, Hervé Jégou, Wanlei Zhao.

We have tackled the problem of image linking. One of the most successful method to link all similar images within a large collection is min-Hash, which is a way to significantly speed-up the comparison of images when the underlying image representation is bag-of-words. However, the quantization step of min-Hash introduces important information loss. In [66], we proposed a generalization of min-Hash, called Sim-min-Hash, to compare sets of real-valued vectors. We demonstrated the effectiveness of our approach when combined with the Hamming embedding similarity. Experiments on large-scale popular benchmarks demonstrated that Sim-min-Hash is more accurate and faster than min-Hash for similar image search. Linking a collection of one million images described by 2 billion local descriptors is done in 7 minutes on a single core machine.

6.5. Natural language processing in multimedia data

6.5.1. Text detection in videos

Participants: Khaoula Elagouni, Pascale Sébillot.

Texts embedded in multimedia documents often provide high level semantic clues that can be used in several applications or services. We thus aim at designing efficient Optical Character Recognition (OCR) systems able to recognize these texts. During the last three years, we have proposed three novel approaches, robust to text variability (different fonts, colors, sizes, etc.) and acquisition conditions (complex background, non-uniform lighting, low resolution, etc.). The first approach relies on a segmentation step and computes nonlinear separations between characters well adapted to the local morphology of images. The two other ones, called segmentation-free approaches, avoid the segmentation step by integrating a multi-scale scanning scheme: The first one relies on a graph model, while the second one uses a particular connectionist recurrent model able to handle spatial constraints between characters. In 2013, a precise evaluation and comparison between these approaches was conducted and published in [16].

6.5.2. Combining lexical cohesion and disruption for topic segmentation

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

Topic segmentation classically relies on one of two criteria, either finding areas with coherent vocabulary use or detecting discontinuities. We proposed a segmentation criterion combining both lexical cohesion and disruption, enabling a trade-off between the two [58]. We provide the mathematical formulation of the criterion and an efficient graph based decoding algorithm for topic segmentation. Experimental results on standard textual data sets and on a more challenging corpus of automatically transcribed broadcast news shows demonstrate the benefit of such a combination. Gains were observed in all conditions, with segments of either regular or varying length and abrupt or smooth topic shifts. Long segments benefit more than short segments.However the algorithm has proven robust on automatic transcripts with short segments and limited vocabulary reoccurrences.

6.5.2.1. Information extraction and text mining

Participants: Vincent Claveau, Marie Béatrice Arnulphy.

Following the work initiated in the previous period, we have kept on working on relation extraction. During this year, we have proposed a new prototype that still relies on a supervised machine learning approach but we now rely on the sequence built from the shortest syntactic path between the entities, as it is done in many studies. These paths of lemmas are then used in a kNN whose similarity score is based on language modeling techniques. Based on this new prototype, we have participated to several tracks of the BioNLP challenges concerning the automatic extraction of relations in a specialized corpus. Results obtained with this simple and non-domain specific technique were relatively good, with a second and fourth ranks among the participants for the two tasks concerned [26].

We also pursued previous work on supervised techniques for entity extraction and classification. Instead of working on complex machine learning approaches, we rather use simple methods but the focus is set on clever similarity computing between training examples and candidates for which we make the most of existing information retrieval techniques. Our approach has been evaluated through our participation to BioNLP-ST13 competition, where it has been ranked first [26].

We have also proposed unsupervised techniques for knowledge discovery, more precisely, to bring out coherent groups of entities. Existing techniques are usually based on clustering; the challenge is then to define a notion of similarity between the relevant entities. In this work, we have proposed to divert conditional random fields (CRF) in order to calculate indirectly the similarities among text sequences. Our approach consists in generating artificial labeling problems on the data to be processed to reveal regularities in the labeling of the entities. The good results obtained shows the validity of our approach [27] and opens many research avenues for other knowledge discovery tasks.

6.5.3. *Unsupervised approaches to fine-grained morphological analysis*

Participants: Vincent Claveau, Ewa Kijak.

Following the work initiated in the previous years, we have kept on studying fine-grained morphological analysis for biomedical information retrieval. In the biomedical field, the key to access information is the use of specialized terms (like *photochemotherapy*). These complex morphological structures may prevent a user querying for *gastrodynia* to retrieve texts containing *stomachalgia*. The original unsupervised technique proposed in 2012 has been further developed and tested. In particular, during this year, we have shown that it largely outperforms state-of-the-art tools (*e.g.*, Morfessor and Derif) for morphological segmentation tasks. It also offers indirect morpho-lexical resources that are more reliable than hand-coded ones used in most state-of-the-art tools [11].

6.5.4. *Tree-structured named entities recognition*

Participants: Christian Raymond, Davy Weissenbacher.

Many natural language processing tasks needs the production of tree-structured outputs, like syntactic parsing, named entities recognition or language understanding. Currently, only machine learning based systems are robust enough to process the raw and noisy automatic transcribed speech while no machine learning paradigm are able to learn directly the tree structure in a reasonable time. In this work, we studied a solution to tackle the problem of predicting tree structured named entities from speech contents. We investigate a fast and robust decomposition strategy that was implemented and ranked best at the ETAPE NER evaluation campaign with results far better than those of the other participant systems [54].

6.5.5. *Fast machine learning algorithm for efficient combination of various features*

Participant: Christian Raymond.

Currently, in the field of natural language processing the machine learning algorithm "boosting over decision stumps" is often designed as the best off-the-shell classifier. It's actually widely used for his abilities to work on relatively big dataset, to operate intrinsically feature selection and to produce very good decision rules. We investigated a slight modification of this algorithm where the decision stumps are replaced by bonsai trees. Bonsai trees are small decision trees (with low depth) that can capture some structure in the data that decision stumps can not. This modification allows the boosting algorithm to exhibits better (or in the worst case similar) performances with a lower number of iteration the original algorithm needs. Thus allows in some cases a big improvement in term of performance for a lower cost in term of learning time. An application on image processing (typed/hand classification) exhibited interesting results in [94]

6.6. Competitions and international evaluation benchmarks

6.6.1. FGcomp'2013, in conjunction with Imagenet

Participants: Philippe-Henri Gosselin, Hervé Jégou.

