INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Project-Team TEXMEX

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

Rennes

THEME SYM

Activity Report

2006

Table of contents

1. Team

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

Head of the team
Patrick Gros [Senior Research Scientist (DR) INRIA, HdR]

Administrative Assistant
Loïc Lesage [Secretary (TR) INRIA, partial position in the project-team]

Faculty Members (University of Rennes 1)
Ewa Kijak [Associate Professor (MCF), since September 1^{st}]
Annie Morin [Associate Professor (MCF), HdR]

Faculty Members (National Institute of Applied Sciences of Rennes)
Pascale Sébillot [Full Professor (PR), HdR]
Anthony Remazeilles [Assistant Professor (ATER), until August 31^{st}]

CNRS Staff Members
Laurent Amsaleg [Research Scientist (CR)]
Vincent Claveau [Research Scientist (CR)]

Research Scientists (partners)
Emmanuelle Martienne [Associate Professor, University of Rennes 2, since October 1^{st}]
Laurent Ughetto [Associate Professor, University of Rennes 2, since October 1^{st}]

Ph.D. Students
Siwar Baghdadi [CIFRE grant with Thomson]
Nicolas Bonnel [Assistant Professor (ATER), National Institute of Applied Sciences of Rennes]
Manolis Delakis [MJENR grant, also with METISS, until October 31^{st}]
Stéphane Huet [MJENR grant, also with METISS]
Zied Jemai [INRIA grant]
Gaël Manson [CIFRE grant with France Télécom, since November 1^{st}]
Fabienne Moreau [Assistant Professor (ATER), University of Rennes 1]
Xavier Naturel [INRIA grant]
Nguyen Khang Pham [also with Institut Francophone pour l'informatique, Hanoï, Vietnam]
Pierre Tirilly [CNRS - Brittany Council grant, since October 1^{st}]
François Tonnin [INRIA - Brittany Council grant, also with TEMICS, until May 31^{st}]

Post-Doctoral Fellows
Antoine Doucet [INRIA grant, since October 1^{st}]
Hervé Jegou [MJENR grant, until August 31^{st}]

Technical Staff
Cédric Dufouil [INRIA Junior technical staff]
Arnaud Dupuis [INRIA Technical Staff]
Pierre-Hugues Joalland [AceMedia project]

2. Overall Objectives

2.1. Overall Objectives

Keywords: *Databases*, *Document Content-Based Access*, *Exploration*, *Image Recognition*, *Indexing*, *Machine Learning*, *Multimedia*, *Natural Language Processing*, *Search*.

The explosion of the quantity of numerical documents raises the problem of the management of these documents. Beyond the storage, we are interested in the problems linked to the management of the contents: how to exploit the large databases of documents, how to classify documents, how to index them in order to search efficiently their contents, how to visualize their contents? To tackle these problems, we propose a multi-field work gathering within the same team specialists of the various media: image, video, text, and specialists in data and related metadata exploitation techniques such as the database techniques, statistics, and information retrieval. Our work is at the intersection of these fields and relates more particularly to 3 points: i) searching in large image databases, ii) adding semantics to search engines, and iii) coupling media for multimedia document description.

Exploiting the content of large databases of digital multimedia documents is a problem with multiple facets. Moreover, the construction of a system exploiting such databases calls upon many techniques: study and description of documents, organization of the bases, search algorithms, classification, visualization, but also needs an efficient management of the primary and secondary memories, as well as nice interfaces and interactions with the user.

The five major challenges of the field TEXMEX aims at tackling are the following ones:

- it is necessary, first of all, to be able **to process large sets of documents**: it is important to develop techniques that scale up gracefully with respect to the quantity of documents taken into account (millions of images, months of videos), and to evaluate their results in quality as well as in speed;
- multimedia documents are not a simple juxtaposition of independent media, and it is important **to better exploit the existing links between the various media** composing a unique document;
- **multimedia document databases are evolutionary**: it is important to take into account that the sets of documents evolve, as do the document description techniques and the modes of questioning, which modifies in turn the way the databases are used;
- towards queries of a semantic nature for their majority, description techniques have only access to the document syntax; it is thus necessary to find means for **reducing this difference between semantic needs and syntactic description tools**;
- **the user-system interaction** is a central point: the user must be able to translate his/her needs efficiently and simply but very precisely, to guide the system or to evaluate the results; he/she must be the one who controls the system.

We have adopted a matricial organization for laying out our research. On the one hand, we have expertise in two main fields, automatic document description and exploitation of these descriptions, and on the other hand, we defined three transverse axis of research. The underlying idea is to focus our work on the questions where the team's multidisciplinarity appears to be an asset to obtain original results.

Our First Field of Competence: Document Description Documents are generally not exploitable directly for search or indexing tasks: it is necessary to use intermediate descriptions which must carry the maximum of information on document semantics, but must also be automatically computable. To the documents and their descriptors, one can add metadata, which we define here as all additional information which inform, supplement or qualify the data with which they are associated.

Our Second Field of Competence: Description Exploitation The question is to define the techniques which make it possible to apprehend, handle and exploit large volumes of data, metadata and descriptors, which have been extracted from the documents: i) **organization and management of the multimedia databases**, including the control of logical and temporal consistency, strategies of computation and selection of descriptors and metadata; ii) **statistical techniques** for the exploration of large volumes of data; iii) **indexing techniques** aiming at confining in the smallest possible area the exploitation of the data and thus avoiding an exhaustive processing whose cost is certainly controlled but prohibitive; iv) **system problems** related to the physical organization of large volumes of data, like disk access management or cache memory management requiring new techniques which are adapted to the characteristics of the descriptors and to the way they are used.

First Axis of Research: Searching in Large Image Bases Going from corpora of a few thousands of images to corpora containing a few millions remains a research challenge today. The solution can neither solely come from new description schemes nor new indexing schemes, but it requires to take into account all the various components of the system and their articulations. Thus, we work on:

- data description, especially in the case of compressed or watermarked images,
- indexing and search algorithms,
- database organization and use of the metadata,
- system and hardware support,

and on the merging of these various techniques to improve the performances of the current systems in speed as well as in quality of recognition.

Second Axis of Research: Towards More Semantic Search Engines Search engines are extensively used tools, but they appear to be disappointing most of the time, due to their syntactic approach based on keywords searching. Natural language processing (NLP) tools could however offer more semantic capabilities, by allowing word sense disambiguation and the possibility to recognize the various formulations of a same concept. It is thus advisable to merge NLP and traditional keyword-based approaches.

However, this union is not so simple. On the one hand, it requires to provide query and document extension strategies to search engines and then to translate these extensions in terms of similarity. On the other hand, natural language processing tools must work in much broader environments than the ones in which they are usually used. The contribution of such a modification of the engines must also be established, which requires a precise work on the evaluation of information retrieval systems.

Third Axis of Research: Multimedia and Cross-Media We study media coupling along three directions. Within the framework of video, we are interested in descriptions which jointly use the sound and image tracks of the video. Such techniques can be applied to automatic video structuring, but also to improve people detection and recognition techniques, whether it is by their face or their voice. Another interesting direction consists of using NLP techniques on texts produced by speech transcriptions. As a matter of fact, speech carries a lot of semantic information and NLP techniques are among the most efficient ones for extracting semantics from textual data.

In addition, we study the interactions between text and image in the documents where these two media are tightly coupled, a common case in scientific bibliographical databases, on the web, in newspapers, in art books or technical documents. One goal is to connect, in the same document, the image and the text associated with images. This could help in obtaining an automatic and semantic description of the images, to link different documents, either by searching for visually similar images, or by searching for texts about a same subject, and thus to improve the description of the images and to remove possible ambiguities in the comprehension of the text.

Moreover, we have also begun the study of the interactions between speech and text together with the METISS Team. This work aims at adapting and inserting methods existing in the text analysis domain into speech recognition models to improve their performances in order to give indexing methods a better access to information such speeches may contain.

3. Scientific Foundations

3.1. Background

The work within the team needs two kinds of competencies: to exploit the content of documents, one should first be able to access this content, *i.e.*, to characterize or describe it. One should also be able to use this description in order to fulfill the tasks related to these documents. Finally, both the descriptors and exploitation techniques must satisfy the user's needs (and proving this simple fact is not so trivial).

Finding a solution requires the use of document description techniques based on text, image or video processing (sound and speech processing are studied by the METISS team with which we closely collaborate). It is also necessary to exploit the correlation and complementarity between the different media, since they do not bring the same information and do not suffer from the same limitations.

After this description stage, it is necessary to exploit the descriptions to satisfy the user's query. At this second stage, are needed sorting, indexing, retrieving algorithms which must provide good and fast results, that are two usually conflicting constraints.

These two aspects are not independent and any solution with only one of the two aspects cannot solve any real problem. The combination of the two in the context of large databases raises many difficult, but interesting, questions, and their solution only comes from a confrontation of people and ideas coming from both sides.

3.2. Document Description and Metadata

Keywords: *Low-level Descriptor*, *Metadata*.

All the multimedia documents have the ambivalent characteristic to be, on the one hand, very rich semantically and, on the other hand, very poor, especially when considering the elementary components which constitute them (sets of characters or of pixels). More concise and informative descriptions are needed in order to handle these documents.

3.2.1. Image Description

Keywords: *Image Indexing*, *Image Matching*, *Image Recognition*, *Invariants*.

Computing image descriptors has been studied for about thirty years. The aim of such a description is to extract indices called descriptors whose distance reflects those of the images they are computed from. This problem can be seen as a coding problem: how images should be coded such that the similarity between the codes reflects the similarity between the original images?

The first difficulty of the problem is that image similarity is not a well-defined concept. Images are polysemic, and their similarity level will depend on the user who judges this similarity, on the problem this user tries to solve, and on the set of images he/she uses at this moment. As a consequence, there does not exist a single descriptor which can solve every problem.

The problem can be specialized with respect to the different kinds of users, databases and needs. As an example, the problem of professional users is usually very specific, when domestic users need more generic solutions. The same difference occurs between databases composed of very dissimilar images and those only composed of images of the same kind (*e.g.*, fingerprints or X-ray images). Finally, retrieving one particular image from an excerpt or browsing a database to choose a set of images may require very different descriptors.

To solve these problems, many descriptors have been proposed in the literature. The most frequent frame of use considered is that of image retrieval from a large database using the query-by-example paradigm. In this case, the descriptors integrate information on the whole image: color histograms in various color spaces, texture descriptors, shape descriptors (with the major drawback that it requires an automatic image segmentation). This field of research is still active: color histograms provide too poor information to solve any problem as soon as the size of the database increases [102]. Several solutions have been proposed to remedy this problem: correlograms [79], weighted histograms [51]...

Texture histograms are usually useful for one kind of texture, but they fail to describe all the possible textures, and no technique exists to decide in which category a given texture falls, and thus which descriptor should be used to describe it properly. Shape descriptors suffer from a lack of robustness.

Many other researches have been carried out in the case of specific databases. Face detection and recognition is the most classical and important case, but other studies concern medical images for example.

In the team, we work with a different paradigm based on local descriptors: one image is described by a set of descriptors. This solution opens the possibility of partial recognitions, like object recognitions independently of the background [100].

The main stages of the method are the following. First, simple features are extracted from each image (interest points in our case, but edges and regions can be used too). The most widely used extractor is the Harris point detector [77] which provides not very precise but "repeatable" points. Other detectors exist, even for points [86].

The similarity between images are then translated into the concept of invariance: measurements of the image invariants to some geometric (rotation, translations, scalings) or photometric (intensity variations) transformations are searched for. In practice, this concept of invariance is usually replaced by the weaker concept of quasi-invariance [47] or by properties only experimentally established [67], [66].

In the case of points, the classical technique consists in characterizing the signal around each point by its convolution with the Gaussian and its first derivatives and by mixing these measurements in order to obtain the invariance properties. The invariance with respect to rotations, scalings and affine transformations was obtained respectively by Florack [68], Dufournaud [61] and Mikolajczyk [92], photometric invariance was demonstrated for grey-levels by Schmid [100] and for color by Gros [74]. The difficult point is that not only invariant quantities have to be computed, but that the feature extractor has to be invariant itself to the same set of transformations.

One of the main difficulties of the domain is the evaluation and the comparison of the methods. Each one corresponds to a slightly different problem. Comparing them is difficult and usually unfair: the results depend on the used databases, especially when these are quite small. In this case, a simple syntactic criterion can give the feeling of a good semantic description, but this does not tell anything about what would happen with a larger database.

3.2.2. *Video Description*

Keywords: *Key-Events*, *Structuring*, *Video Indexing*.

Professional and domestic video collections are usually much bigger than the corresponding still image collections: 1000 is a common factor between them. If video images often have a weaker quality than still images (motion, fuzzy images...), they present a temporal redundancy which can be exploited to gain some robustness.

Video indexing is a large concept which covers different topics of research: video structuring consists of finding the temporal units of a video (shots, scenes) and is a first step to compute a table of contents of a video; key-event detection is more oriented to the creation of an index of the video; finally, all the extracted elements can be characterized with various descriptors: motion descriptors [63], or still image-based descriptors, but which can use the image temporal redundancy [76].

Many contributions have been proposed in the literature in order to compute a temporal segmentation of videos, and especially to detect shot boundaries and transitions [52], [69]. Nevertheless, shots appear to be a too low-level segment for many applications since a video typically contains more than 3000 shots. Scene segmentation, or what is called macro-segmentation is a solution, but it remains an open problem. The combination of media is probably an important axis of research to progress on this topic.

3.2.3. *Text Description*

Keywords: *Corpus-Based Acquisition of Linguistic Resources*, *Exploratory Data Analysis*, *Lexical Semantics*, *Machine Learning*, *Natural Language Processing*.

