INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Project-Team TEMICS

TraitEment, Modélisation d'Images et CommunicationS

Rennes - Bretagne-Atlantique

Theme : Vision, Perception and Multimedia Understanding

Activity Report

2009

Table of contents

1. Team

Research Scientist

Christine Guillemot [Team Leader, DR, Inria, HdR]
Teddy Furon [CR, INRIA, on leave at Thomson]
Cédric Herzet [CR, INRIA]
Claude Labit [DR, INRIA, 20%, HdR]
Aline Roumy [CR, INRIA]
Caroline Fontaine [CR, CNRS, 80%, until 31st August 2009]
Gagan Rath [CR, INRIA, until June 2009]

Faculty Member

Jean-Jacques Fuchs [Professor, SPM, University of Rennes 1, HdR]
Olivier Le Meur [MdC, IFSIC, University of Rennes 1, since 1st Sept. 2009]
Simon Malinowski [ATER]

Technical Staff

Laurent Guillo [IR, CNRS]

PhD Student

Simon Bos [INRIA grant, Joint Lab. INRIA-Alcatel Lucent]
Ana Charpentier [INRIA grant, MEDIEVALS project]
Thomas Colleu [CIFRE contract with France Télécom]
Angélique Drémeau [INRIA grant]
Josselin Gauthier [Ministry of research grant, since 1st Oct. 2009]
Vincent Jantet [INRIA grant, Futurimages project]
Aurélie Martin [CIFRE contract with Thomson, until 31st Oct. 2009]
Jonathan Taquet [Inria-Région ARED grant]
Velotiaray Toto-Zarasoa [Ministry of research grant]
Mehmet Turkan [INRIA grant, CORDIS]
Fuchun Xie [INRIA grant, CORDIS]
Joaquin Zepeda [INRIA grant, ANR-ICOS-HD project, also with TexMex]

Post-Doctoral Fellow

Fethi Smach [INRIA grant, ANR-ICOS-HD project, since 22nd July 2009]

Administrative Assistant

Huguette Béchu [TR, INRIA, shared with VISTA]

Other

Ronan Boitard [INRIA, since 15th Sept. 2009]
Mathieu Desoubeaux [INRIA, until 15th Nov. 2009]
Gabrielle Ombrouk [INRIA, Until 14th Sept. 2009]
Yann Serrand [INRIA]
Tangui Poirier [INRIA, since 15th Sept. 2009]
Gael Sourimant [INRIA grant, Fururimages project]
Denis Kubasov [CNRS grant, ESTIVAL project, until 30th June 2009]

2. Overall Objectives

2.1. Overall Objectives

The goal of the TEMICS project-team is the design and development of algorithms and practical solutions in the areas of analysis, modelling, coding, communication and watermarking of images and video signals.The TEMICS project-team activities are structured and organized around the following research directions :

- *3D modelling and representations of multi-view video sequences.*

The emergence of new video formats, allowing panoramic viewing, free viewpoint video (FTV), and Three-Dimensional TV (3DTV) on immersive displays are creating new scientific and technological problems in the area of video content modelling and representation. Omni-directional video, free viewpoint video and stereoscopic or multi-view video are formats envisaged for interactive and 3DTV. Omni-directional video refers to a 360-degree view from one single viewpoint or a spherical video. The notion of "free viewpoint video" refers to the possibility for the user to choose an arbitrary viewpoint and/or view direction within a visual scene, creating an immersive environment. A multi-view video together with depth information allows, by using view synthesis techniques, the generation of virtual views of the scene from any viewpoint. This property can be used in a large diversity of applications, including 3DTV, FTV, security monitoring and tracking. This type of 3D content representation is also known as MVD (Multi-View plus Depth). The TEMICS project-team focuses on several algorithmic problems to analyze, represent, compress and render multi-view video content. The team first addresses the problem of depth information extraction. The depth information is associated with each view as a depth map, and transmitted in order to perform virtual view generation and allow the inter-operability between capture (with N cameras) and display (of P views) devices. The huge amount of data contained in multi-view sequences motivates the design of efficient representation and compression algorithms.

- *Sparse representations, compression, feature extraction and texture description.*

  Low rate as well as scalable compression remains a widely sought capability. Scalable video compression is essential to allow for optimal adaptation of compressed video streams to varying network characteristics (e.g. to bandwidth variations) as well as to heterogeneous terminal capabilities. Wavelet-based signal representations are well suited for such scalable signal representations. Special effort is thus dedicated to the study of motion-compensated spatio-temporal expansions making use of complete or overcomplete transforms, e.g. wavelets, curvelets and contourlets, and more genrally of sparse signal approximation and representation techniques. The sparsity of the signal representation depends on how well the bases match with the local signal characteristics. Anisotropic waveforms bases, based on directional transforms or on sets of bases optimized in a sparsity-distortion sense are studied. Methods for texture analysis and synthesis, for prediction and for inpainting, which are key components of image and video compression algorithms, based on sparse signal representations are also developed. The amenability of these representations for image texture description is also investigated and measures of distance between sparse vectors are designed for approximate nearest neighbors search and for image retrieval. Beyond sparse image and video signals representations, the problem of quantization of the resulting representations taking into account perceptual models and measures, in order to optimize a trade-off between rate and perceptual quality, is studied.