Joint participation with Naila Murray and Florent Perronnin, Xerox Research Center Europe.

We have participated the the FGCOMP'2013 challenge and obtained the best results among all participants, see http://sites.google.com/site/fgcomp2013 Although the proposed system follows most of the standard Fisher classification pipeline, we have evaluated and used several key features and good practices that improve the accuracy when specifically considering fine-grained classification tasks [75]. In particular, we consider the late fusion of two systems both based on Fisher vectors, but that employ drastically different design choices that make them very complementary. Moreover, we show that a simple yet effective filtering strategy significantly boosts the performance for several class domains. The method is described in a technical report.

6.6.2. Hyperlink generation in broadcast videos

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

Joint participation with Camille Guinaudeau, Heidelberg Institute of Technology (currently LIMSI-CNRS).

Following up on our 2012 participation, we participated in the Search and hyperlinking task implemented in the framework of the Mediaeval 2013 benchmark initiative. We limited ourselves to hyperlink generation, building on research results in natural language processing, information retrieval and topic segmentation, focusing our contribution on the selection of precise target segments for hyperlinks.

6.6.3. Maurdor campaign

Participant: Christian Raymond.

Joint participation with Yann Ricquebourg, Baptiste Poirriez, Aurélie Lemaitre and Bertrand Coüasnon, IRISA/Intuidoc.

We are participating to the ongoing MAURDOR campaign http://www.maurdor-campaign.org which aims at evaluating systems for automatic processing of written documents. The contribution of TEXMEX comes from the machine learning system based on boosting over bonsai trees we implemented. In the context of this campaign, we investigate the usefulness of this algorithm to combine efficiently features on a relatively big dataset. The very first result shows that this system get state-of-the-art performance while it is much faster than traditional SVM approaches.

6.6.4. Information extraction challenge at BioNLP-ST13

Participant: Vincent Claveau.

BioNLP Shared Task is a community-wide effort to address fine-grained, structural information extraction from biomedical literature. This year, several tasks were proposed and 22 teams participated. TexMex has proposed runs for three main tasks concerning entity extraction and categorization, and relation extraction. The methods proposed by our team are based on machine learning and information retrieval components. Although they do not exploit specialized or domain-specific knowledge, we obtained good results and ranked first, first and third according to the tasks.

7. Bilateral Contracts and Grants with Industry

7.1. Bilateral Grants with Industry

- CIFRE Ph. D. thesis of Ludivine Kuznik with Institut National de l'Audiovisuel
- CIFRE Ph. D. thesis of Bingqing Qu with Institut National de l'Audiovisuel
- CIFRE Ph. D. thesis of Mohamed-Haykel Boukadida with Orange Labs
- CIFRE Ph. D. thesis of Cédric Penet with Technicolor

8. Partnerships and Cooperations

8.1. National Initiatives

8.1.1. ANR FIRE-ID

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

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

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

8.1.2. ANR Secular

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

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

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

8.2. European Initiatives

8.2.1. Collaborations in European Programs, except FP7

Program: Eurostars

Project title: Forensic Image Identifier and Analyzer

Duration: 03/2011 - 07/2014

Coordinator: Videntifier Technologies

Other partners: Videntifier Technologies (Iceland), Forensic Pathways (UK)

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

8.2.2. Quaero

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

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

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

In 2013, a key fact is the best paper award obtained by Cédric Penet at CBMI 2013.

8.3. International Initiatives

8.3.1. Inria International Partners

8.3.1.1. Informal International Partners

- Intelligent Systems Lab Amsterdam (ISLA), University van Amsterdam
- Pontifical Catholic Univeristy of Minas Gerais, Brazil
- National Institute for Informatics, Japan
- Prague Technical University, Czech Republic
- National Technical University of Athens, Greece

8.4. International Research Visitors

8.4.1. Visits of International Scientists

- Michael Rabbat

  Dates: November 2013 (1 month)

  Subject: Continuous Associative Memories

  Institution: Mc Gill University, Canada

8.4.2. Internships

- Giorgos Tolias

  Dates: October 2012–January 2013 (5 months)

  Subject: Large scale visual search

  Institution: National Technical University of Athens (Greece)

8.4.3. Visits to International Teams

- Mihir Jain

  Dates: June 2013–September 2013

  Subject: Action Recognition and Event Retrieval

  Institution: Intelligent Systems Lab Amsterdam (ISLA), University van Amsterdam

9. Dissemination

9.1. Scientific Animation

Laurent Amsaleg

- was a reviewer for TPAMI in 2013;
- was a reviewer for Information Systems in 2013;
- was a reviewer for Multimedia Tools and Applications in 2013;
- was a program committee member of ICMR 2013;
- was a program committee member of MMM 2013;
- was a program committee member of CBMI 2013;
- was a program committee member of ICME 2013;
- was a program committee member of SISAP 2013;
- was a program committee member of SITIS 2013.

Vincent Claveau

- is a member of the editorial board of the journal Traitement Automatique des Langues;
- is a board member of Association pour la Recherche d'Information et Applications;
- was a program committee member of Conf. francophone en Traitement automatique des langues naturelles, France;
- was a program committee member of Conf. en Recherche d'Information et Applications, Switzerland;
- was a program committee member of ISCA/IEEE Workshop on Speech, Language and Audio in Multimedia, France;
- was a reviewer for the journal BioInformatics;
- was a reviewer for the journal Technique et Science Informatique;
- was a editorial committee member of the special issue *Fouille de Données et Humanités Numériques* of the RNTI journal;
- was a reviewer for Hubert Curien program.

Teddy Furon

- was an associate editor for IEEE Trans. on Information Forensics and Security;
- was an associate editor for Elsevier Digital Signal Processing Journal;
- was an associate editor for Hindawi Scientific World Journal;
- was a member of the IEEE Information Forensics and Security Technical Committee;
- was a program committee member of IHMMSEC 2013;
- was a program committee member of CMS 2013;
- was a program committee member of WIFS 2013.