3.2.3.1. Lexical semantics for Information Retrieval

Automating indexing of textual documents [99] has to tackle two main problems: first choosing indexing terms, *i.e.* simple or complex words automatically extracted from a document, that "represent" its semantic content and make its detection possible when the document database is questioned; second, dealing with the fact that the representation *is* word-based and not concept-based. Therefore information retrieval has to overcome two semantic problems: various possibilities to formulate a same idea (how to match a concept in a text and a query expressed with different words); word ambiguity (a same word can cover different concepts). In addition to these difficulties, the meaning of a word, and thus the semantic relations that link it to other words, varies from one domain to another. One solution is to use domain-specific linguistic resources, both to disambiguate words and to propose equivalent formulations. These domain-specific resources are however not pre-existing and must be automatically extracted from corpora (collections of texts). Moreover, if one wants to use resources really adapted to one's text collection, prior to acquire them, one has to adopt a linguistic framework defining the semantic elements that are to be collected from corpora. In this respect, work in lexical semantics provides different theoretical models; let us cite three of them that are used in the TEXMEX project.

F. Rastier's differential semantic theory [96] is a linguistic theory in which the meaning of a word is defined through the differences it presents with the other meanings in the lexicon. Within a given semantic class —group of words that can be exchanged in some contexts—, words share *generic semes* (*i.e.*, generic semantic features) that characterize their common points and are used to build the class (*e.g.* the generic seme */to seat/* is associated with the class {*chair, armchair, stool*...}), and *specific* ones that explicit their differences (*/has arms/* differentiates *armchair* from the two others).

In J. Pustejovsky's Generative lexicon theory [95], a so-called *qualia structure* is defined. In this structure, words are described in terms of semantic roles. For example, the *telic* role indicates the purpose or function of an item (*cut* for *knife*), the agentive role its creation mode (*build* for *house*)... The qualia structure of a noun is mainly made up of verbal associations, encoding relational information.

The Meaning-Text Theory is a broad linguistic framework [90], whose lexicology part defines *Lexical Functions* [91] (LFs). LFs are designed to encode every semantic relation of a word, such as syntagmatic relations (*e.g. mouse–to click*, *shower–to have*...) or paradigmatic ones (*e.g. professor–student*, *bee–honey*...).

Concerning the corpus-based acquisition of lexical resources, 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 [41]. Relying on both methods, machine learning solutions are being developed in TEXMEX; they aim at being automatic and generic enough to give the possibility to extract, from a corpus, the kind of lexical elements required by a given task (for example, query expansion in an information retrieval system).

3.2.3.2. Characterization of Huge Sets of Thematically Homogeneous Texts

A collection of texts is said to be thematically homogeneous if the texts share some domain of interest. We are concerned by the indexing and analysis of such texts. The research of relevant keywords is not trivial: even in thematically homogeneous sets, there is a high variability in the used words and even in the concerned sub-fields. Apart from the indexing of the texts, it is valuable to detect thematic evolutions in the underlying corpus.

Generally, textual data are not structured and we must suppose that the files we are concerned with have either a minimal structure, or a general common thema. The method we use is factorial correspondence analysis. We get clusters of documents and their characteristic words.

3.2.4. Retrieval and Description Evaluation

Keywords: *Discriminating Power*, *Evaluation*, *Performance*.

The situation on this subject is very different according to the concerned media. Reference test bases exist for text, sound or speech, and regular evaluation campaigns are organized (NIST for sound and speech recognition, TREC for text in English, CLEF for text in various European languages, SENSEVAL or ROMANSEVAL for text disambiguation...).

In the domain of images and videos, the BENCHATLON provides a database to evaluate image retrieval systems while TREC provides test database for video indexing. A system to evaluate shot transition algorithms has been developed by G. Quenot and P. Joly [98].

Setting protocols of evaluation to compare different content-based information systems (CBIR) is a very hard task, especially when considering the relevance feedback from users who submit an image or a video as query-by-example to the CBIR system. In this context, our idea is to automatically learn user profiles during the searching scenarii and to correlate some feedback indicators (non-intrusively collected) with the sets of descriptors used in the query to compute the results. Finally, the objective is to adapt the next query execution or the image/video browsing, by dynamically taking into account the last feedback.

3.2.5. *Metadata Management*

Keywords: *Automatic Selection*, *Integration*, *MPEG-21*, *MPEG-7*, *Metadata*, *Metadata Management*, *Standard*, *TV-Anytime*.

To improve the data organization or to define the strategies to compute some descriptors, it may be advisable to use additional information, called metadata. Metadata (data about the data) must describe the data well enough as to be used as a surrogate for the data when making decisions regarding description and use of the data. Metadata can give complex information concerning structure description, semantics and contents of data items, their associated processes and, more widely, the respective domains of this various information.

Metadata are: i) data describing and documenting data, ii) data about datasets and their usage aspects, iii) the content, quality, constraints, and other characteristics of data.

The documenting role of metadata is fundamental. This information can provide decision elements in order to choose the most appropriate dataset or processing techniques and also, the most appropriate data presentation mode. In the case of large amounts of data, it is difficult to analyze data content in a straightforward way. Metadata then give appreciation or description informative elements of the dataset.

However, metadata role is not restricted to documenting information. Metadata must also allow:

- **Data acquisition and transformation** that are complex steps for data producers. Metadata can, on one hand, represent the production memory by describing operations carried out during data acquisition and transformation process, and it can, on the other hand, prevent a data producer from repeating the production step of an already existing dataset;
- **Description of structure and role of data**, in order to allow its interpretation and treatment by a user, especially during transfer steps.

V. Kashyap and A. Sheth [81] proposed a first classification of metadata for multimedia documents in two main classes: metadata which contain external information (date, localization, author...) and metadata which contain internal information directly dependent on the content (such as low-level descriptors) or describing the content independently (such as keywords annotations) [43], [54], [56], [72], [80]. Many standardized metadata such as in MPEG-7, TV-Anytime, *etc.* and also *ad hoc* content-descriptive metadata can be included in this classification [87], [89].

The key elements of the metadata managed by TEXMEX include (but are not limited to):

- Media description metadata (such as global descriptors —color, texture, motion, shape, *etc.*— or local descriptors) extracted from the images or from the bitstream and that are eventually formated as MPEG-7 or TV-AnyTime metadata or MPEG-21 Digital Item Declaration,
- Media usage metadata (such as relevance feedback in searching scenarii, access rights, availability, encryption, conditional access, *etc.*),
- User metadata (such as user preferences, usage history, *etc.*) and natural environment characteristics metadata (such as location, audio environment, illumination characteristics, *etc.*).

The selection and organization of metadata is highly application-dependent and also depends on the various objectives of metadata consumption that can facilitate: data access, data summary, data interoperability, media or content presentation and adaptation...

Metadata are a privileged way to keep information relative to a document or its descriptors in order to facilitate future processing. They appear to be a key point in a coherent exploitation of large multimedia databases.

3.3. Efficient Exploitation of Descriptors and Metadata

Keywords: *Data Analysis*, *Data Quality*, *Indexing*, *Statistics*.

Even if the description of the documents can be done automatically, this is not enough to build a complete indexing and retrieval system usable in practice. As a matter of fact, the system must be able to answer a query in a reasonable amount of time, and thus needs tools to guarantee this aspect. The section is devoted to some of these tools.

On-line and off-line processing define the two main categories of exploitation. On one hand, off-line processing corresponds usually to techniques which need to consider all the data, and the complexity in time is thus not the main issue. On the other hand, on-line processing needs to go really fast. To gain such a performance, these procedures use the result of the off-line processing to limit the treatment to the smallest data subset necessary to answer the query.

3.3.1. Statistics and Data Quality over Huge Datasets

Keywords: *Data Quality Metrics*, *Exploratory Data Analysis*, *Sampling*, *Statistics*.

The situation where we have few available data has been well studied but a huge amount of data generates different kinds of problems: for instance, the use of classical inferential statistics results in hypothesis testing concludes rather often to reject the null hypothesis. Besides, the methods of models identification fail very often or the quality of the model is overestimated. The question is: how can we set a representative sampling in such datasets? We must add also that some clustering algorithms are unusable with such large datasets. Therefore, it is clear that working with huge datasets is difficult because of their computational complexity, because of the data quality and because of the scaling problem in inferential statistics.

However, statistical methods can be used, with caution, if the data quality is good. So the first step is the cleaning and the checking of data to be sure of their coherence. The second step depends on our goal. Either we want to build a global model, or we are looking for hidden structures in the data. In the first case, we can work on a sample of the data and use methods such as clustering, segmentation, regression models. In case we are looking for hidden structures, sampling is not appropriate and we need to use other heuristics.

Exploratory data analysis (EDA) is an essential tool to deal with huge amounts of data. EDA describes data in an interactive way, without *a priori* hypothesis and provides useful graphical representations. Visualization methods when the dimension of the data is greater than three is also necessary: for instance, parallel coordinates. All these previous methods watch the data to discover their properties.

Let us add that most of the available data mining programs are very expensive, and that their contents are very disappointing and poor for most of them.

3.3.2. Multidimensional Indexing Techniques

Keywords: *Approximate Searches*, *Curse of Dimensionality*, *Databases*, *Multidimensional Indexing Techniques*, *Nearest-Neighbors (NN)*.

This section gives an overview of the techniques used in databases for indexing multimedia data (focusing on still images, however). Database indexing techniques are needed as soon as the space required to store all the descriptors gets too big to fit in main memory. Database indexing techniques are therefore used for storing descriptors on disks and for accelerating the search process by using multi-dimensional indexing structures. Their goal is mainly to minimize the resulting number of I/Os. This section first gives an overview of traditional multidimensional indexing approaches achieving exact nearest-neighbors searches. We especially focus on the filtering rules these techniques use to dramatically reduce their response times. We then move to approximate NN-search schemes.

3.3.2.1. Traditional Approaches, Cells and Filtering Rules

Traditional database multidimensional indexing techniques typically divide the data space into cells containing vectors. Cell construction strategies can be classified in two broad categories: *data-partitioning* indexing methods [45], [105] that divide the data space according to the distribution of data and *space-partitioning* [78], [104] indexing methods that divide the data space along predefined lines regardless of the actual values of data and store each descriptor in the appropriate cell.

Data-partitioning index methods, like the SS-Tree [105] or the SR-Tree [82], all derive from the seminal R-Tree [75], originally designed for indexing bi-dimensional data used in Geographical Information Systems.

Space-partitioning techniques like grid-file [93], K-D-B-Tree [97], LSD^h-Tree [78] typically divide the data space along predetermined lines regardless of data clusters. Actual data are subsequently stored in the appropriate cells.

NN-algorithms typically use the geometrical properties of cells to eliminate cells that cannot have any impact on the result of the current query [53]. Eliminating irrelevant cells avoids having to subsequently analyze all the vectors they contain, which, in turn, reduces response times. Eliminating irrelevant cells is often enforced at run-time by applying two rather similar *filtering rules*.

The first rule is applied at the very beginning of the search process and identifies irrelevant cells as follows:

$$\text{if } \text{dmin}(q,C_i) \geq \text{dmax}(q,C_j) \text{ then } C_i \text{ is irrelevant,} \tag{1}$$

where dmin(q,C_i) is the minimum distance between the query point q and the cell C_i and dmax(q,C_j) the maximum distance between q and cell C_j.

The search process ranks the remaining cells on their increasing distances to q. It then accesses the cells, one after the other, fetches all the vectors each cell contains, and computes the distance between q and each vector of the cell. This may possibly update the current set of the k best neighbors found so far.

The second filtering rule is applied to stop the search as soon as it is detected that none of the vectors in any remaining cell can possibly impact the current set of neighbors; all remaining cells are skipped. This second rule is:

$$\text{if } dmin(q,C_i) \geq d(q, nn_k) \text{ then stop,} \tag{2}$$

where C_i is the cell to process next, $d(q, nn_k)$ is the distance between q and the current k^{th}-NN.

Unfortunately, the “curse of dimensionality” phenomenon makes these filtering rules ineffective in high-dimensional spaces [104], [46], [94], [53], [83].

3.3.2.2. Approximate NN-Searches

This phenomenon is particularly prevalent when performing *exact* NN-searches. There is therefore an increasing interest in performing *approximate* NN-searches, where result quality is traded for reduced query execution time. Many approaches to approximate NN-searches have been published.

Dimensionality Reduction Approaches. Dimension reduction techniques have been used to overcome the "curse of dimensionality" phenomenon. These techniques, such as PCA, SVD or DFT [70] exploit the underlying correlation of vectors and/or their self similarity [83], frequent with real datasets. NN-search schemes using dimension reduction techniques are approximated because the reduction only coarsely preserves the distances between vectors. Therefore, the neighbors of query points found in the transformed feature space might not be the ones that would be found using the original feature space. These techniques introduce imprecision on the results of NN-searches which cannot be controlled nor precisely measured. In addition, such techniques are effective only when the number of dimensions of the transformed space become very small, otherwise the "curse of dimensionality" phenomenon remains. This makes their use problematic when facing very high-dimensional datasets.

Early Stopping Approaches. Weber and Böhm with their approximate version of the VA-File [103] and Li *et al.* with Clindex [85] perform approximate NN-searches by interrupting the search after having accessed an arbitrary, predetermined and fixed number of cells. These two techniques are efficient in terms of response times, but give no clue on the quality of the result returned to the user. Ferhatosmanoglu *et al.* [65] combine this approach with a dimensionality reduction technique:

it is possible to improve the quality of an approximate result by either reading more cells or by increasing the number of dimensions for distance calculations. Yet, this scheme suffers from the drawbacks mentioned here and above.

Geometrical Approaches. Geometrical approaches typically consider an approximation of the sizes of cells instead of considering their exact sizes. They typically account for an additional ε value when computing the minimum and maximum distances to cells, making somehow cells "smaller". Shrunk cells make the filtering rules more effective, which, in turn, increases the number of irrelevant cells. Cells containing interesting vectors might be filtered out, however.

The VA-BND scheme [103] empirically estimates ε by sampling database vectors. It is shown that this ε is big enough to increase the filtering power of the rules while small enough in the majority of cases to avoid missing the true nearest-neighbors. The main drawback of this approach is that the same ε is applied to all existing cells. It does not account for the very different data distributions possible in cells.