- *Joint source-channel coding*. The advent of Internet and wireless communications, often characterized by narrow-band, error and/or loss prone, heterogeneous and time-varying channels, is creating challenging problems in the area of source and channel coding. Design principles prevailing so far and stemming from Shannon's source and channel separation theorem must be re-considered. The separation theorem holds only under asymptotic conditions where both codes are allowed infinite length and complexity. If the design of the system is heavily constrained in terms of complexity or delay, source and channel coders, designed in isolation, can be largely suboptimal. The project objective is to develop theoretical and practical solutions for image and video transmission over heterogeneous, time-varying wired and wireless networks. Many of the theoretical challenges are related to understanding the tradeoffs between rate-distortion performance, delay and complexity for the code design. The issues addressed encompass the design of error-resilient source codes, joint source-channel codes and multiply descriptive codes, minimizing the impact of channel noise (packet losses, bit errors) on the quality of the reconstructed signal, as well as of turbo or iterative decoding techniques.

- *Distributed source and joint source-channel coding.* Current compression systems exploit correlation on the sender side, via the encoder, e.g. making use of motion-compensated predictive or filtering techniques. This results in asymmetric systems with respectively higher encoder and lower decoder complexities suitable for applications such as digital TV, or retrieval from servers with e.g. mobile devices. However, there are numerous applications such as multi-sensors, multi-camera vision systems, surveillance systems, with light-weight and low power consumption requirements that would benefit from the dual model where correlated signals are coded separately and decoded jointly. This model, at the origin of distributed source coding, finds its foundations in the Slepian-Wolf and the Wyner-Ziv theorems. Even though first theoretical foundations date back to early 70's, it is only recently that concrete solutions have been introduced. In this context, the TEMICS project-team is working on the design of distributed prediction and coding strategies based on both source and channel codes. Although the problem is posed as a communication problem, classical channel decoders need to be modified. Distributed joint source-channel coding refers to the problem of sending correlated sources over a common noisy channel without communication between the senders. This problem occurs mostly in networks, where the communication between the nodes is not possible or not desired due to its high energy cost (network video camera, sensor network...). For independent channels, source channel separation holds but for interfering channels, joint source-channel schemes (but still distributed) performs better than the separated scheme. In this area, we work on the design of distributed source-channel schemes.

- *Data hiding and watermarking*.

  The distribution and availability of digital multimedia documents on open environments, such as the Internet, has raised challenging issues regarding ownership, users rights and piracy. With digital technologies, the copying and redistribution of digital data has become trivial and fast, whereas the tracing of illegal distribution is difficult. Consequently, content providers are increasingly reluctant to offer their multimedia content without a minimum level of protection against piracy. The problem of data hiding has thus gained considerable attention in the recent years as a potential solution for a wide range of applications encompassing copyright protection, authentication, steganography, or a mean to trace illegal usage of the content. This latter application is referred to as fingerprinting. Depending on the application (copyright protection, traitor tracing or fingerprinting, hidden communication), the embedded signal may need to be robust or fragile, more or less imperceptible. One may need to only detect the presence of a mark (watermark detection) or to extract a message. The message may be unique for a given content or different for the different users of the content, etc. These different applications place various constraints in terms of capacity, robustness and security on the data hiding and watermarking algorithms. The robust watermarking problem can be formalized as a communication problem : the aim is to embed a given amount of information in a host signal, under a fixed distortion constraint between the original and the watermarked signal, while at the same time allowing reliable recovery of the embedded information subject to a fixed attack distortion. Applications such as copy protection, copyright enforcement, or steganography also require a security analysis of the privacy of this communication channel hidden in the host signal.

Given the strong impact of standardization in the sector of networked multimedia, TEMICS, in partnership with industrial companies, seeks to promote its results in standardization (JPEG, MPEG). While aiming at generic approaches, some of the solutions developed are applied to practical problems in partnership with industry (Thomson, France Télécom) or in the framework of national projects (RIAM ESTIVALE), ANR ESSOR, ANR ICOS-HD, ANR MEDIEVALS, ANR PERSEE, DGE/REGION FUTURIMAGES) and European projects (IST-NEWCOM++). The application domains addressed by the project are networked multimedia applications (on wired or wireless Internet) via their various requirements and needs in terms of compression, of resilience to channel noise, or of advanced functionalities such as navigation, protection and authentication.