Guillaume Gravier

- is president of the administration council of the French-speaking speech communication association (AFCP);
- is the coordinator of the Special Interest Group Speech and Language in Multimedia of the Intl. Speech Communication Association;
- is a member of the informal coordination group of the Mediaeval benchmarking initiative;
- was a member of the steering committee of Interspeech 2013, in charge of tutorials, special sessions and program coordination;
- was the general chair of the ISCA/IEEE Workshop on Speech, Language and Audio in Multimedia;
- was a program committee member of ACM Conf. on Multimedia;
- was a program committee member of IEEE Intl. Conf. on Multimedia Signal Processing;
- was a program committee member of ISCA Intl. Workshop on Non-Linear Speech Processing;
- was a program committee member of Interspeech;
- was a program committee member of IEEE Intl. Conf. on Multimedia and Expo;
- was a program committee member of Intl. Workshop on Content-based Multimedia Indexing;
- was a program committee member of Intl. Language Resources and Evaluation Conf.;
- was a program committee member of GRETSI;

Patrick Gros

- is scientific officer of the Inria research center of Rennes – Bretagne Atlantique, and thus member of the executive board of the center;
- is a member of the scientific board of Université européenne de Bretagne;
- is a member of the Evaluation Board of Inria;
- was appointed as an expert by the French National Agency;
- was a program committee member of the Intl. Conf. on Creative Content Technologies, Spain;
- was a program committee member of Conférence en Recherche d'Information et Applications, Switzerland;
- was a Technical Program Committee member of the European Signal Processing Conf. Morocco;
- was a program committee member of the ACM Int. Conf. on Multimedia Retrieval, USA;
- was a program committee member of Int. Conf. on Machine Learning and Data Mining, Germany;
- was a program committee member of the MUSCLE Intl. Workshop on Computational Intelligence for Multimedia Understanding, Turkey;
- was a program committee member of GRETSI workshop, France.

Hervé Jégou

- has co-organized a tutorial on Large-Scale Visual Recognition at CVPR 2013;
- was a program committee member of CVPR 2013;
- was a program committee member of ICCV 2013;
- was a program committee member of CBMI 2013;
- was a program committee member of CORESA 2013;
- was a program committee member of FGCV 2013;

François Poulet

- was a program committee member of VINCI'2013;
- was a program committee member of AusDM'2013;
- was a program committee member of KDIR'2013;
- was a program committee member of KICSS'2013;
- was a program committee member of IC3K'2013;
- is a reviewing committee member of AKDM, Springer;
- was a reviewer for DAMI, Data Mining, Springer;
- was a reviewer for IJSS, International Journal of System Science;
- was a reviewer for RNTI, Revue des Nouvelles Technologies de l'Information;

Christian Raymond

- is a member of the editorial board of the journal Discours;
- was a member of the organization committee of ISCA/IEEE Workshop on Speech, Language and Audio in Multimedia;
- was a program committee member of Interspeech;
- was a reviewer for IEEE Transaction on Speech, Language an Audio Processing;
- was a reviewer for the Journal Traitement Automatique des Langues;
- was a program committee member of Conf. francophone en Traitement Automatique des Langues Naturelles;
- was a program committee member of Intl. Conf. on Machine Learning and Applications.

Pascale Sébillot

- was a member of the program committee of Conf. francophone en Traitement automatique des langues naturelles, France;
- was a member of the program committee of Workshop on Distributional Semantics, France;
- was a member of the program committee of Conf. Terminologie et Intelligence Artificielle, France;
- was a member of the program committee of ISCA/IEEE Workshop on Speech, Language and Audio in Multimedia, France;
- is a member of the permanent steering committee of Conf. francophone en Traitement automatique des langues naturelles
- is an editorial committee member of the Journal Traitement Automatique des Langues;
- was a member of the reading committee of several issues of the Journal Traitement Automatique des Langues.

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

9.2.1.1. Course and track responsibilities

Ewa Kijak is head of the Image engineering track of ESIR (Ecole Supérieure d'Ingénieur de Rennes), the engineering school of University of Rennes 1, France.

Guillaume Gravier is coordinator of the course "Data analysis and probabilistic models" of the Master by research in Computer Science (2nd year), University of Rennes 1, France.

Patrick Gros is coordinator of the track "From Data to Knowledge: Machine Learning, Modeling and Indexing Multimedia Contents and Symbolic Data" of the Master by research in Computer Science (2nd year), University of Rennes 1, France.

Francois Poulet is head of Master 2 Mitic, Computer Science Methods and ICT.

9.2.1.2. List of courses

Master: Laurent Amsaleg, High-Dimensional Indexing, 13h, M2R, University Rennes 1, France

Master: Laurent Amsaleg, DataBase tuning, 10h, M2, ENSAI, France

Master: Vincent Claveau, Indexing and multimedia databases, 15h, M2, ENSSAT

Master: Vincent Claveau, Natural Language Processing, 36h, M2, Univ. Rennes 1

Master: Vincent Claveau, Data-Based Knowledge Acquisition 2: Symbolic Methods, 27 hours, M1, INSA de Rennes

Master: Vincent Claveau, Symbolic and sequential data, 10h, M2R, University of Rennes 1

Master: Vincent Claveau, Multimedia indexing and machine learning, M2, 10 hours, University Rennes 1

Master: Ewa Kijak, Image analysis and classification, 30h, M1, ESIR

Master: Ewa Kijak, Image processing, 34h, M1, ESIR

Master: Ewa Kijak, Supervised learning, 16h, M2R, University Rennes 1

Master: Ewa Kijak, Statistical data mining, 26h, M2, University Rennes 1

Master: Ewa Kijak, Indexing and multimedia databases, 15h, M2, ENSSAT

Master: Guillaume Gravier, Data analysis and probabilistic models, 20h, M2, Univ. Rennes 1