The AC-NN scheme for M-Trees [57] also relies on a single value ε set by the user. Here, ε represents the maximum relative error allowed between the distance from q to its exact NN and the distance from q to its approximate NN. In this scheme, setting ε is far from being intuitive. The experiments showed that, in general, the actual relative error is always much smaller than ε. Ciaccia and Patella also present an extension to AC-NN called PAC-NN which uses a probabilistic technique to determine an estimation of the distance between q and its NN. It then stops the search as soon as it finds a vector closer than this estimated distance. Unfortunately, AC-NN and PAC-NN cannot search for k neighbors.

Hashing-based Approaches. Approximate NN-searches using locality sensitive hashing (LSH) techniques [71] project the vectors into the Hamming cube and then use several hash functions such that co-located vectors are likely to collide in buckets. LSH techniques tune the hash functions based on a value for ε which drives the precision of searches. As for the above schemes, setting the right value for ε is key and tricky. The maximum distance between any query point and its NN is also key for tuning the hash functions. While finding the appropriate setting is, in general, very hard, it was observed [71] that choosing only one value for this maximum distance gives good results in practice. This, however, makes more difficult any assessment on the quality of the returned result. Finally, the LSH scheme [71] might, in certain cases, return less than k vectors in the result.

Probabilistic Approaches. DBIN [44] clusters data using the EM (Expectation Maximization) algorithm. It aborts the search when the estimated probability for a remaining database vector to be a better neighbor than the one currently known falls below a predetermined threshold. DBIN bases its computations on the assumption that the points are IID samples from the estimated mixture-of-Gaussians probability density function. Unfortunately, DBIN can not search for k neighbors.

P-Sphere Trees [73] investigate the trading of (disk) space for time when searching for the approximate NN of query points. In this scheme, some vectors are first picked from a sample of the DB, and each picked vector becomes the center of one hypersphere. Then, the DB is scanned and all the vectors that have one particular center as nearest neighbor go into the corresponding hypersphere. Vectors belonging to overlapping hyperspheres are replicated. Hyperspheres are built in such a manner that the probability of finding the true NN can be enforced at run time by solely scanning the sphere whose center is the closest to the query point. P-Sphere Trees can neither search for k neighbors.

To our knowledge, no technique linking the precision of the search to a probability of improving the result can search for k neighbors.

Rank Aggregation-based Approaches. Recently, Fagin *et al.* [64] proposed a framework for very efficiently evaluating single descriptor nearest-neighbor queries over high-dimensional collections. This framework is based on projecting the descriptors onto a limited set of random lines. Each random line is used to give a ranking of the database descriptors with respect to the query descriptor. These rankings are then efficiently aggregated to produce a fairly good approximation of the actual Euclidean k-nearest neighbors. The fastest algorithm to aggregate the rankings was called OMEDRANK.

The OMEDRANK algorithm has several nice properties: it is based on a cheap aggregation of rankings instead of a complex distance function; it uses standard B^+-trees to index the data, therefore handling updates gracefully; and it allows for a clever dimensionality reduction, by varying the number of random lines that are indexed.

4. Application Domains

4.1. Still Image Database Management

Keywords: *Digital Pictures*, *Image Databases*, *Medical Imagery*, *Photo Agencies*, *Text-Image Indexing*.

We are particularly interested in large image bases, like those managed by photo agencies. These agencies have between five hundred thousands and twelve millions of images. The Andia Press agency has a million of images, Sigma twelve millions, the Corbis agency which gathers the whole Bill Gate's acquisitions has thirty six millions of images. These agencies work according to two modes. In the first one, they respond to a customer's query by sending him/her a set of images. The customer pays for the images that he/she publishes. In the second mode, the customers are subscribers ant the agencies send them their new photographs, the mode of payment being identical. This working method is that of the AFP or Reuters.

One of the concerns of the agencies is of course the digital rights management, and the fact that they are not unduly used by people or institutions while not having discharged the rights. Watermarking and indexing are two techniques planned to control image diffusion, either by seeking a watermark of property in the images, or by checking by indexing that the image is not a fragment of an image of the agency base.

Another important field where the management of the images acquires an increasing importance is that of the medical images. The access to the medically interesting contents of the image is a true difficulty, so is the level of quality imposed by this field to the recognition system. The applications of content-based methods are thus still to come in this field.

Traditional image indexing consists in automatically extracting from an image numerical descriptors representing the color, texture, interest points or other similar information. However, such descriptors are not relevant to tackle the problem of a "semantic" querying of an image database: how can a customer find the pictures of a sunset or the pictures of his/her daughter learning to swim? How can an archivist in a news agency find a relevant picture to illustrate an article dealing with poverty in India? One way to address this problem is to make the most of existing documents in which images and texts appear together, and then use relevant parts of the texts to index the images; nonetheless, the definition of such text-image indexing schemes are up to now under study.

4.2. Video Database Management

Keywords: *Video Bases*, *Video Structuring*.

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 Franch TV channel TF1 switched to an entirely digitized production, the cameras remaining the only analogical stage of the production. 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.

Then, 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. If the effort of management of such collections will be probably weak, without rigorous method, there is a great need for tools to help the user: automatic creation of summaries and synopses to allow finding information easily, or to have within few minutes a general idea of a program. Even if the service is rustic, it is initially evaluated according to the appreciation which it brings to a system (video tape recorder, decoder), will have to remain not very expensive, but will benefit from a large diffusion.

On the other hand, are professional files: TV channel archives, registration of copyright, 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 documentalists, 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 cost. 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

Keywords: *Bibliography*, *Indexing*.

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 by the way 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.

4.4. Robotics and Visual Servoing

Keywords: *Planning*, *Robotics*, *Visual Memory*, *Visual Servoing*.

If collaboration between robotics and vision is an already old subject, it underwent an important change of paradigm in the five last years. Hitherto, collaboration was considered on the level of planning: a camera observed the world around a robot to enable it to plan its displacements. The results appeared to be not so satisfactory.

The field of collaboration then moved towards control: the vision is not any more used to plan a movement, but to ensure its follow-up and good execution, by setting up a closed loop of control including vision [62], [55], [88]. The results are completely promising and many industrial applications already exist.

Some difficulties remain: the tasks to be achieved are specified using a target image that should be reached, but that assumes that the robot is able to establish a link between this image and the current image provided by the camera. This is a classical image matching problem. If these two images do not have anything in common, it will be necessary to use a collection of intermediate images, which define intermediate positions of the robot before reaching the final position.

The control problem drives to an image collection management problem, with dynamic collections to follow the evolution of the environment of the robot, and needs for fast access for recognition. This application appears important because it widely opens the experimental use conditions of visual servoing: once an environment collected in a base, the robot can start from any position to go towards any target. If this kind of approach presents little interest for articulated arm for which the articular co-ordinates can be read directly, an autonomous vehicle can benefit from it in restricted environments such as car parks. In this case, the systems of positioning as the GPS do not offer sufficient relative precision and do not give information of orientation.

5. Software

5.1. Softwares

5.1.1. I-Description

Participant: Patrick Gros [correspondant].

I-DESCRIPTION allows a user to compute local or global image descriptors: differential local invariants, global and local color histograms or weighed histograms. It was deposited with the “Agence pour la Protection des Programmes” (APP) under the number IDDN.FR.001.270047.000.S.P.2003.000.21000.

5.1.2. Asares

Participant: Vincent Claveau [correspondant].

ASARES is a symbolic machine learning system (based on inductive logic programming) that automatically infers, from descriptions of pairs of linguistic elements (noun-noun, noun-verb...) found in a corpus in which the components are linked by a given semantic relation (synonymy, hyperonymy, qualia, lexical function...), corpus-specific morpho-syntactic and semantic patterns that convey the target relation. The patterns are explanatory and linguistically motivated, and can be applied to a corpus to efficiently extract resources and populate semantic lexicons. Two semi-supervised versions of ASARES also exist, that rely on a combination of the supervised symbolic pattern learner and a statistical extraction technique. These semi-supervised versions rival ASARES’s supervised one. ASARES was deposited with the “Agence pour la Protection des Programmes” (APP) under the number IDDN.FR.001.0032.000.S.C.2005.000.20900

5.1.3. Faestos

Participant: Pascale Sébillot [correspondant].

FAESTOS (Fully Automatic Extraction of Sets of keywords for TOpic characterization and Spotting) is a tool composed of a sequence of statistical treatments that extracts from a morpho-syntactically tagged corpus sets of keywords that characterize the main topics that corpus deals with. The system exploits the distribution of words of the corpus over its paragraphs, and requires neither human intervention nor given knowledge about the number or nature of the topics of the corpus. The extracted lists of keywords are employed in order to detect the presence of a topic in a paragraph, revealed by a keyword cooccurrence. Morevover, the system extracts from each keyword class a triple of words that permits an intuitive designation of the underlying topic. FAESTOS was deposited with the “Agence pour la Protection des Programmes”.

5.1.4. 2PAC

Participant: Pascale Sébillot [correspondant].

2PAC (2-Pass Acquisition of semantic Classes) brings together words used in a similar way in a topical sub-corpus such as that extracted by FAESTOS so as to build classes of words of similar meanings ("semantic classes") specific to the use that is made of them in that given topic. It works by first computing general semantic proximities from the whole corpus, then using that information to perform a more in-depth analysis of one of the topical sub-corpora. The result is a classification tree from which classes must be extracted manually; a graphical user interface has been developed to ease that task of manual exploitation. 2PAC was deposited with the “Agence pour la Protection des Programmes”.

5.1.5. tmx-DIVA-solution

Participants: Cédric Dufouil, Arnaud Dupuis [correspondant], Patrick Gros.

The library TMX-DIVA-SOLUTION (DIVA stands for *Di*stant *V*ideo *A*ccess) enables an easy access to a collection of videos stored on a remote machine. It contains a server on the remote machine, in charge of extracting from video files the image or sound data needed by the user and sending it to the client. The client part of the library decodes what it receives from the server and provides the user with the exact data needed (the exact image or sound information). A caching system allows the system to transfer a GOP only once when several neighbouring images are queried by a user. The server and the client can work on different systems (Windows, Linux or MacOSX). TMX-DIVA-SOLUTION was deposited with the “Agence pour la Protection des Programmes” (APP) under the number IDDN.FR.001.320006.000.S.P.2006.000.40000

5.1.6. NaviTex

Participants: Cédric Dufouil [correspondant], Arnaud Dupuis, Patrick Gros.

The NAVITEX tool was developed with the help of TMX-DIVA SOLUTION. It allows a user to index and annotate videos using TV-Anytime (XML) metadata format. Those metadata can be exported to the DIVA MANAGER, in order to easily access some parts of a video directly from the web page. NAVITEX runs on Linux and Windows OS. NAVITEX is in the process of being deposited with the “Agence pour la Protection des Programmes” (APP).

5.2. Experimental Platform

Participants: Laurent Amsaleg, Patrick Gros, Arnaud Dupuis, Cédric Dufouil, Pascale Sébillot.

Until 2005, we used various computers to store our data and to make our experiments. In 2005, we began some work to specify and set-up dedicated equipment to the experiments on very large collections of data. During 2006, we specified our first complete platform, which is organized around a very large storage capacity (between 64 and 100Tb), and containing acquisition devices (for Digital Terrestrial TV), computing servers and data servers. Two softwares were developed: one to manage the acquisition process and all the data already stored, the second one to allow an easy access to the data from processing programs (tmx-DIVA-solution). The platform will be completed with dedicated software to manage large collections of still images and all the metadata associated with the data. This platform is paid by a joint effort of INRIA, INSA Rennes and University of Rennes 1.

6. New Results

6.1. Image Retrieval for Large Databases

Our work on image description does not aim at finding new general descriptors. The IMEDIA and LEAR teams are very active in this field, and we use their results. The originality of our work comes from the size of the database we want to handle. In large databases, most images will be compressed. Is it possible to describe an image without decompressing it? Without sticking too tightly to the JPEG’2000 format, we try to find new description schemes based on wavelet decomposition of images. This is our first direction of research.

A second direction concerns the combination of descriptors: when documents are described by many descriptors, how a query should be processed in order to provide an answer as fast as possible? To answer this question, we study the information that each descriptor can provide about the other ones. The aim is to determine the order in which the descriptors should be considered by using data mining techniques applied to visual descriptors.

The third direction tackles the problem of indexing and retrieving large collections of descriptors. In the local description scheme, 1 million of images can give raise to 600 millions of descriptors, and retrieving any information among such an amount of data requires really fast access techniques, whatever the aim of this access.

A fourth direction is due to our collaboration with the roboticians of the LAGADIC team. They work on visual servoing and using a database is a good way to improve the applicability of their techniques to large displacements. Our description technique appears to be particularly well suited to such an application where a matching between images is required, and not only a global link of similarity between images.

6.1.1. Image Description, Compression and Watermarking

Keywords: *Image Compression*, *Image Description*, *Image Indexing*.

Participants: Patrick Gros, François Tonnin.

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

During the last two decades, image representations obtained with various transforms, *e.g.*, Laplacian pyramid, separable wavelet transforms, curvelets and bandlets have been considered for compression and de-noising applications. Yet, these critically-sampled transforms do not allow the extraction of low level signal features (points, edges, risdges, blobs) or of local descriptors. Many visual tasks such as segmentation, motion detection, object tracking and recognition, content-based image retrieval, require prior extraction of these low level features. The Gaussian scale space is almost the unique image representation used for this detection problem. Management of large databases are therefore uneasy, as the extraction of features requires first to decompress the whole database and then convert the images in the Gaussian scale space. It is thus desirable to find representations suitable for both problems: compression and signal feature extraction. However, their design criteria are somewhat antagonist. Feature extraction requires the image representation to be covariant under a set of admissible transformations, which ideally is the set of perspective transformations. Reducing this set of transformations to the group of isometries, and adding the constraint of causality, the image representation is uniquely characterized by the Gaussian scale space. In a compression perspective, one searches to reconstruct the image from a minimal amount of information, provided by quantized transform coefficients. Thus, the image representation should be sparse, critically-sampled (or minimally redundant), and transform coefficient should be as independent as possible. However, critically-sampled representations suffer from shift-variance, thus are not adapted for feature extraction.