Master: Guillaume Gravier, Speech and audio processing, 4h, M1, ESILV

Master: Patrick Gros, Mathematics workshop, 12h, ISTIC, University of Rennes 1

Doctorate: Patrick Gros, Introduction to projective geometry for computer vision and imagery, 12h, University of Rennes 1

Master: François Poulet, Introduction to Data Mining, 21h, M2, ISTIC, University of Rennes 1

Master: François Poulet, Visual Data Mining, 14h, M2, ISTIC, University of Rennes 1

Master: François Poulet, Machine Learning for Multimedia Data, 28h, M2 ISTIC, University of Rennes 1

Master: François Poulet, Visual Analytics, 21h, M2, ISTIC, University of Rennes 1

Master: François Poulet, Large Scale Classification, 11h, M2, ISTIC, University of Rennes 1

Master: Pascale Sébillot, Advanced Databases and Modern Information Systems, 70h, M2, INSA de Rennes

Master: Pascale Sébillot, Data-Based Knowledge Acquisition 2: Symbolic Methods, 18h, M1, INSA de Rennes

9.2.2. Supervision

PhD: Khaoula Elagouni, Combining neural-based approaches and linguistic knowledge for text recognition in multimedia documents, INSA de Rennes, May 28, 2013, Pascale Sébillot, Christophe Garcia (LIRIS, Lyon) and Franck Mamalet (Orange Labs, Rennes)

PhD: Cédric Penet, De l'indexation d'évènements dans des films - Application à la détection de violence, Université Rennes 1, October 10, 2013, Guillaume Gravier, Patrick Gros and Claire-Hélène Demarty (Technicolor)

PhD: Gylfi Gudmunson, Parallelism and distribution for very large scale content-based image retrieval, Université Rennes 1, September 12, 2013, Laurent Amsaleg and Patrick Gros

PhD: Thanh-Nghi Doan: Large Scale Support Vector Machine Algorithms for Visual Classification, University of Rennes 1, Nov.2013, Francois Poulet

PhD in progress: Raghavendran Balu, started October 2013, Teddy Furon and Laurent Amsaleg

PhD in progress: Petra Bosilj, Video description and retrieval, started October 2012, Ewa Kijak and Sebastien Lefèvre (IRISA/SEASIDE)

PhD in progress: Mihir Jain, Video description and retrieval, started February 2011, Hervé Jegou

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

PhD in progress: Abir Ncibi, Machine learning models for TV stream structuring, started October 2011, Vincent Claveau, Guillaume Gravier, Patrick Gros

PhD in progress: Bingqing Qu, Structure discovery of TV programs from an homogeneous collection, started October 2012, Guillaume Gravier and Jean Carrive (INA)

PhD in progress: Anca Roxana Simon, Hierarchical semantic structuring of video collections, started October 1, 2012, Guillaume Gravier and Pascale Sébillot

PhD in progress: Stefan Ziegler, Landmark-driven speech recognition, started September 2010, Guillaume Gravier (with PANAMA)

9.2.3. Juries

Laurent Amsaleg, PhD, Zineddine Kouahla, Univ. Nantes

Guillaume Gravier, PhD, reviewer, A. Abduraman, Eurecom

Guillaume Gravier, PhD, M. Bouallègue, Univ. Avignon

Guillaume Gravier, HDR, reviewer, P. Paroubek, Univ. Paris-Sud

Hervé Jégou, PhD, Stefan Gammeter, ETHZ

Hervé Jégou, PhD, Relja Arandjelovic, Oxford University

Hervé Jégou, PhD, Albert Gordo, UAB

Francois Poulet, PhD, reviewer, Sébastien Heymann, UMPC-LIP6

Pascale Sébillot, HDR, reviewer, Pierre Beust, Caen Basse Normandie university

Pascale Sébillot, PhD, president, Antoine Boutet, Rennes 1 university

Pascale Sébillot, PhD, reviewer, Wei Wang, Paris-Sud university

Pascale Sébillot, PhD, reviewer, Prajol Shresta, Nantes university

Pascale Sébillot, PhD, reviewer and president, Fanny Lalleman, Toulouse university

Pascale Sébillot, PhD, reviewer, Nicolas Foucault, Paris-Sud university

9.3. Popularization

- Guillaume Gravier, Patrick Gros and Hervé Jégou: Invited talks at INA, January 2013
- Guillaume Gravier: Invited speaker at Séminaire Délégation Générale pour l'Armement, Paris
- Guillaume Gravier: Invited talk at PUC Minas, Brazil, April 2013

- Guillaume Gravier: Seminar on Speech and Audio Processing, M. Sc. Students, INSA Rennes, France
- Patrick Gros and Pascale Sébillot: Invited speakers at the workshop on Big Data, ENSAI, Rennes, France, November 2013.
- Guillaume Gravier: Invited speaker at Colloquium Rennais des Sciences du numérique, Rennes
- Hervé Jégou: Invited talk at Universita Autonoma de Barcelona, January 2013
- Hervé Jégou: Invited talk at University of Caen, January 2013
- Hervé Jégou: Invited talk at Oxford University, Visual Geometry Group, UK, November 2013

10. Bibliography

Major publications by the team in recent years

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

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

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

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

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

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

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

Publications of the year

Doctoral Dissertations and Habilitation Theses

[8] G. T. GUDMUNDSSON. , *Parallelism and distribution for very large scale content-based image retrieval*, Université Rennes 1, September 2013, http://hal.inria.fr/tel-00926069

[9] C. PENET. , *De l'indexation d'évènements dans des films - Application à la détection de violence*, Université Rennes 1, October 2013, http://hal.inria.fr/tel-00909117

Articles in International Peer-Reviewed Journals

[10] P. BAS, T. FURON. *A New Measure of Watermarking Security: The Effective Key Length*, in "IEEE Transactions on Information Forensics and Security", July 2013, vol. 8, n° 8, pp. 1306 - 1317, http://hal.inria.fr/hal-00836404