This year, the collaboration between TEMICS (Christine Guillemot) and TEXMEX (Patrick Gros) through the thesis of François Tonnin came to an end with the defense of the thesis in June. This work has led to the design of a feature point extractor and of a local descriptor in signal representations given by the over-sampled steerable transforms. Although the steerable transforms due to their properties of covariance under translations and rotations, and due to their angular selectivity, provide signal representations well-suited to feature point and descriptor extraction, the opposite constraints of image description and compression were not fully solved.

The final problem is rarely addressed in the literature and consists in the proper quantization of transformed coefficients for both good image reconstruction and preservation of description quality. As the transform is redundant, one image has many possible representations. We first used POCS (Projection onto Convex Sets) to find a sparser representation and we adapted the classical technique in order to preserve the content of the neighborhoods of extracted points. Then, we designed a compression scheme allowing the reconstruction of steerable coefficients from the information required by description, which is reduced to an energy and an orientation coefficient for each spatial point. The final step is the quantization of these coefficients. This compression scheme allows us to detect illegal copies in image bases compressed at one bit per pixel.

Videos appear to be the next challenge. On the one hand, HDTV represents sets of images even bigger than what is present in most still image collections. On the other hand, some functionalities are required by the professionals of the domain in order to develop their services: scalable coding to allow an easy distribution of the content on many platforms, copy detection as a complementary tool to DRM. These aspects form the core of the ICOS-HD project that will begin in 2007.

6.1.2. Scalability of Local Image Descriptors

Keywords: *Local image descriptors*, *PvS-framework*, *high-dimensional indexing*, *median rank aggregation*, *scalability*.

Participants: Laurent Amsaleg, Hervé Jegou.

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

With the proliferation of digital media and online access, multimedia retrieval and image retrieval, in particular, is growing in importance. The computer vision community has recently started a trend towards advanced image description schemes using *local descriptors* (*e.g.*, SIFT). The applications of local descriptor schemes include face recognition, shape recognition and image copyright protection. With these schemes, each image yields many descriptors (several hundreds for high-quality images), where each descriptor describes a small "local" area of the image. Two images are typically considered similar when many of their descriptors are found to be similar.

All of these approaches, however, have only been studied and compared at a small scale. Typically, less than 10,000 images are used, which makes it hard to predict how they will perform with collections of hundreds of thousands of images or more.

There are three primary issues associated with scalability of image descriptors.

- First, it is key to answer queries efficiently, even with very large image (and descriptor) collections. This issue demands efficient database support techniques.
- Second, it is necessary to ensure effectiveness; that queries return useful results from very large collections. Note that studying effectiveness at a large scale requires first having efficient database support.
- Finally, the descriptors must be created efficiently, in particular in a high-throughput environment.

That work addresses all three issues.

In [84], we have demonstrated that our PvS-framework achieves efficient query processing for large collections of local descriptors. We therefore decided to compare three major local descriptor schemes (SIFT, PCA-SIFT and RDTQ) to study their recognition power at large scale. This comparison included a fourth scheme that we designed, and called Eff^2. Using a collection of almost thirty thousand images, we showed that our new descriptor scheme gives the best results in almost all cases. We then gave two stop rules to reduce query processing time and show that in many cases only a few query descriptors must be processed to find matching images. Finally, we test our descriptors on a collection of over three hundred thousand images[1], resulting in over 200 million local descriptors, and show that even at such a large scale the results are still of high quality, with no change in query processing time [32], [31].

[1] We have signed, in December 2005, a formal agreement of cooperation between IRISA, Reykjavík University and Morgunbladid, the main newspaper in Iceland. This agreement defines the terms under which Reykjavík University and the TEXMEX team can obtain access to the image collection of Morgunbladid while Morgunbladid will have use of PvS software. This collection consists of about 300,000 high-resolution images. The images were delivered to us after being thumbnailed to 512x512 pixels, which is sufficient for performing extensive recognition-based performance measurements. We can keep the images for two years. Then, we have to destroy them. Their descriptions, however, can be kept as long as needed for research and development purposes, since this format does not allow for any presentation or reconstruction of the images. It is extremely difficult to get access to real image collections, and signing this aggreement gave us a real push since we were able to conduct a series of experiments at a scale never reached.

Based on the experience gained with the PvS-framework, we have designed a more sophisticated and general index which is also based on ranking, projections and partitions. This index is called the NV-tree (Nearest Vector tree) and we are in the process of patenting it. With respect to the PvS-framework, the NV-tree yields better performance and space utilization, is better able to capture the real distribution of data by self-tuning the projection and partitioning strategies, copes with on-the-fly updates of the descriptor collections, can be used stand-alone or by aggregating the results from two or more indices, and lends itself effectively to distributed processing to further reduce response times. All in all, the NV-tree yields efficient query processing and good result quality with extremely large descriptor collections.

6.1.3. Describing Sequences for Audio/Video Retrieval

Participant: Laurent Amsaleg.

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

People working in such organizations typically want to annotate incoming video or audio streams before archiving. Those annotations are then used by any subsequent search since they are at the roots of document matching. It is key to note that document annotation is an entirely manual process and to understand that this process can not scale with the constantly increasing number of streams to annotate. Therefore, one salient application is the automated segmentation of multimedia streams into separate units, then the automatic annotation of each unit, before archiving the documents. It is thus necessary to perform searches in streams to detect for example jingles, trailers, or the periodic broadcast of elements, etc. Those searches are more complex then searching simply for the repetition of identical patterns since it is necessary to find correlations despites distortions, duration variations, super-imposition of noise, text, additional music, inclusion of multiple side-streams, etc.

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

One of the most prevalent difficulties comes from the temporal aspect of video and/or audio descriptions. Describing video and audio means creating sequences of descriptions in which the notion of order between descriptions is central. That notion of order is ignored by all traditional search techniques that only search for independent elements that are, at most, very loosely coupled.

We therefore try to understand how multidimensional indexing techniques can integrate in their principles the notion of sequences of descriptions. This needs to be done to make possible searches by content in very large archives of video and/or audio documents. We have started to work on this topic. We have implemented few techniques from the state of the art (exhaustive search, dynamic time warping, mixture of Gaussian models and SVM-based modelling) and ran performance evaluations on audio recognition. Using a collection of real audio sample, we checked the ability of each technique to handle recognition despite time shifts, time distortions and some other signal distortions. It turns out that SVM-based models perform quite nicely but are very inefficient in terms of response time. This open room for improvement.

6.1.4. Navigation in Personal Image Collections

Participant: Laurent Amsaleg.

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

In recent years, the world has seen a tremendous increase in the capability to create, share and store digital images. As a result, personal image collections are growing at an astounding rate and it is clear that in the future individuals will need to access tens of thousands, or even hundreds of thousands, of digital images. It is therefore imperative to start studying ways to access these images in a useful and interesting manner. What is needed is software that will allow users to seamlessly organize, search and browse their images.

In many households, organizing a home photo collection has long been a neglected task. This is still true even with the latest digital photo browsers that typically simply dump pictures into folders, an electronic version of the good old shoe-boxes our parents were using for paper-printed pictures. They offer no support for browsing and searching by image contents, and therefore are inadequate for handling such large collections. Despite numerous features (effective packing on thumbnails on screen, identifying representative images, zoomable user interfaces...), all current photo browsers share limitations such as using a time-line view or a folder view at each time, failing to use the two dimensions of the screen. Most have clumsy annotations capabilities and more than anything else completely separate the search and browsing functions. This key flaw is not unique to image browsers: on the Web, browsing is clicking hyperlinks while searching is through Google or others, typically returning a flat list of results from which browsing can start. Overall, presentation is typically linear and the contents of the images are not used to guide the search and presentation.

Each image may be described by a number of attributes, based on image contents and image meta-data (such as camera and time information, stored in so-called EXIF headers). Some of these attributes may be linear or spatial, such as time and location of taking the image, while others may be textual, hierarchical or categorical. These attributes may be considered dimensions in an image hyper-space, which we must be able to traverse dynamically to fully enjoy our digital images. In on-line analytical processing (OLAP), multi-dimensional data is dealt with by considering a few dimensions at a time and pivoting between dimensions when necessary. In advanced computer games such as EVE online, large three-dimensional worlds are explored by simulating space-travel. Both approaches have been very successful in keeping their users occupied and focused on their task for a long time. We propose that a browsing interface for images should merge these features into a multi-dimensional interface that allows flexible space-travel like exploration of the image hyperspace. In order to begin exploring the possibilities of such a browsing interface we have implemented a prototype, based on the PartiView browser, which allows us to browse images in a three-dimensional space. The dimensions may be based on image contents and image meta-data and different dimensions may be combined in an arbitrary manner. Our conclusion is that while the prototype has shortcomings, this is a very promising research direction that merits further exploration. What is novel in this work is that we want to integrate to an image browser OLAP browsing concepts, such as pivoting and filtering that have typically been designed to facilitate the browsing of huge financial data collections.

6.1.5. General Frame for Clustering High Dimensional Datasets using Random Projections

Participants: Laurent Amsaleg, Zied Jemai, Annie Morin.

Random Projections (RPs) have recently emerged as a powerfull method for dimension reduction. Random projections is computationally significantly less expensive than the other techniques and preserves distances quite nicely. It has been shown that RP yields results comparable to conventional dimensionnality reduction techniques [42], [48], [58]. RPs, however, are highly unstable, which becomes problematic when used for data clustering purposes.

We made an experimental observation showing that data points belonging to a "natural" cluster are very likely to be clustered together in many possible different clusterings. Multiplying projections and segmentations of the data space onto many lines gives many different clustering of the points. Note this method much relies on our previous PvS work [84]. Therefore, it is possible to create a matrix of co-association of points across all the produced clusterings. The rational of our approach is to weight associations between data points by the number of times they co-occur in a cluster. Clusters are produced by independent runs of RP-based clusterings. This matrix becomes the support for determining consistent and stable clusters, subsequently refined using a classical hierarchical approach.

We applied this RP-based clustering framework on 4 different databases with different dimensionality and cardinalities (up to 43 million 128-d descriptors). So far, this method seems to be scalable and can address easily the large-scale clustering problem in data-mining.

6.1.6. Intensive Use of Factorial Analysis for Image Mining

Keywords: *Correspondance Analysis*, *Visualization*.

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

We need to define "words" in images to use the same tools for image analysis as we do in textual data analysis. We use local descriptors to describe still images. For each image, we get a large number of descriptors lying in a high dimension space. The first step is to cluster the descriptors and replace each descriptor by its cluster number. We build a frequency table crossing the images and the clusters. In a cell (i,j), we have the number of descriptors of the image i belonging to the cluster j. To make the link with textual analysis, an image is equivalent to a document while a descriptor cluster is equivalent to a word. We then use factorial correspondence analysis to process the frequency table and to get groups of "words" defining a concept and related to the previous concepts, groups of images very closed to each other. This method is very similar to the one used by Andrew Zisserman *et al.* [101]. The work is just starting with the Ph.D. thesis of N. K. Pham.

6.1.7. Coupling Action and Perception by Image Indexing and Visual Servoing

Keywords: *Robot Motion control*, *Visual Servoing*.

Participants: Patrick Gros, Anthony Remazeilles.

This is a joint work with the LAGADIC team (F. Chaumette).

Our work aims at developing an integrated approach combining image recognition to allow a robot to localize itself with respect to an image collection depicting its environment and image servoing to control the motions and displacements of the robot.

This year, the work of Anthony Remazeilles was mainly dedicated to the implementation of the system developed on the Cycab platform (see Lagadic project annual report). Although the system worked well in indoor environments, outdoor scenes appeared to be problematic for the image recognition module. The lighting conditions result in big photometric differences between the images of the database and images to be recognized that are beyond the robustness capacity of the descriptors used.

Since we have now the possibility to manage very large image collection without additional cost in terms of online retrieval time, the solution is probably to add much more images in the database, images which should correspond to the various illumination conditions encountered by the robot (morning, mid-day and evening images, sunny and cloudy images...). This aspect will constitute our contribution to the collaboration in the future.

6.2. Text Retrieval for Large Databases

6.2.1. Natural Language Processing and Machine Learning

Keywords: *Corpus-Based Acquisition of Lexical Relations*, *Information Retrieval*, *Machine Learning*, *Natural Language Processing*.

Participants: Laurent Amsaleg, Vincent Claveau, Antoine Doucet, Fabienne Moreau, Pascale Sébillot, Laurent Ughetto.

Our research in this field focuses on three points: elaboration of fully automatic and generic machine learning solutions —using both symbolic and statistical approaches— to extract from textual corpora linguistic resources needed by a given application; exploitation of lexical resources in information retrieval systems, trying to discover accurate solutions to the problem of their integration into the systems; quest for fast systems in spite of this integration.

During 2006, our work has especially concerned the 3 following aspects:

1. **Machine-learning based acquisition of morphological variants of words**.

   Information retrieval systems (IRSs) usually suffer from a low ability to recognize a same idea that is expressed in different forms. A way of improving them is to take into account morphological variants (such as `compile/compilation/recompiling`...). We proposed a simple yet effective method to recognize these variants that are further used so as to enrich queries. It is based on an unsupervised machine learning approach using analogies to decide if two words are morphologically related. In comparison with already published methods, our system does not need any external resources or *a priori* knowledge, and thus supports many languages. This new approach has been evaluated against several IR collections, 6 different languages. Reported results show a significant and systematic improvement of the whole IRS efficiency both in terms of precision and recall for every language [18].

2. **Linguistic resources and information retrieval (IR)**.