[11] V. CLAVEAU, E. KIJAK. *Analyse morphologique non supervisée en domaine biomédical. Application à la recherche d'information*, in "Traitement Automatique des Langues", October 2013, vol. 54, n° 1, pp. 54-1, http://hal.inria.fr/hal-00912301

[12] J. D. CRUZ GOMEZ, C. BOTHOREL, F. POULET. *Community detection and visualization in social networks: integrating structural and semantic information*, in "ACM Transactions on Intelligent Systems and Technology", December 2013, vol. 5, n° 1, pp. 11-26 [*DOI :* 10.1145/2542182.2542193], http://hal.inria.fr/hal-00763931

[13] J. D. CRUZ GOMEZ, C. BOTHOREL, F. POULET. *Analyse intégrée des réseaux sociaux pour la détection et la visualisation de communautés*, in "Technique et Science Informatiques (TSI)", 2014, vol. 33, http://hal.inria.fr/hal-00937849

[14] T.-N. DOAN, T. N. DO, F. POULET. *Large Scale Classifiers for Visual Classification Tasks*, in "Multimedia Tools and Applications", 2014, 24 p. , http://hal.inria.fr/hal-00952765

[15] T.-N. DOAN, T. N. DO, F. POULET. *Parallel Incremental Power Mean SVM for Classification of Large Scale Visual Datasets*, in "International Journal of Multimedia Information Retrieval", 2014, 15 p. , http://hal.inria.fr/hal-00943187

[16] K. ELAGOUNI, C. GARCIA, F. MAMALET, P. SÉBILLOT. *Text Recognition in Multimedia Documents: A Study of two Neural-based OCRs Using and Avoiding Character Segmentation*, in "International Journal on Document Analysis and Recognition, IJDAR", 2013, 13 p. [*DOI :* 10.1007/s10032-013-0202-7], http://hal.inria.fr/hal-00867225

[17] V. F. MOTA, E. D. A. PEREZ, S. MACIEL LUIZ MAURÍLIO DA, M. B. VIEIRA, P.-H. GOSSELIN. *A tensor motion descriptor based on histograms of gradients and optical flow*, in "Pattern Recognition Letters", April 2014, vol. 39, pp. 85-91 [*DOI :* 10.1016/J.PATREC.2013.08.008], http://hal.inria.fr/hal-00861395

[18] R. TAVENARD, L. AMSALEG. *Improving the Efficiency of Traditional DTW Accelerators*, in "Knowledge and Information Systems", 2013, http://hal.inria.fr/hal-00862176

Invited Conferences

[19] P.-H. GOSSELIN, D. PICARD. *Machine Learning and Content-Based Multimedia Retrieval*, in "European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning", Bruges, Belgium, April 2013, pp. 251-260, http://hal.inria.fr/hal-00864824

International Conferences with Proceedings

[20] M. BADR, D. VODISLAV, D. PICARD, S. YIN, P.-H. GOSSELIN. *Multi-criteria Search Algorithm: An Efficient Approximate K-NN Algorithm for Image Retrieval*, in "IEEE Int. Conf. on Image Processing ICIP2013", Melbourne, Australia, September 2013, pp. 2901-2905, http://hal.inria.fr/hal-00832196

[21] R. BALU, T. FURON, H. JÉGOU. *Beyond "project and sign" for cosine estimation with binary codes*, in "ICASPP - International Conference on Acoustics, Speech, and Signal Processing", Florence, Italy, IEEE, February 2014, http://hal.inria.fr/hal-00942075

[22] R. BENMOKHTAR, J. DELHUMEAU, P.-H. GOSSELIN. *Efficient Supervised Dimensionality Reduction for Image Categorization*, in "IEEE International Conference on Acoustics, Speech, and Signal Processing", Vancouver, Canada, May 2013, pp. 2425-2428, http://hal.inria.fr/hal-00807483

[23] S.-A. BERRANI, M. H. BOUKADIDA, P. GROS. *Constraint satisfaction programming for video summarization*, in "IEEE International Symposium on Multimedia", Anaheim, California, United States, IEEE, December 2013, http://hal.inria.fr/hal-00909370

[24] P. BOSILJ, S. LEFÈVRE, E. KIJAK. *Hierarchical Image Representation Simplification Driven by Region Complexity*, in "International Conference on Image Analysis and Processing", Naples, Italy, 2013, pp. 562-571 [*DOI :* 10.1007/978-3-642-41181-6_57], http://hal.inria.fr/hal-00921674

[25] C. CARTON, P. GROS. *Fast Repetition Detection in TV Streams using Repetition Patterns*, in "Content-Based Multimedia Indexing", Veszprem, Hungary, June 2013, http://hal.inria.fr/hal-00844099

[26] V. CLAVEAU. *IRISA participation to BioNLP-ST 2013: lazy-learning and information retrieval for information extraction tasks*, in "BioNLP Workshop, colocated with ACL 2013", Bulgaria, August 2013, pp. 188-196, http://hal.inria.fr/hal-00912308

[27] V. CLAVEAU, A. NCIBI. *Découverte de connaissances dans les séquences par CRF non-supervisés*, in "20ème conférence sur le Traitement Automatique des Langues Naturelles, TALN", Sables d'Olonne, France, June 2013, vol. 1, volume 1 p. , http://hal.inria.fr/hal-00912314

[28] J. D. CRUZ GOMEZ, C. BOTHOREL, F. POULET. *Layout Algorithm for Clustered Graphs to Analyze Community Interactions in Social Networks*, in "International Network for Social Network Analysis", Hamburg, Germany, May 2013, 2 p. , http://hal.inria.fr/hal-00780515

[29] L. DE OLIVEIRA, Z. K. DO PATROCÍNIO JR., S. J. GUIMARÃES, G. GRAVIER. *Searching for near-duplicate video sequences from a scalable sequence aligner*, in "IEEE International Symposium on Multimedia", United States, 2013, 4 p. , http://hal.inria.fr/hal-00906327

[30] J. DELHUMEAU, P.-H. GOSSELIN, H. JÉGOU, P. PÉREZ. *Revisiting the VLAD image representation*, in "ACM Multimedia", Barcelona, Spain, October 2013, http://hal.inria.fr/hal-00840653