   Our aim in this domain is to explore methods that enable information retrieval systems to capture the semantics of natural language texts, and to exploit the semantic information that natural language processing (NLP) techniques can automatically extract from textual documents. Currently, for example, systems based on Salton's vector space model (VSM) represent documents and queries as sets of words, without regard for the relationship (or even linear order) between them. Fabienne Moreau's Ph.D. thesis [13] aims at adapting current IR models to allow information gleaned from NLP to inform IR. This year has been dedicated to the following points:

   – elaboration of a prototype of linguistically informed-IRS, based on VSM, to integrate in parallel multiple kinds of linguistic knowledge, belonging to the morphological, syntactic, or semantic levels of language. This tool has enabled us to demonstrate the effective interest of several sorts of knowledge (especially morphological and semantic);

   – through an original analysis of correlations between the various linguistic index in terms of complementarity and redundancy to retrieve documents, bringing to light of the relevance of some mono-level and some multi-level index couplings;

   – in order to automatically detect the best way to combine those informations within an IRS, proposition of a supervised machine-learning technique (based on a neural network) that merges the lists of documents produced by each linguistic index, and automatically adapts its behavior to the characteristics of the queries. This combination results in higher performances than those obtained by the most efficient single index, especially in terms of stability;

   – adaptation of our previously-described method for the acquisition of morphological variants to IRS. Using its productions to expand queries has led to high scores on various collections of documents, and several languages.

3. **Efficient but rapid IR systems**.

   Concurrently to this aim of efficiency, we also study a second aspect of the quest of performances for IRS: systems have to be relevant and to provide better answers, but they also have to be rapid and to answer quickly to users, even when they are questioning huge textual databases. Antoine Doucet has joined TEXMEX in October 2006 for a one-year Inria post-doctoral position; the research aims at finding a way to merge the models that are the most suitable to integrate linguistic information and the best search algorithms developed within the scientific field of text databases (TDB). In this domain, very efficient algorithms for IR on huge textual databases have been developed, that however use over-simplified models of linguistic knowledge.

6.2.2. *Intensive Use of Factorial Analysis for Text Mining: Indicators and Displays*

Keywords: *Correspondance Analysis*, *Visualization*.

Participant: Annie Morin.

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

6.2.3. *Visualization and Web Mining*

Keywords: *3D Metaphors*, *Human-Computer Interfaces*, *Information Retrieval*, *Interface Evaluation*, *Search Result Visualization*, *Self-Organizing Maps*, *Visual Categorization of Web Pages*.

Participants: Nicolas Bonnel, Annie Morin.

This work deals with the dynamic generation of interactive 3D presentations of web search results. Here the issue is how to effectively represent the results matching a query on textual search engine. These researches were carried out in the context of a thesis in cooperation with the R&D division of France Télécom (CIFRE grant) [10].

Textual information retrieval is one of the main tasks related to the Web. This task relies generally on search engines which have become an essential tool of the Web. Indeed users start at a search engine 88% of the time when they have a new task to complete on the Web. However, confronted with the huge increase of information available on the Web and the lack of significant evolution of the search process, the amount of documents matching a query becomes awfully important. It is therefore difficult for the user to effectively interpret all these results. This problem of representing search result is addressed through information user interfaces (IUI). Much work has been carried out on search result visualization since a decade, without any real impact on the most popular user interfaces. Our approach focuses on the necessity to dynamically create interactive 3D presentations based on visualization metaphors adapted not only to the end-user, but also to the task to complete as well as to the data. Two main steps can be distinguished in our work [24].

The first one consists of organizing effectively the results of a web query. For this purpose, on-the-fly clustering methods are investigated and only statistical and deterministic approaches are considered. More precisely we focus on a particular unsupervised clustering method: the self-organizing maps. This method enables to cluster the results *and* to organize the clusters thanks to a non-linear projection on a predefined map. To correctly adapt this method to our context, we discuss on many points such as the distances to use or the weighting schemes to apply. Another important point is the quality evaluation of the proposed classification and organization.

The second step concerns the visualization of the organized search results. The goal is to define cognitive 3D metaphors of visualization allowing for a richer space representation which efficiently and effectively helps users in their tasks. Various interactive and adaptive metaphors are then proposed but the main one is based on the city concept. This metaphor enables the user to interact with web search results which are represented in a 3D virtual city and organized on the city ground according to the computation of the self-organizing map. All the proposed interfaces are hybrid (*i.e.* composed of a 2D part and a 3D scene) and integrated in the SmartWeb prototype which uses the X-VRML language that enables to effectively design 3D visualization metaphors and automatically generate interactive 3D content.

A user study of this interface was carried out and an evaluation framework for IUI is proposed in order to be able to successfully evaluate this kind of interface [22], [21]. Some key points for designing successful 3D metaphors are also discussed [49].

Finally our approach can be considered as a post-processing step for a search engine. Obviously the necessity to have an effective indexing and an effective retrieval process is a complementary work to this one [50].

6.3. Multimedia Document Description

Keywords: *Multimedia Description*.

The term multimedia documents is broadly used and covers in fact most documents. It is in fact more and more appropriate since any document are now truly multimedia and contain several media: sound, image, video, text. The description of these documents, videos for example, remains quite difficult. Research groups are often monodisciplinary and specialist of only one of these media, and the interaction between the different media of a same document is not taken into account. Nevertheless, it is clear that this interaction is a very rich source of information and allows us to avoid the limitations of the techniques devoted to a single medium since the limits vary according to the concerned medium.

The combination of all the media remains a difficult objective. Each pair of media present specific difficulties. On the other hand, as few studies have been published in this domain, many of the difficulties are still to be discovered. To handle this fact, we decided to study particular combinations of media for problems of reduced size. The four ongoing studies concern the audio-video combination for both fine grain and coarse grain segmentation of videos, the description of documents containing text and images, and the collaboration between speech transcription and natural language processing.

This work is done in close collaboration with the METIS group of IRISA which brings its knowledge of the sound and speech media as well as in stochastic modeling.

6.3.1. Stochastic Models for Video Description

Keywords: *Dynamic Bayesian Networks*, *Hidden Markov Models*, *Image-Sound Interaction*, *Segment models*, *Video Structuring*.

Participants: Siwar Baghdadi, Manolis Delakis, Ewa Kijak, Patrick Gros, Pascale Sébillot.

Our work on this topic is done in close collaboration with the METISS team of IRISA and Thomson as external partner.

Our first axis of research on multimedia description concerns the structuration of sport videos. M. Delakis's thesis was devoted to the use of Segment Models (SMs) to extract the dense structure of tennis broadcasts. Three years ago, E. Kijak's thesis showed the interest of multimodal integration for this kind of videos, but the Hidden Markov Models (HMMs) used at that time was lacking of the needed flexibility. The aim of our work was thus to test a more general model. In order to extract individual events from a video, we investigate the use of Dynamic Bayesian Networks in the frame of S. Baghdadi's thesis.

6.3.1.1. Segment models for dense tennis broadcast structuring

The aim of this study is the automatic construction of the table of contents of a tennis broadcast. In this type of video, game rules as well as production rules define a type of tennis video syntax that can be modeled by Markovian models and parsed with dynamic programming techniques. The video is thus segmented into scenes, which are video portions that share a unique semantic content and serve as the basic building blocks of the table of contents of the video.

Motivated by the need for more efficient multimodal representations, the use of segmental features in the framework of Segment Models was proposed [60], [59], instead of the frame-based features of Hidden Markov Models. Considering each scene of the video as a segment, the synchronization points between different modalities are extended to the scene boundaries. Visual features coming from the produced broadcast video and auditory features recorded in the court view are processed before fusion in their own segments, with their own sampling rates and models. On video only data or with synchronous audiovisual fusion, SMs demonstrated a performance improvement over HMMs, traded with a negligible extra computation cost. Possibilities of asynchronous fusion of auditory models with SMs were also examined, even though performance did not finally improve.

Beside the audiovisual integration, we are also concerned with the fusion of high-level information coming from the game structure and rules and from the score announcements [25]. The first dictates that the solution provided by Viterbi decoding should be consistent with the tennis rules and the second one that it must be consistent with the actual game evolution, as it is recorded in the score announcements. The incorporation of tennis rules is easily performed by extending the flat topologies of both HMMs and SMs to hierarchical ones, in order to follow the hierarchical decomposition of a tennis game. Hierarchical topologies also allow for the detection of high-level segments in the video, like a set or a game. Unfortunately, the number of transition probabilities in a hierarchical topology is very high and thus they have been manually set to arbitrary values. For this reason, the hierarchical topologies did not outperform the flat ones.

Regarding the score announcements, their fusion as an extra feature (or observation) at the shot level (HMMs) or at the scene level (SMs) was firstly considered. In both case, a marginal performance improvement was noticed. The reason is that score announcement may appear a lot delayed or may not appear at all after a game event. The probability distributions of this feature become thus almost uniform, carrying no useful information. The Score-Oriented Viterbi Search was introduced and used instead, in order to fully exploit the semantic content of the score announcements and their positions as well. The algorithm finds out the most likely path that is consistent with the score announcements in the expense of a computation overhead a little superior to the standard Viterbi decoding for both HMMs and SMs. The fusion of the score announcements as such yielded a clear performance improvement in both HMMs and SMs and with both flat or hierarchical topologies.

Finally, we explored the idea of using a SM-RNN hybrid, *i.e.*, the use of Recurrent Neural Networks as a segmental scorer. To this end, the newly introduced LSTM topology was favorably compared to BPTT-trained RNNs and used in the hybrid. The hybrid performed however visibly inferior to the standard SM but still with a promising performance. In fact, what makes the difference is that the HMM-based segmental scorers can use prior knowledge on the task directly into their topology, while the LSTM scorers were built from scratch.

Future work involves firstly the addition of useful information in the feature set, like the outcomes of a player tracking process. Tennis rules dictate that the players have to change position between successive exchanges. Tracking players between successive exchanges can thus convey helpful information. We also plan to extend the framework of SMs to other genres of video where structure analysis is required. News broadcasts can be such a video genre, defining a segment as a news story. Information sources from multiple modalities, like image, sound and text could be asynchronous between them but, by definition, they are synchronous inside the story unit boundaries.

6.3.1.2. Dynamic Bayesian Networks for video event detection

The aim of this work is to study the additional flexibility brought by DBN with respect to what is possible with Segments Models. The first application considered is advertisement detection. Like Segment models, DBN should allow us to preserve the dynamics of each modality while taking better into account the dependance between these modalities. It also allows us to learn automatically the correlations between random variables, and thus to learn the structure of the network. But the possibility to use such a learning in practical situations has still to be investigated. Our first work consists in testing how the temporal aspect of our problem can be handled using DBNs.

6.3.2. Automatic Structuring and Labeling of Television Streams

Keywords: *Digital TV*, *Video Mining*, *Video Stream*.

Participants: Patrick Gros, Gaël Manson, Xavier Naturel.

Our work on this topic is done in collaboration with INA and the Metiss team (IRISA).

This topic investigates the possibility of automatically structuring large television streams. Structuring means finding the structure of the TV stream at a relatively coarse level, *e.g.* extracting the succession of programs in a large corpus of weeks or months of TV. This is of major interest for television archives, which are usually not structured nor documented, thus making it very difficult to retrieve any information. Other applications of structuring include monitoring or consumer-oriented TV services. Commercial monitoring is already a mature subject, a few solutions from industry based on watermarking or fingerprinting indeed exist. Apart from this classical commercial monitoring application, our solution enables to monitor a channel in its entirety: computing the amount of commercials vs. programs, checking if programs are scheduled on time... This is of interest for institutions like CSA (Conseil supérieur de l'audiovisuel) which in France monitors channels to detect potential frauds to the French law.

Our previous work enabled us to segment a TV stream into programs and to tag them with labels coming from a TV program guide. The program guide indeed provides a rough schedule of the programs to come and albeit errors it contains very useful information for tagging the segmented stream. The link between the extracted segments and TV program guide information is done using a Dynamic Time Warping algorithm (DTW), which computes a global alignment between the segmentation and the program guide.

Extensions of this simple algorithm have been proposed, in particular to take into account information coming from a shot recognition process. Such a recognition which aims at detecting repeated video segments provides "synchronization points" called anchors, which constrain the path of the DTW algorithm, and thus provide a nice way to include side information. A new post-processing algorithm has been designed for resolving conflicts in the case where labels proposed by the DTW and the recognition process differ. It is based on a simple yet effective Bayesian decision criterion.

The problem of the video database update has also been investigated. As a matter of fact, the database used in the recognition process needs to be updated in order to keep correct labeling results over time. We have shown that a naïve approach actually decreases the labeling results due to over-segmentation by trailers. A more subtle approach was then proposed which structures and eventually labels the new segments before including them in the database. As a result, with carefully selected segments, an automatic update of the database is possible and the results remain consistent over a three weeks period.

6.3.3. Image and Text Joint Description

Keywords: *Image-Text Interaction*.

Participants: Vincent Claveau, Patrick Gros, Emmanuelle Martienne, Pierre Tirilly.

In text retrieval engines, images are not taken into account; conversely, image retrieval engine treats the images independently of the text surrounding them. Of course, it should be better to couple these two engines or, at least, to couple the information that both media can provide.

The first way to reach this goal is to determine the parts of the texts which are related to images. This could lead to textual descriptions of images, and thus to the possibility of textual queries to retrieve images, in a much richer way than what is currently offered by systems using simple keywords associated with the images. Moreover, it is possible to find two documents containing a same image and to use both texts to disambiguate or improve the understanding of the text.

Our first work in this framework was to build up a huge corpus of news articles containing texts and images. Moreover, in this corpus, each image is associated with a brief description given by news agencies documentalists describing the content of the image. This corpus allowed us to test, in "real-world" conditions, the relevance of simple approaches trying to combine visual and textual pieces of evidence to propose a meaningful description of images. Indeed, in our first experiments, we jointly used a face detection system and named entity recognition system to annotate images representing a person with its name. Very good results are obtained, both in terms of precision and recall.

6.3.4. Text and Speech Joint Description

Keywords: *Text-Speech Interaction*.

Participants: Vincent Claveau, Stéphane Huet, Pascale Sébillot.