[31] A. DELVINIOTI, H. JÉGOU, L. AMSALEG, M. HOULE. *Image Retrieval with Reciprocal and shared Nearest Neighbors*, in "VISAPP–International Conference on Computer Vision Theory and Applications", Barcelone, Portugal, 2014, http://hal.inria.fr/hal-00907481

[32] C.-H. DEMARTY, C. PENET, M. SCHEDL, I. BOGDAN, V. L. QUANG, Y.-G. JIANG. *The MediaEval 2013 Affect Task: Violent Scenes Detection*, in "MediaEval 2013 Working Notes", Spain, October 2013, 2 p. , http://hal.inria.fr/hal-00932551

[33] T.-N. DOAN, T. N. DO, F. POULET. *Large Scale Image Classification with Many Classes, Multi-features and Very High-Dimensional Signatures*, in "1st International Conference on Computer Science, Applied

Mathematics and Applications", Varsaw, Poland, SCI-479, Studies in Computational Intelligence, Springer Verlag, May 2013, pp. 105-116, http://hal.inria.fr/hal-00907859

[34] T.-N. DOAN, T. N. DO, F. POULET. *Large Scale Visual Classification with Many Classes*, in "9th International Conference on Machine Learning and Data Mining in Pattern Recognition", New-York, United States, P. PERNER (editor), Lecture Notes in Artificial Intelligence - 7988, Springer Verlag, July 2013, pp. 629-643 [*DOI :* 10.1007/978-3-642-39712-7_48], http://hal.inria.fr/hal-00907847

[35] T.-N. DOAN, T. N. DO, F. POULET. *Large Scale Visual Classification with Parallel, Imbalanced Bagging of Incremental LIBLINEAR SVM*, in "DMIN'13, the 9th International Conference on Data Mining", Las Vegas, United States, July 2013, 7 p. , http://hal.inria.fr/hal-00907868

[36] T.-N. DOAN, T. N. DO, F. POULET. *Parallel incremental SVM for classifying million images with very high-dimensional signatures into thousand classes*, in "International Joint Conference on Neural Networks", Dallas, United States, P. ANGELOV, D. LEVINE, P. ERDI (editors), July 2013, pp. 2976-2983 [*DOI :* 10.1109/IJCNN.2013.6707121], http://hal.inria.fr/hal-00907881

[37] T.-N. DOAN, F. POULET, T. N. DO. *Multi-way Classification for Large Scale Visual Object Dataset*, in "11th International Workshop on Content-Based Multimedia Indexing, CBMI 2013", Veszprém, Hungary, L. CZÚNI (editor), June 2013, pp. 185-190 [*DOI :* 10.1109/CBMI.2013.6576579], http://hal.inria.fr/hal-00907851

[38] J. A. DOS SANTOS, O. PENATTI, R. DA SILVA TORRES, P.-H. GOSSELIN, S. PHILIPP-FOLIGUET, A. X. FALCAO. *Remote Sensing Image Representation based on Hierarchical Histogram Propagation*, in "IEEE International Geoscience and Remote Sensing Symposium", Melbourne, Australia, July 2013, 4 p. , http://hal.inria.fr/hal-00861379

[39] M. DOUZE, J. REVAUD, C. SCHMID, H. JÉGOU. *Stable hyper-pooling and query expansion for event detection*, in "ICCV 2013 - IEEE International Conference on Computer Vision", Sydney, Australia, IEEE, October 2013, http://hal.inria.fr/hal-00872751

[40] M. ESKEVICH, G. J. F. JONES, R. ALY, R. J. ORDELMAN, S. CHEN, D. NADEEM, C. GUINAUDEAU, G. GRAVIER, P. SÉBILLOT, T. DE NIES, P. DEBEVERE, R. VAN DE WALLE, P. GALUSCAKOVA, P. PECINA, M. LARSON. *Multimedia Information Seeking through Search and Hyperlinking*, in "ACM International Conference on Multimedia Retrieval, ICMR 2013", Dallas, United States, April 2013, 8 p. , http://hal.inria.fr/hal-00867090

[41] J. FELLUS, D. PICARD, P.-H. GOSSELIN. *Decentralized K-means using randomized Gossip protocols for clustering large datasets*, in "International Workshop on Knowledge Discovery Using Cloud and Distributed Computing Platforms", Dallas, Texas, United States, December 2013, 8 p. , http://hal.inria.fr/hal-00915822

[42] T. FURON, H. JÉGOU, L. AMSALEG, B. MATHON. *Fast and secure similarity search in high dimensional space*, in "IEEE International Workshop on Information Forensics and Security", Guangzhou, China, 2013, http://hal.inria.fr/hal-00857570

[43] M. JAIN, H. JÉGOU, P. BOUTHEMY. *Better exploiting motion for better action recognition*, in "CVPR - International Conference on Computer Vision and Pattern Recognition", Portland, United States, April 2013, http://hal.inria.fr/hal-00813014

[44] B. Þ. JÓNSSON, Á. ERÍKSDÓTTIR, Ó. WAAGE, G. TÓMASSON, H. SIGURÞÓRSSON, L. AMSALEG. *M3 + P3+ O3 = Multi-D Photo Browsing*, in "International Conference on MultiMedia Modeling", Dublin, Ireland, 2014, http://hal.inria.fr/hal-00925464

[45] O. KIHL, D. PICARD, P.-H. GOSSELIN. *A Unified Formalism for Video Descriptors*, in "IEEE Int. Conf. on Image Processing ICIP2013", Melbourne, Australia, September 2013, pp. 2416-2419, http://hal.inria.fr/hal-00832190

[46] J. KRAPAC, F. PERRONNIN, T. FURON, H. JÉGOU. *Instance classification with prototype selection*, in "ICMR - ACM International Conference on Multimedia Retrieval", Glasgow, United Kingdom, February 2014, http://hal.inria.fr/hal-00942275