A lot of sound documents contain speech. Speech recognition is used to automatically index and exploit them, particularly when the data to analyze are voluminous. Automatic speech recognition (ASR) systems produce a textual transcription from sound documents, thanks to acoustic and linguistic clues. However, most of current ASR systems are more based on purely statistical methods (study of n-grams in corpora) than on linguistics. Stéphane Huet's Ph.D. thesis (under the double supervision of Guillaume Gravier, from the METISS Project-Team, and Pascale Sébillot) aims at using Natural Language Processing (NLP) techniques to help the transcription of sound documents.

This year, after publishing a state-of-the art of the use of linguistic knowledge in ASR systems as an internal report [39], we have focused our interest on the study of the integration of parts of speech (PoS) (categorial tags identifying a word as a noun, a verb, an adjective, *etc.*, and at the singular or plural form, conjugated, *etc.*) to improve the quality of the output of ASR systems. Our choice is mostly motivated by the following reasons:

- PoS can be reliably obtained from automatic NLP tools. The main difficulties we are faced with when NLP techniques are applied to spoken documents are their lack of structure, since there are no paragraphs or even sentences, and the noise of automatic transcriptions, as the hypotheses made by ASR systems may include many erroneous words. These phenomena complicate the use of a high-level syntactic analysis but only slightly disturb PoS tagging. With only minor adjustments to the specificities of spoken documents, our experiments showed than more than 91% of words of automatic transcriptions are correctly tagged, while the usual results for written documents are of about 95%;
- PoS tagging brings interesting information to correct errors commonly made by ASR systems. For instance, in the French language, the plural and singular forms, as well as the masculine and feminine ones of many words, are homophones and PoS tagging brings new information to discriminate them;
- disambiguating words according to their PoS is usually necessary before extracting other information from a text like terms or named entities.

We showed by experiments the interest of PoS to correct transcription errors [30]. We proposed and evaluated several approaches to use PoS tag information in a postprocessing stage to rescore and reorder the 100-best hypotheses produced by the METISS Project-Team's ASR system; we noticed an absolute decrease of 0.9% of the word error rate when we include PoS knowledge into the score computed for each hypothesis [29].

We plan to further use PoS tagging to help us to segment outputs of ASR systems into topically cohesive stories, which will imply to find the most relevant keywords to discriminate the different parts of a given broadcast.

7. Contracts and Grants with Industry

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

7.1.1. Pôle de Compétitivité

Participant: Patrick Gros.

The French government organized in 2005 competitiveness poles (*pôles de compétitivité*) in France to strengthen ties in given region between industries (big and small companies), research labs (both public and private ones) and teaching institutions (universities and schools of engineering). We play an active role in the pole called "Images and networks" whose main actors are Thomson and France Télécom and which is located in Brittany and Pays de la Loire. In this frame, we are part of a common project called "Semimges" with France Télécom, Thomson, TDF, and the R2D2 team (IRISA). This project should be 2 year long and should start in 2007.

7.1.2. *Quaero*

Participants: Laurent Amsaleg, Vincent Claveau, Patrick Gros, Pascale Sébillot.

The French and German governments have launched a new initiative to build big common projects. We are part of the project named Quaero and which aims at developing technologies to create and exploit multimedia documents. The main partners are Thomson, France Télécom, DGA, CNRS. Many SME participates to this project like LTU, Exalead, Versys, Jouve, Bertin...The project is organized in four parts: content creation, content finding, core technologies and corpora. We are involved in the third of these parts. Our participation will mainly concern natural language processing, multimedia analysis and fast search algorithms.

7.1.3. *Contract with Thomson*

Participants: Patrick Gros, Siwar Baghdadi.

Duration: 36 months, starting in December 2005.

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

7.1.4. *Contract with France Télécom*

Participants: Patrick Gros, Gaël Manson.

Duration: 36 months, starting in November 2006.

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

7.2. European Initiatives

7.2.1. *European Network of Excellence MUSCLE: Multimedia Understanding through Semantics, Computation, and Learning*

Keywords: *Images*, *Multimedia*, *Naturel Language Processing*, *Video*.

Participants: Patrick Gros, Laurent Amsaleg, Pascale Sébillot, Manolis Delakis, Ewa Kijak.

Duration: 4 years, starting in April 2004. 42 partners. Prime: ERCIM, scientific coordinator: Nozha Boujemaa, INRIA.

This project aims at developing the collaboration in the domain of automatic multimedia document analysis, in particular to be able to handle and exploit their meaning. The project is thus concerned by all content-based analysis tools available for every media (text, sound and speech, image and video), but also by the techniques which allow us to combine the information extracted from each media, and by the common techniques needed to handle such data (optimization, classification, intensive computation).

TEXMEX is mainly active in the WP6 (multimodal analysis) through the work of M. Delakis (see Section 6.3.1).

7.2.2. *European Integrated Project aceMedia: Integrating Knowlege, Semantics and Content for User-Centered Intelligent Media Services*

Keywords: *Multimedia*, *Video*, *Video Indexing*.

Participants: Patrick Gros, Laurent Amsaleg, Zied Jemai, Pierre-Hugues Joalland.

Duration: 48 months, started in January 2004. 15 partners. Prime: Motorola Ltd.

The goal of this project is to encode multimedia document for their diffusion on networks like Internet, telecommunication networks or broadcasting systems. This new encoding scheme is based on autonomous entities called ACEs (standing for Autonomous Content Entity). Each entity is made of data, related metadata and an intelligence layer.

ACEs are dedicated to storing, retrieving and communicating documents in an efficient and autonomous way. It supports self-organization, self-annotation and self-adaptation according to the current user's preferences and devices. Additional embedded mechanisms are semantic detection, fast retrieval and relevance feedback.

TEXMEX team is responsible for the methods related to indexing, intelligent search and fast retrieval of ACE documents. The ACE documents will be described by both conceptual and content-based descriptors. Our algorithms compute the list of the most similar documents according to a query, by matching their numerical descriptors. Currently, the provided methods are the optimised exhaustive search and a based-on kd-trees method. Other methods will be evaluated.

8. Other Grants and Activities

8.1. National Initiatives

8.1.1. ACI masses de données Remix: Indexing Large Amounts of Data with Reconfigurable Chips

Participant: Laurent Amsaleg.

Duration: 3 years, starting in September 2003. Joint work with Symbiose and R2D2.

This project aims at developing new technologies to access very large databases like DNA or image databases. The core technology used in the project is based on FPGA chips.

8.1.2. ACI masses de données MDP2P: Managing Large Amounts of Data in P2P Systems

Participant: Laurent Amsaleg.

Duration: 3 years, starting in September 2003. Joint work with PARIS and ATLAS.

This project aims at studying the problem arising when managing lots of data on P2P systems like PC clusters. These problems are studied from a system point of view (memory management) and from an algorithmic point of view (parallelization of algorithms).

8.1.3. ACI masses de données DEMI-TON: Multimodal Description for Automatic Structuring of TV Streams

Participants: Manolis Delakis, Stéphane Huet, Patrick Gros, Xavier Naturel, Pascale Sébillot, Arnaud Dupuis, Cédric Dufouil.

Duration: 3 years, starting in April 2005. Partners: INA, METISS project-team.

This project concerns the development of new techniques to index large collections of TV programs. INA records and index more than 50 channels 24 hours a day. As the number of available documentalists did not increase as fast as the number of channels to index, they have to rely on more automatic processes. The first need is to verify that the programs in the stream correspond effectively to what was announced in the TV program guide and to synchronize the stream with this program guide. In a second stage, some programs like news reports have to be indexed to the topics that were tackled by the program and which, of course, could not be announced in the program guide.

The first year of this project was dedicated to define places of collaborations between the three involved groups and to start the work in these three directions:

- A first common work has been carried out to coordinate the effort done by the three partners to develop experimental platforms. This work will continue what has been achieved within the FERIA project, with new developments to take the hardware available at each partner's site;
- Macrosegmentation of TV streams has been studied: the learning based approach of INA will be coupled with the signal based approach of TEXMEX;
- The interaction between speech transcription and natural language processing tools has been studied in a collaboration between TEXMEX and METISS.

At IRISA, a first platform to record and analyze TV streams has been developed (see Section 5.2). It is based on a Windows server in charge of recording and distributing the content and two Linux PCs in charge of processing tasks. It is now available to all the members of METISS, TEXMEX and VISTA through a simple web site.

8.1.4. Participation to National Working Groups

- P. Sébillot is a member of the thematic network "Information and knowledge: discovering and abstracting" of the STIC department of CNRS.
- P. Sébillot is member of the AS "Semantic Web" of the STIC department of CNRS.
- P. Sébillot participates to AS "Text Mining" of the STIC department of CNRS.
- P. Sébillot is a member of AFIA Café (Collège apprentissage, fouille et extraction).
- P. Sébillot is a member of the working group A3CTE: Application, Learning and Knowledge Acquisition from Electronic Texts of GDR I3.
- P. Sébillot is a member of the working group PRC I3-AFIA TIA (terminologie et intelligence artificielle)

8.2. International Collaborations

8.2.1. Collaboration with Reykjavík University, Iceland

Keywords: *Approximate Search Schemes*, *Content-Based Image Retrieval Systems*, *Curse of Dimensionality*, *Local Descriptors*, *Random Projections*, *User Interface*.

Participants: Laurent Amsaleg, Hervé Jégou.

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

Image databases, and content-based image retrieval systems in particular, have become increasingly important in many applications areas. Moreover, new applications exploiting fine detail of images are now fast emerging thanks to recent and modern image processing techniques. While extremely effective (they return high quality results), these image processing techniques are very inefficient (they answer very slowly) due to their complexity and because of the inadequacy of traditional lower layers of software. This is particularly prevalent at large scale when dealing with image collections of realistic sizes. The goal of this project is to research and develop new database support that integrates efficiency and effectiveness for modern large-scale computer-vision related applications and problems.

Together, we came up with the PvS-framework that provides an efficient and scalable support for local description based recognition applications. While this work is still very active, we have initiated another thread of research by investigating the browsing of personnal image collections. Today, everyone can witness the tremendous increase in the capability to create, share and store digital images. As a result, personal image collections are growing at an astounding rate and it is clear that in the future individuals will need to access tens of thousands, or even hundreds of thousands, of digital images. It is therefore imperative to start studying ways to access these images in a useful and interesting manner. Addressing this topic is a new development in our cooperation.

8.2.2. Collaboration with Croatia and Slovenia

Participant: Annie Morin.

Partners: Andrija Stampar School of Public Health, Medical School, University of Zagreb, department of Electronics, Microelectronics, Computer and Intelligent systems, University of Zagreb, Zagreb, Croatia; Faculty of Computer and Information Science, University of Ljubljana, Slovenia; ERIC lab., University of Lyon; Rudjer Boskovic Institute, Zagreb, Croatia.

A. Morin keeps on working with the university of Zagreb and the university of Ljubljana in spite of the end of the Egide cooperation program. She was invited in April to give a course to Master students at the university of Zagreb. One of the Master student wrote his thesis on the use of factorial correspondence analysis in text mining and its implementation in the free software Orange, a data mining library for Python developed by Blaz Zupan at the university of Ljubljana.

8.2.3. *Collaboration with Dublin City University, Ireland*

Participants: Laurent Amsaleg, Patrick Gros.

Our collaboration in the frame of aceMedia outlined the opportunity for a deeper collaboration on large image base indexing and retrieval. The first objective of this new collaboration is to make a first experience of indexing and retrieval with a database of one million images. DCU has a good knowledge on image descriptors and retrieval system interfaces which is complementary to ours. On the other hand, our indexing technology is more advanced. A corpus of several hundreds of thousands of images has been shared between DCU and TEXMEX to make common experiments.

8.2.4. *Collaboration with University of Montreal, Quebec, Canada*

Participants: Vincent Claveau, Stéphane Huet, Fabienne Moreau, Pascale Sébillot.

A one-year collaboration agreement has been signed between Montreal (OLST research project) and Rennes 1 (TEXMEX research project) universities mid-2005. The cooperation aims at automatically detecting in textual corpora paradigmatic relations between lexical units or terms.

8.3. Visits of foreign researchers, Invitations to foreign labs

8.3.1. *Visit from members of the Reykjavík University, Iceland*

Participant: Laurent Amsaleg.

Partners: Bjorn Jónsson, Herwig Lejsek, Harnar Olafsonn, Fridrik Asmundsonn, Kari Hardarson.

In November 2005, Bjorn Jónsson, Herwig Lejsek, Harnar Olafsonn, Fridrik Asmundsonn spent a week at IRISA in the context of the Eff^2 associate team. In May 2006, Kari Hardarson visited Texmex as well.

8.3.2. *Visit to the Reykjavík University, Iceland*

Participants: Laurent Amsaleg, Hervé Jegou.

Laurent Amsaleg made several visits to the Reykjavík University for work and for participating to master committees. Hervé Jégou spent a full week with Laurent Amsaleg in June working with the Icelandic partners.

- L. Amsaleg was a member of the board for Hafthor Gudnason's Master's thesis in June 2006 and entitled "Median Rank in Face Recognition", University of Reykjavík, Iceland.
- L. Amsaleg was a member of the board for Fridrik Asmundsonn's Master's thesis in august 2006 and entitled "The NV-Network: A Distributed Architecture for High Throughput Image Retrieval", University of Reykjavík, Iceland.

8.3.3. *Visit of A. Polguère and M.-C. L'Homme, University of Montreal*

Participants: Vincent Claveau, Pascale Sébillot.

Partners: Alain Polguère, Professor at the University of Montreal, Canada, and head of OLST lab and Marie-Claude L'Homme, Professor at the University of Montreal, Canada.

In the framework of the collaboration between TEXMEX and OLST, A. Polguère and M.-C. L'Homme spent a week at IRISA. During this week, they worked on the lexical acquisition project and gave a seminar on the development of lexical and terminological resources.

8.3.4. *Visit of J. Puranen, University of Helsinki*

Participant: Annie Morin.

Partners: Juha Puranen, Professor at the University of Helsinki, Finland.

Juha Puranen spent 10 days at IRISA working with Annie Morin on Java applets for data mining lectures. He came here with a grant from the AFFRST, Association franco-finlandaise pour la recherche scientifique et technique and the applets can be found on http://noppa5.pc.helsinki.fi/koe/flash/flash-fra.html.

8.3.5. Visit to the University of Pernambouc, Brazil

Participant: Annie Morin.

After the ICOTS conference in July 2006 which has been hold in Salvador de Bahia, Annie Morin was invited by Francisco de Carvalho at the Computing centre of the federal university of Pernambouc (UFPE). She spent a week there working with students and giving a seminar on text mining.

9. Dissemination

9.1. Conference, Workshop and Seminar Organization

- A. Morin organized a workshop on "Statistical visualization: from PDA to huge screen" on behalf the SFdS (Société Française de Statistique) at IRISA in January 2006.

9.2. Involvement with the Scientific Community

- L. Amsaleg:
  - was a program committee member of the 22nd Bases de données avancées, BDA 2006, Lille, France;
  - was a program committee member of the 3rd conférence en Recherche d'Informations et Applications, CORIA'06, Lyon, March 2006;
  - was a member of the editorial board of the ACM Transactions on Multimedia Computing, Communications, and Applications, 2006;
  - was a program committee member of the 12th International Workshop on Multimedia Information Systems, MIS 2006;
  - was a program committee member of the 4th Manifestation des jeunes chercheurs francophones dans les domaines des STIC, MajecStic'06, Lorient, France, November 2006;
  - was reviewer for ARA's MDCA (Masses de Donn??es et Connaissances Ambiantes) call to project;
  - is the 2006 ACM SIGMOD Workshops Coordinator.
- N. Bonnel:
  - was a program committee member of the 4th Manifestation des jeunes chercheurs francophones dans les domaines des STIC, MajecStic'06, Lorient, France, November 2006.
- V. Claveau:
  - was a program committee member of the 3rd COnference on Information Retrieval and Application CORIA'06, Lyon, March 2006;
  - was a program committee member of the special issue *Traitements automatisés des corpus spécialisés : contexte et sens* of the journal *Glottopol*, n^{o} 8, 2006;
  - was reviewer for ARA's MDCA (Masses de Données et Connaissances Ambiantes) call to project.
- P. Gros:

  - is a member of the editorial board of the French journal entitled "Traitement du signal";
  - was program committee member of the third COnference on Information Retrieval and Application CORIA'06, Lyon, March 2006;
  - was a program committee member of the third International Symposium on Image/Video Communications over fixed and mobile networks ISIVC 2006, Hammamet, Tunisia, September 2006;
  - was a program committee member of the Pacific Rim Conference on Multimedia PCM 2006, Hangzhou, China, November 2006.
- E. Kijak
  - is a member of the organizing committee of the campaign ARGOS (Programme Techno Vision) for benchmarking of video content analysis methods.
- F. Moreau:
  - was a program committee member of MajecStic'06 (4^{th} Manifestation des jeunes chercheurs francophones dans les domaines des STIC), Lorient, France, November 2006.
- A. Morin:
  - was a member of the program committee of ITI 2006 (Information technology interfaces);
  - is elected President of the GMF IIS (group of the French members of the international statistical institute) until may 2007;
  - is a member of the CNU (National Council of the University) in the computer science section.
- P. Sébillot:
  - is an associate editor of the journal In Cognito - Cahiers Romans de Sciences Cognitives.
  - is an associate editor of Jedai (Journal Électronique d'Intelligence Artificielle).
  - was a member of the program committee of JADT'06 (8^{th} International Journées internationales d'analyse statistique des données textuelles), Besançon, France, April 2006.
  - was a member of the program committee of DEFT'06 (4^{th} Défi fouille de textes), Fribourg, Switzerland, September 2006.
  - was a member of the program committee of MajecStic'06 (4^{th} Manifestation des jeunes chercheurs francophones dans les domaines des STIC), Lorient, France, November 2006.
  - is a member of the board of ATALA (Association pour le traitement automatique des langues).
  - is a member of the scientific committee of the TCAN program of CNRS (Traitement des connaissances, apprentissage et NTIC).

9.3. Teaching Activities

- L. Amsaleg: Advanced Databases, Professional Master in Computer Science, 2^{nd} year, IFSIC, University of Rennes 1.
- L. Amsaleg: Management of Multimedia Documents, Professional Master in Computer Science, 2^{nd} year, IFSIC, University of Rennes 1.
- L. Amsaleg: Advanced Databases & Modern Information Systems, Diic3, IFSIC, University of Rennes 1.

- E. Kijak and P. Sébillot: Multimedia Indexing: Techniques and Applications, Research Master in Computer Science, 2nd year, IFSIC, University of Rennes 1.
- E. Kijak and F. Moreau: Digital Documents Indexing and Retrieval, Professional Master in Computer Science, 2nd year, IFSIC, University of Rennes 1.
- A. Morin and V. Claveau: Textual data analysis, Research Master "Biomedical information processing, 2nd year, University of Rennes 1.
- P. Sébillot and V. Claveau: Advanced Databases and Modern Information Systems, 5th year, Computer Science, INSA Rennes.

9.4. Participation to Seminars, Workshops, Invited Conferences

- S. Huet and G. Gravier (from the METISS Project-Team) gave an invited talk at the Faculty of Information Technology, Brno University of Technology, Czech Republic, about the use of POS tagging in speech recognition systems, in September 2006.
- P. Sébillot gave an invited talk at France Télécom R&D seminar, Lannion, about machine-learning based acquisition of linguistic resources for multimedia information retrieval, in July 2006.

9.5. Popular work, Press articles

- In March 2006, L. Amsaleg published an article describing his work in the issue nb. 53 of the Newsletter of INRIA. This article, entitled "Un logiciel pour identifier rapidement les images piratées – A Software for Fast Identification of Pirated Images" was for some time in the front page of INRIA, as a typical example of good international cooperation.
- In March 2006, L. Amsaleg also published an article in a French popular scientific magazine called "Science & Vie Junior", entitled "Des images piratées débusquées en 1 seconde" (unfortunately, subscription is needed to read the article).
- Also, L. Amsaleg had a article in the *Les Échos* economic newspaper on April 12, entitled "À la recherche d'images piratées".
- L. Amsaleg had another article in the scientific journal of the CNRS, published in the issue number 197 of June 2006, entitled "Un logiciel contre le vol d'images".
- Finaly, L. Amsaleg were asked by an editor named "Techniques pour l'ingénieur" to write an extended article describing not only his research work but also the context within which enforcing copyright protection is key. This article, entitled "Contrer le piratage d'images : un logiciel précis et rapide" will be published in February 2007.

10. Bibliography

Major publications by the team in recent years

[1] L. AMSALEG, P. GROS. *Content-based Retrieval Using Local Descriptors: Problems and Issues from a Database Perspective*, in "Pattern Analysis and Applications", vol. 2001, no 4, March 2001, p. 108-124.

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

[3] S.-A. BERRANI, L. AMSALEG, P. GROS. *Approximate Searches: k-Neighbors + Precision*, in "Proceedings of the 12th ACM International Conference on Information and Knowledge Management", 2003, p. 24-31.

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

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

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

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

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

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

Year Publications

Doctoral dissertations and Habilitation theses

[10] N. BONNEL. *Génération dynamique de présentations interactives en multimédia 3D, de données, pour des applications en ligne*, Thèse de doctorat, Université de Rennes 1, December 2006.

[11] M. DELAKIS. *Multimodal Tennis Video Structure Analysis with Segment Models*, Thèse de doctorat, Université de Rennes 1, France, October 2006.

[12] A. KOUOMOU-CHOUPO. *Améliorer la recherche par similarité dans une grande base d'images fixes par des techniques de fouille de données*, Thèse de doctorat, Université de Rennes 1, February 2006, http://www.irisa.fr/texmex/publications/versionElect/2006/these_anicet.pdf.

[13] F. MOREAU. *Revisiter le couplage traitement automatique des langues et recherche d'information*, Thèse de doctorat, Université de Rennes 1, December 2006.

[14] F. TONNIN. *Description locale des images fixes dans le domaine compressé*, Thèse de doctorat, Université de Rennes 1, June 2006, http://www.irisa.fr/texmex/people/tonnin/these.ps.

Articles in refereed journals and book chapters

[15] B. BACHIMONT, P. GROS. *Recherche : des défis scientifiques*, in "Les nouveaux dossiers de l'audiovisuel, numéro spécial : Internet : quelle place pour la vidéo ?", n° 9, March 2006, p. 28-30.

[16] V. CLAVEAU, M.-C. L'HOMME. *Discovering and Organizing Noun-Verb Collocations in a Specialized Corpora Using Inductive Logic Programming*, in "International Journal of Corpus Linguistics", vol. 11, n^{o} 2, 2006, p. 209–241.

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

[18] F. MOREAU, V. CLAVEAU. *Extension de requêtes par relations morphologiques acquises automatiquement*, in "Revue I3 (Information - Interaction - Intelligence)", to appear, 2006.

[19] S. OZDOWSKA, V. CLAVEAU. *Inférence de règles de propagation syntaxique pour l'alignement de mots*, in "TAL (Traitement Automatique des Langues)", to appear, 2006.

[20] P. SÉBILLOT. *Symbolic Machine Learning: A Different Answer to the Problem of the Acquisition of Lexical Knowledge from Corpora*, in "TripleC (Cognition, Communication, Co-operation), special issue: selected papers from ECAP 2005 - European Computing and Philosophy Conference 2005", vol. 4, n^{o} 2, 2006, p. 277-283.

Publications in Conferences and Workshops

[21] N. BONNEL, M. CHEVALIER. *Critères d'évaluation pour les interfaces des systèmes de recherche d'information*, in "Actes de la 3ème Conférence en Recherche d'Information et Applications (CORIA'06), Lyon, France", March 2006, p. 109-114, http://hal.inria.fr/docs/00/06/31/97/PDF/Bonnel_CORIA06_web.pdf.

[22] N. BONNEL, M. CHEVALIER. *Evaluation des Interfaces Utilisateur d'Information*, in "Actes du 4ème Atelier Visualisation et Extraction de Connaissances - EGC'06", 2006, p. 23-38, http://hal.inria.fr/docs/00/06/31/98/PDF/Bonnel_EGC06_AtelierVisu_web.pdf.

[23] N. BONNEL, G. LE GUERNIC. *Système de recherche de méthodes Java basé sur leur signature*, in "Actes de la 4ème Manifestation des Jeunes Chercheurs francophones dans les domaines des STIC (MajecSTIC'06), Lorient, France", November 2006, à paraître.

[24] N. BONNEL, V. LEMAIRE, A. COTARMANAC'H, A. MORIN. *Effective Organization and Visualization of Web Search Results*, in "Proceedings of the IASTED International Conference on Internet and Multimedia Systems and Applications (EuroIMSA'06)", A. BOUCOUVALAS (editor). , ACTA Press, 2006, p. 209-216, http://hal.inria.fr/docs/00/06/31/99/PDF/Bonnel_EuroIMSA06_web.pdf.

[25] M. DELAKIS, G. GRAVIER, P. GROS. *Score oriented Viterbi search in sport video structuring using HMM and segment models*, in "Proceedings of the International Workshop on Multimedia Signal Processing (MMSP'06), Victoria, Canada", October 2006, http://www.irisa.fr/texmex/publications/versionElect/2006/mmsp06.pdf.

[26] A. DUPUIS, C. DUFOUIL. *Une solution client/serveur pour l'analyse de corpus audiovisuels*, in "Actes de la conférence Systèmes d'Informations Audio-Visuels, SIAV'06, Fribourg, Switzerland", September 2006.

[27] G. GEORGES, S. DERRIEN, S. RUBINI, F. RAIMBAULT, L. AMSALEG. *ReMIX: une architecture pour la recherche dans les masses de données indexées*, in "Actes de Sympa 2006, Symposium en Architecture de Machines, Perpignan, France", 2006.

[28] P. E. HADJIDOUKAS, L. AMSALEG. *Parallelization of a Hierarchical Data Clustering Algorithm using OpenMP*, in "Proceedings of IWOMP, Second International Workshop on OpenMP, Reims, France", June 2006.

[29] S. HUET, G. GRAVIER, P. SÉBILLOT. *Are Morphosyntactic Taggers Suitable to Improve Automatic Transcription?*, in "Proceedings of Text, Speech and Dialogue (TSD), Brno, Czech Republic", Lecture Notes in Computer Science, vol. 4188/2006, September 2006, p. 391-398, http://www.springerlink.com/content/e50760087267418r/fulltext.pdf.

[30] S. HUET, G. GRAVIER, P. SÉBILLOT. *Peut-on utiliser les étiqueteurs morphosyntaxiques pour améliorer la transcription automatique ?*, in "Actes des 26èmes Journées d'Études sur la Parole (JEP), Dinard, France", June 2006, http://jep2006.irisa.fr/openconf/author/final/final-86.pdf.

[31] H. LEJSEK, F. ÁSMUNDSSON, B. ÞÓR-JÓNSSON, L. AMSALEG. *Blazingly Fast Image Copyright Enforcement*, in "Proceedings of the 14th ACM International Conference on Multimedia, demonstrations, Santa Barbara, CA, USA", October 2006, http://www.irisa.fr/texmex/publications/versionElect/2006/de05-lejsek.pdf.

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

[33] F. MOREAU, V. CLAVEAU. *Extension de requêtes par relations morphologiques acquises automatiquement*, in "Actes de la 3ème Conférence en Recherche d'Informations et Applications, (CORIA'06), Lyon, France", March 2006, p. 181-192.

[34] A. MORIN. *Intensive use of factorial correspondence analysis for text mining: application with statistical education publications*, in "Proceedings of ICOTS-7 (International Conference on Teaching Statistics), Salvador, Bahia, Brazil", July 2006.

[35] X. NATUREL, G. GRAVIER, P. GROS. *Fast Structuring of Large Television Streams using Program Guides*, in "Proceedings of the 4th International Workshop on Adaptive Multimedia Retrieval (AMR), Geneva, Switzerland", July 2006, http://www.irisa.fr/texmex/publications/versionElect/2006/amr_naturel.pdf.