[47] B. MATHON, T. FURON, L. AMSALEG, J. BRINGER. *Secure and Efficient Approximate Nearest Neighbors Search*, in "1st ACM Workshop on Information Hiding and Multimedia Security", Montpellier, France, A. UHL (editor), IH & MMSec '13, June 2013, pp. 175–180 [*DOI :* 10.1145/2482513.2482539], http://hal.inria.fr/hal-00817336

[48] D. MOISE, D. SHESTAKOV, G. T. GUDMUNDSSON, L. AMSALEG. *Indexing and Searching 100M Images with Map-Reduce*, in "ACM International Conference on Multimedia Retrieval", Dallas, United States, 2013, http://hal.inria.fr/hal-00796475

[49] D. MOISE, D. SHESTAKOV, G. Þ. GUÐMUNDSSON, L. AMSALEG. *Terabyte-scale image similarity search : experience and best practice*, in "IEEE International Conference on Big Data", Santa Clara, United States, 2013, http://hal.inria.fr/hal-00857577

[50] A. NCIBI, E. MARTIENNE, V. CLAVEAU, G. GRAVIER, P. GROS. *Robust TV Stream Labelling with Conditional Random Fields*, in "MMEDIA - 5th International Conference on Advances in Multimedia", Venise, Italy, IARIA, April 2013, pp. 88-95, http://hal.inria.fr/hal-00844640

[51] *Best Paper*
C. PENET, C.-H. DEMARTY, G. GRAVIER, P. GROS. *Audio Event Detection in Movies using Multiple Audio Words and Contextual Bayesian Networks*, in "CBMI - 11th International Workshop on Content Based Multimedia Indexing - 2013", Veszprém, Hungary, June 2013, http://hal.inria.fr/hal-00822022.

[52] C. PENET, C.-H. DEMARTY, G. GRAVIER, P. GROS. *Technicolor/Inria team at the MediaEval 2013 Violent Scenes Detection Task*, in "MediaEval 2013 Working Notes", Spain, 2013, 2 p. , http://hal.inria.fr/hal-00906300

[53] R. PRIAM, M. NADIF, G. GOVAERT. *Gaussian topographic co-clustering model*, in "IDA 2013 : The Twelfth International Symposium on Intelligent Data Analysis", France, 2013, pp. 345-356, http://hal.inria.fr/hal-00933243

[54] C. RAYMOND. *Robust tree-structured named entities recognition from speech*, in "Proceedings of the International Conference on Acoustic Speech and Signal Processing", Vancouver, Canada, 2013, http://hal.inria.fr/hal-00830142

[55] J. REVAUD, M. DOUZE, C. SCHMID, H. JÉGOU. *Event retrieval in large video collections with circulant temporal encoding*, in "CVPR 2013 - International Conference on Computer Vision and Pattern Recognition", Portland, United States, IEEE, March 2013, pp. 2459-2466 [*DOI :* 10.1109/CVPR.2013.318], http://hal.inria.fr/hal-00801714

[56] Y. RICQUEBOURG, C. RAYMOND, B. POIRRIEZ, A. LEMAITRE, B. COÜASNON. *Boosting bonsai trees for handwritten/printed text discrimination*, in "Document Recognition and Retrieval (DRR)", San Francisco, United States, 2014, http://hal.inria.fr/hal-00910718

[57] D. SHESTAKOV, D. MOISE, G. T. GUDMUNDSSON, L. AMSALEG. *Scalable high-dimensional indexing with Hadoop*, in "CBMI—International Workshop on Content-Based Multimedia Indexing", Veszprém, Hungary, 2013, http://hal.inria.fr/hal-00817378

[58] A.-R. SIMON, G. GRAVIER, P. SÉBILLOT. *Leveraging lexical cohesion and disruption for topic segmentation*, in "International Conference on Empirical Methods in Natural Language Processing, EMNLP 2013", Seattle, United States, October 2013, 11 p. , http://hal.inria.fr/hal-00867011

[59] A.-R. SIMON, G. GRAVIER, P. SÉBILLOT. *Un modèle segmental probabiliste combinant cohésion lexicale et rupture lexicale pour la segmentation thématique*, in "TALN - Conférence sur le traitement automatique des langues naturelles", Les Sables d'Olonne, France, ATALA, June 2013, http://hal.inria.fr/hal-00844112

[60] H. TABIA, D. PICARD, H. LAGA, P.-H. GOSSELIN. *3D Shape Similarity Using Vectors of Locally Aggregated Tensors*, in "IEEE International Conference on Image Processing", Melbourne, Australia, September 2013, pp. 2694-2698, http://hal.inria.fr/hal-00832182

[61] H. TABIA, D. PICARD, H. LAGA, P.-H. GOSSELIN. *Fast Approximation of Distance Between Elastic Curves using Kernels*, in "British Machine Vision Conference", United Kingdom, September 2013, 11 p. , BMVC 2013, http://hal.inria.fr/hal-00861369

[62] G. TOLIAS, Y. AVRITHIS, H. JÉGOU. *To aggregate or not to aggregate: selective match kernels for image search*, in "ICCV - International Conference on Computer Vision", Sydney, Australia, September 2013, http://hal.inria.fr/hal-00864684

[63] M. TREVISIOL, H. JÉGOU, J. DELHUMEAU, G. GRAVIER. *Retrieving Geo-Location of Videos with a Divide & Conquer Hierarchical Multimodal Approach*, in "ICMR - International Conference of Multimedia Retrieval", Dallas, United States, ACM, April 2013, http://hal.inria.fr/hal-00801698

[64] D. WEISSENBACHER, Y. SASAKI. *Which Factors Contributes to Resolving Coreference Chains with Bayesian Networks?*, in "14th International Conference on Intelligent Text Processing and Computational Linguistics", Samos, Greece, March 2013, pp. 200-212, http://hal.inria.fr/hal-00844450

[65] W.-L. ZHAO, H. JÉGOU, G. GRAVIER. *Oriented pooling for dense and non-dense rotation-invariant features*, in "BMVC - 24th British Machine Vision Conference", Bristol, United Kingdom, September 2013, http://hal.inria.fr/hal-00841590