[36] A. REMAZEILLES, F. CHAUMETTE, P. GROS. *3D navigation based on a visual memory*, in "Proceedings of the IEEE International Conference on Robotics and Automation, Orlando, Floride, USA", May 2006.

[37] A. REMAZEILLES, N. MANSARD, F. CHAUMETTE. *Qualitative Visual Servoing: application to the visibility constraint*, in "Proceedings of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS'06, Beijing, China", October 2006.

[38] S. SEGVIC, A. REMAZEILLES, F. CHAUMETTE. *Enhancing the Point Feature Tracker by Adaptive Modelling of the Feature Support*, in "Proceedings of the European Conf. on Computer Vision, ECCV'2006, Graz, Austria", Lecture Notes in Computer Science, vol. 3952, May 2006, p. 112–124.

Internal Reports

[39] S. HUET, P. SÉBILLOT, G. GRAVIER. *Introduction de connaissances linguistiques en reconnaissance de la parole : un état de l'art*, Technical report, n^{o} 1804, IRISA, May 2006, https://hal.inria.fr/inria-00077747.

[40] X. NATUREL, P. GROS. *Detecting Repeats for Video Structuring*, Technical report, n^{o} 1790, IRISA, March 2006, https://hal.inria.fr/inria-00001154.

References in notes

[41] 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.

[42] D. ACHLIOPTAS. *Database-friendly random projections*, in "Proceedings of the 20th ACM Symposium on Principles of Database Systems", 2001.

[43] J. ANDERSON, M. STONEBRAKER. *Sequoia 2000 Metadata Schema for Satellite Images*, in "ACM SIGMOD Record, Special issue on Metadata for Digital Media", vol. 23, n^{o} 4, 1994.

[44] K. BENNETT, U. FAYYAD, D. GEIGER. *Density-Based Indexing for Approximate Nearest-Neighbor Queries*, in "Proceedings of the 5th ACM International Conference on Knowledge Discovery and Data Mining, San Diego, CA, USA", August 1999.

[45] S. BERCHTOLD, D. KEIM, H. KRIEGEL. *The X-tree : An Index Structure for High-Dimensional Data*, in "VLDB", 1996.

[46] K. BEYER, J. GOLDSTEIN, R. RAMAKRISHNAN, U. SHAFT. *When Is "Nearest Neighbor" Meaningful?*, in "Proceedings of the 8th International Conference on Database Theory, London, United Kingdom", January 1999.

[47] T. BINFORD, T. LEVITT. *Quasi-Invariants: Theory and Exploitation*, in "Proceedings of DARPA Image Understanding Workshop", 1993, p. 819-829.

[48] E. BINGHAM, H. MANNILA. *Random projection in dimensionality reduction: Applications to image and text data*, in "Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining", 2001.

[49] N. BONNEL, A. COTARMANAC'H, A. MORIN. *Visualisation 3D des résultats de recherche : quel avenir ?*, in "Créer, jouer, échanger : expériences de réseaux (Actes de la 8e Conférence H2PTM : Hypermedias Hypertexts, Products, Tools and Methods)", J. CLÉMENT, S. IMAD (editors). , Hermes Science Publications, November 2005, p. 325-339.

[50] N. BONNEL, F. MOREAU. *Quel avenir pour les moteurs de recherche ?*, in "Actes de la 3e Manifestation des Jeunes Chercheurs francophones dans les domaines des STIC (MajecSTIC)", November 2005, p. 291-299.

[51] N. BOUJEMAA, S. BOUGHORBEL, C. VERTAN. *Soft Color Signatures for Image Retrieval by Content*, in "Eusflat'2001", vol. 2, 2001, p. 394-401.

[52] P. BOUTHEMY, M. GELGON, F. GANANSIA. *A Unified Approach to Shot Change Detection and Camera Motion Characterization*, in "IEEE Transactions on Circuits and Video Technology", vol. 9, n^{o} 7, October 1999, p. 1030-1044.

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

[54] K. BÖHM, T. RAKOW. *Metadata for Multimedia Documents*, in "ACM SIGMOD Record Special Issue on Metadata for Digital Media", vol. 23, n^o 4, 1994, p. 21-26.

[55] F. CHAUMETTE. *De la perception à l'action : l'asservissement visuel, de l'action à la perception : la vision active*, Habilitation à diriger des recherches, Université de Rennes 1, January 1998.

[56] F. CHEN, M. HEARST, J. KUPIEC, J. PEDERSON, L. WILCOX. *Metadata for Mixed-Media Access*, in "ACM SIGMOD Record, Special issue on Metadata for Digital Media", vol. 23, n^o 4, 1994.

[57] P. CIACCIA, M. PATELLA. *PAC Nearest Neighbor Queries: Approximate and Controlled Search in High-Dimensional and Metric Spaces*, in "Proceedings of the 16th International Conference on Data Engineering, San Diego, CA, USA", February 2000.

[58] S. DASGUPTA. *Experiments with random projections*, in "In Proceedings of the 16th Conf. Uncertainty in Artificial Intelligence", 2000.

[59] M. DELAKIS, G. GRAVIER, P. GROS. *Audiovisual Fusion with Segment Models for Video Structure Analysis*, in "Proceedings of the 2nd European Workshop on the Integration of Knowledge, Semantic and Digital Media Technologies (EWIMT'05)", 2005.

[60] M. DELAKIS, G. GRAVIER, P. GROS. *Multimodal Segmental-Based Modeling of Tennis Video Broadcasts*, in "Proceedings of IEEE international Conference on Multimedia and Expo", 2005.

[61] Y. DUFOURNAUD, C. SCHMID, R. HORAUD. *Appariement d'images à des échelles différentes*, in "Actes du 12e Congrès Francophone AFRIF-AFIA de Reconnaissance des Formes et Intelligence Artificielle, Paris, France", vol. 2, February 2000, p. 327-336.

[62] B. ESPIAU, F. CHAUMETTE, P. RIVES. *A New Approach to Visual Servoing in Robotics*, in "IEEE Transactions on Robotics and Automation", vol. 8, n^o 3, June 1992, p. 313-326.

[63] R. FABLET. *Modélisation statistique non paramétrique et reconnaissance du mouvement dans des séquences d'images : application à l'indexation vidéo*, Thèse de doctorat, Université de Rennes 1, France, July 2001.

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

[65] H. FERHATOSMANOGLU, E. TUNCEL, D. AGRAWAL, A. E. ABBADI. *Approximate Nearest Neighbor Searching in Multimedia Databases*, in "Proceedings of the 17th International Conference on Data Engineering, Heidelberg, Germany", April 2001, p. 503–511.

[66] G. FINLAYSON, S. CHATTERJEE, B. FUNT. *Color Angular Indexing*, in "Proceedings of the 4th European Conference on Computer Vision, Cambridge, Angleterre", 1996, p. 16-27.

[67] G. FINLAYSON, M. DREW, B. FUNT. *Color Constancy: Generalized Diagonal Transforms Suffice*, in "Journal of the Optical Society of America A", vol. 11, n^o 11, November 1994, p. 3011-3019.

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

[69] U. GARGI, R. KASTURI, S. ANTANI. *Performance characterization and comparison of video indexing algorithms*, in "Proceedings of the Conference on Computer Vision and Pattern Recognition, Santa Barbara, Californie, États-Unis", June 1998, p. 559-565.

[70] J. GERBRANDS. *On The Relationships Between SVD, KLT and PCA*, in "Pattern Recognition", vol. 14, n^o^ 1-6, 1981, p. 375–381.

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

[72] U. GLAVITSCH, P. SCHAUBLE, M. WECHSLER. *Metadata for Integrating Speech Documents in a Text Retrieval System*, in "ACM SIGMOD Record, Special issue on Metadata for Digital Media", vol. 23, n^o^ 4, 1994.

[73] J. GOLDSTEIN, R. RAMAKRISHNAN. *Contrast Plots and P-Sphere Trees: Space vs. Time in Nearest Neighbor Searches*, in "Proceedings of the 26th Int. Conf. on Very Large Data Bases, Cairo, Egypt", September 2000, p. 429–440.

[74] P. GROS. *Experimental Evaluation of Color Illumination Models for Image Matching and Indexing*, in "Proceedings of the RIAO'2000 Conference on Content-Based Multimedia Information Access", April 2000, p. 567-574.

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

[76] R. HAMMOUD, R. MOHR. *Mixture Densities for Video Objects Recognition*, in "Proceedings of the 15th International Conference on Pattern Recognition, Barcelone, Espagne", vol. 2, IAPR, September 2000, p. 71-75.

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

[78] A. HENRICH. *The LSD^h-Tree: An Access Structure for Feature Vectors*, in "ICDE", 1998.

[79] J. HUANG, S. R. KUMAR, M. MITRA, W. ZHU, R. ZABIH. *Image Indexing Using Color Correlograms*, in "Proceedings of the Conference on Computer Vision and Pattern Recognition, Puerto Rico, USA", June 1997, p. 762-768.

[80] R. JAIN, A. HAMPAPURAM. *Representations of Video Databases*, in "ACM SIGMOD Record, Special issue on Metadata for Digital Media", vol. 23, n^o^ 4, 1994.

[81] V. KASHYAP, A. SHETH. *Semantic Heterogeneity in Global Information Systems: The Role of Metadata, Context and Ontologies*, in "Cooperative Information Systems, San Diego, Californie, États-Unis", M. PAPAZOGLOU, G. SCHLAGETER (editors). , Academic Press, 1998, p. 139-178.

[82] N. KATAYAMA, S. SATOH. *The SR-tree: An Index Structure for High-Dimensional Nearest Neighbor Queries*, in "ACM SIGMOD", 1997.

[83] F. KORN, B. PAGEL, C. FALOUTSOS. *On the 'Dimensionality Curse' and the 'Self-Similarity Blessing'*, in "IEEE Trans. on Knowledge and Data Engineering", vol. 13, n^o 1, January 2001, p. 96–111.

[84] H. LEJSEK, F.-H. ÁSMUNDSSON, B. ÞÓR-JÓNSSON, L. AMSALEG. *Efficient and Effective Image Copyright Enforcement*, in "Actes des Journées Bases de Données Avancées, BDA'05, Saint Malo, France", October 2005.

[85] C. LI, E. CHANG, H. GARCIA-MOLINA, G. WIEDERHOLD. *Clustering for Approximate Similarity Search in High-Dimensional Spaces*, in "IEEE Trans. on Knowledge and Data Engineering", vol. 14, n^o 4, July 2002, p. 792–808.

[86] E. LOUPIAS, N. SEBE, S. BRES, J.-M. JOLION. *Wavelet-based Salient Points for Image Retrieval*, in "Proceedings of the IEEE International Conference on Image Processing, Vancouver, Canada", 2000.

[87] A. LUGMAYR, S. NIIRANEN, S. KALLI. *Digital Interactive TV and Metadata, Future Broadcast Multimedia*, Signals and Communication Technology, Springer, 2004.

[88] E. MALIS, F. CHAUMETTE, S. BOUDET. *2 1/2 D Visual Servoing*, in "IEEE Transactions on Robotics and Automation", vol. 15, n^o 2, April 1999, p. 238-250.

[89] S. MARCUS, V. SUBRAHMANIAN. *Foundations of Multimedia Database Systems*, in "Journal of the ACM", vol. 43, n^o 3, 1996, p. 474-523.

[90] I. MEL'ČUK. *Vers une linguistique Sens-Texte*, January 1997, Leçon inaugurale, Collège de France, Chaire internationale.

[91] I. MEL'ČUK. *Collocations and Lexical Functions*, in "Phraseology. Theory, Analysis, and Applications, Oxford", A. P. COWIE (editor). , chap. 2, Clarendon Press, 1998.

[92] K. MIKOLAJCZYK, C. SCHMID. *An Affine Invariant Interest Point Detector*, in "Proceedings of the 7th European Conference on Computer Vision, Copenhague, Danemark", 2002.

[93] J. NIEVERGELT, H. HINTERBERGER, K. SEVCIK. *The Grid File: An Adaptable, Symmetric Multikey File Structure*, in "ACM TODS", vol. 9, n^o 1, 1984.

[94] B. PAGEL, F. KORN, C. FALOUTSOS. *Deflating the Dimensionality Curse Using Multiple Fractal Dimensions*, in "Proceedings of the 16th International Conference on Data Engineering, San Diego, California, USA", March 2000.

[95] J. PUSTEJOVSKY. *The Generative Lexicon*, MIT Press, Cambridge, 1995.

[96] F. RASTIER. *Sémantique Interprétative*, Second, Presses universitaires de France, 1996.

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

[98] R. RUILOBA, P. JOLY, S. MARCHAND-MAILLET, G. QUENOT. *Towards a Standard Protocol for the Evaluation of Video-to-Shots Segmentation Algorithms*, in "Proceedings of the first European Workshop on Content Based Multimedia Indexing, Toulouse, France, Toulouse, France", October 1999.

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

[100] C. SCHMID, R. MOHR. *Local Grayvalue Invariants for Image Retrieval*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", vol. 19, n^{o} 5, May 1997, p. 530-534, ftp://ftp.inrialpes.fr/pub/movi/publications/schmid_pami97.ps.gz.

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

[102] M. STRICKER, M. SWAIN. *The Capacity of Color Histogram Indexing*, in "Proceedings of the Conference on Computer Vision and Pattern Recognition, Seattle, Washington, USA", 1994.

[103] R. WEBER, K. BÖHM. *Trading Quality for Time with Nearest Neighbor Search*, in "Proceedings of the 7th Conference on Extending Database Technology, Konstanz, Germany", March 2000.

[104] R. WEBER, H. SCHEK, S. BLOTT. *A Quantitative Analysis of Performance Study for Similarity-Search Methods in High-Dimensional Spaces*, in "Proceedings of the 24th International Conference on Very Large Data Bases, New York City, New York, États-Unis", August 1998, p. 194-205.

[105] D. WHITE, R. JAIN. *Similarity Indexing with the SS-tree*, in "ICDE", 1996.