[66] W.-L. ZHAO, H. JÉGOU, G. GRAVIER. *Sim-Min-Hash: An efficient matching technique for linking large image collections*, in "ACM Multimedia", Barcelona, Spain, ACM, October 2013, http://hal.inria.fr/hal-00839921

[67] C.-Z. ZHU, H. JÉGOU, S. SATOH. *Query-adaptive asymmetrical dissimilarities for visual object retrieval*, in "ICCV - International Conference on Computer Vision", Sydney, Australia, October 2013, http://hal.inria.fr/hal-00872957

[68] S. ZIEGLER, G. GRAVIER. *A framework for integrating heterogeneous sporadic knowledge sources into automatic speech recognition*, in "Workshop on Speech, Language and Audio in Multimedia", France, 2013, pp. 37-42, http://hal.inria.fr/hal-00906348

National Conferences with Proceedings

[69] T.-N. DOAN, F. POULET. *Classification d'images à grande échelle*, in "Orasis 2013, 14e journées francophones des jeunes chercheurs en vision par ordinateur", Cluny, France, June 2013, http://hal.inria.fr/hal-00907855

[70] B. MATHON, T. FURON, L. AMSALEG, J. BRINGER. *Recherche approximative de plus proches voisins efficace et sûre*, in "GRETSI", Brest, France, September 2013, http://hal.inria.fr/hal-00823879

Conferences without Proceedings

[71] L. CATANESE, N. SOUVIRAÀ-LABASTIE, B. QU, S. CAMPION, G. GRAVIER, E. VINCENT, F. BIMBOT. *MODIS: an audio motif discovery software*, in "Show & Tell - Interspeech 2013", Lyon, France, August 2013, http://hal.inria.fr/hal-00931227

Scientific Books (or Scientific Book chapters)

[72] , *Proceedings of the 14th Annual Conference of the International Speech Communication Association (Interspeech), 25-29 August 2013, Lyon (France)*, International Speech Communication Association (ISCA), August 2013, 3500 p. , http://hal.inria.fr/hal-00931864

[73] T.-N. DOAN, F. POULET. *Large Scale Image Classification: fast Feature Extraction, Multi-CodeBook Approach and SVM Training*, in "AKDM-4, Advances in Knowledge Discovery and Management, Vol.4", Studies in Computational Intelligence, Springer-Verlag, 2014, pp. 159-176, http://hal.inria.fr/hal-00907533

Research Reports

[74] T. FURON, H. JÉGOU. , *Using extreme value theory for image detection*, Inria, February 2013, n^{o} RR-8244, http://hal.inria.fr/hal-00789804

[75] P.-H. GOSSELIN, N. MURRAY, H. JÉGOU, F. PERRONNIN. , *Inria+Xerox@FGcomp: Boosting the Fisher vector for fine-grained classification*, Inria, December 2013, n^{o} RR-8431, http://hal.inria.fr/hal-00920187

[76] P. KONIUSZ, F. YAN, P.-H. GOSSELIN, K. MIKOLAJCZYK. , *Higher-order Occurrence Pooling on Mid- and Low-level Features: Visual Concept Detection*, September 2013, 20 p. , http://hal.inria.fr/hal-00922524

[77] N. SOUVIRAÀ-LABASTIE, L. CATANESE, G. GRAVIER, F. BIMBOT. , *The MODIS software for word like motif discovery and its use for zero resource audio summarization*, Inria, July 2013, n^{o} RT-0439, http://hal.inria.fr/hal-00848631

[78] G. TOLIAS, H. JÉGOU. , *Local visual query expansion: Exploiting an image collection to refine local descriptors*, Inria, July 2013, n^{o} RR-8325, http://hal.inria.fr/hal-00840721

Other Publications

[79] L. DE OLIVEIRA, Z. K. DO PATROCÍNIO JR., S. J. GUIMARÃES, G. GRAVIER. , *Searching for near-duplicate video sequences from a scalable sequence aligner*, November 2013, http://hal.inria.fr/hal-00906164

[80] C. GUINAUDEAU, A.-R. SIMON, G. GRAVIER, P. SÉBILLOT. , *HITS and IRISA at MediaEval 2013: Search and Hyperlinking Task*, November 2013, http://hal.inria.fr/hal-00906249

[81] R. TAVENARD, H. JÉGOU, M. LAGRANGE. , *Efficient Cover Song Identification using approximate nearest neighbors*, February 2013, http://hal.inria.fr/hal-00672897

References in notes

[82] 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

[83] N. ALAJLAN, M. S. KAMEL, G. H. FREEMAN. *Geometry-Based Image Retrieval in Binary Image Databases*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", 2008, vol. 30, n^{o} 6, pp. 1003–1013

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

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

[86] 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, pp. 518–529

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

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

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

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

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

[92] B. QU, F. VALLET, J. CARRIVE, G. GRAVIER. *Using grammar induction to discover the structure of recurrent TV programs*, in "Intl. Conf. on Advances in Multimedia", 2014

[93] L. RABINER, B.-H. JUANG. , *Fundamentals of speech recognition*, Prentice HallEnglewood Cliffs, NJ, 1993

[94] Y. RICQUEBOURG, C. RAYMOND, B. POIRRIEZ, A. LEMAITRE, B. COÜASNON. *Boosting bonsai trees for handwritten/printed text discrimination*, in "Document Recognition and Retrieval (DRR)", 2014

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

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

[97] E. URBACH, J. B. T. M. ROERDINK, M. H. F. WILKINSON. *Connected Shape-Size Pattern Spectra for Rotation and Scale-Invariant Classification of Gray-Scale Images*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", 2007, vol. 29, n^{o} 2, pp. 272–285

[98] V. VILAPLANA, F. MARQUES, P. SALEMBIER. *Binary Partition Trees for Object Detection*, in "IEEE Transactions on Image Processing", 2008, vol. 17, n^{o} 11, pp. 2201–2216

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