3. Scientific Foundations

3.1. 3D scene modelling based on projective geometry

3D reconstruction is the process of estimating the shape and position of 3D objects from views of these objects. TEMICS deals more specifically with the modelling of large scenes from monocular video sequences. 3D reconstruction using projective geometry is by definition an inverse problem. Some key issues which do not have yet satisfactory solutions are the estimation of camera parameters, especially in the case of a moving camera. Specific problems to be addressed are e.g. the matching of features between images, and the modelling of hidden areas and depth discontinuities. 3D reconstruction uses theory and methods from the areas of computer vision and projective geometry. When the camera $\mathcal{C}_i$ is modelled as a *perspective projection*, the *projection equations* are :

$$\widetilde{p_i} = P_i \widetilde{x}, \tag{1}$$

where $\widetilde{x}$ is a 3D point with homogeneous coordinates $\widetilde{x} = (x\ y\ z\ 1)^t$ in the scene reference frame $\mathcal{R}_0$, and where $\widetilde{p_i} = (X_i\ Y_i\ 1)^t$ are the coordinates of its projection on the image plane I_i. The *projection matrix* P_i associated to the camera $\mathcal{C}_i$ is defined as $P_i = K(R_i|t_i)$. It is function of both the *intrinsic parameters* K of the camera, and of transformations (rotation R_i and translation t_i) called the *extrinsic parameters* and characterizing the position of the camera reference frame $\mathcal{R}_i$ with respect to the scene reference frame $\mathcal{R}_0$. Intrinsic and extrinsic parameters are obtained through calibration or self-calibration procedures. The *calibration* is the estimation of camera parameters using a calibration pattern (objects providing known 3D points), and images of this calibration pattern. The *self-calibration* is the estimation of camera parameters using only image data. These data must have previously been matched by identifying and grouping all the image 2D points resulting from projections of the same 3D point. Solving the 3D reconstruction problem is then equivalent to searching for $\widetilde{x}$, given $\widetilde{p_i}$, i.e. to solve Eqn. (1) with respect to coordinates $\widetilde{x}$. Like any inverse problem, 3D reconstruction is very sensitive to uncertainty. Its resolution requires a good accuracy for the image measurements, and the choice of adapted numerical optimization techniques.

3.2. Frame expansions

Signal representation using orthogonal basis functions (e.g., DCT, wavelet transforms) is at the heart of source coding. The key to signal compression lies in selecting a set of basis functions that compacts the signal energy over a few coefficients. Frames are generalizations of a basis for an overcomplete system, or in other words, frames represent sets of vectors that span a Hilbert space but contain more numbers of vectors than a basis. Therefore signal representations using frames are known as overcomplete frame expansions. Because of their inbuilt redundancies, such representations can be useful for providing robustness to signal transmission over error-prone communication media. Consider a signal $\mathbf{x}$. An overcomplete frame expansion of $\mathbf{x}$ can be given as $F\mathbf{x}$ where F is the frame operator associated with a frame $\Phi_F \equiv \{\varphi_i\}_{i \in I}$, φ's are the frame vectors and I is the index set. The ith frame expansion coefficient of $\mathbf{x}$ is defined as $(F\mathbf{x})_i \equiv \langle \varphi_i, \mathbf{x} \rangle$, for all $i \in I$. Given the frame expansion of $\mathbf{x}$, it can be reconstructed using the dual frame of Φ_F which is given as $\widetilde{\Phi}_F \equiv \{(F^h F)^{-1} \varphi_i\}_{i \in I}$. Tight frame expansions, where the frames are self-dual, are analogous to orthogonal expansions with basis functions. Frames in finite-dimensional Hilbert spaces such as $\mathbf{R}^K$ and $\mathbf{C}^K$, known as discrete frames, can be used to expand signal vectors of finite lengths. In this case, the frame operators can be looked upon as redundant block transforms whose rows are conjugate transposes of frame vectors. For a K-dimensional vector space, any set of N, $N > K$, vectors that spans the space constitutes a frame. Discrete tight frames can be obtained from existing orthogonal transforms such as DFT, DCT, DST, etc by selecting a subset of columns from the respective transform matrices. Oversampled filter banks can provide frame expansions in the Hilbert space of square summable sequences, i.e., $l_2(\mathbf{Z})$. In this case, the time-reversed and shifted versions of the impulse responses of the analysis and synthesis filter banks constitute the frame and its dual. Since overcomplete frame expansions provide redundant information, they can be used as joint source-channel codes to fight against channel degradations. In this context, the recovery of a message signal from the corrupted frame expansion coefficients can be linked to the error correction in infinite fields. For example, for discrete frame expansions, the frame operator can be looked upon as the generator matrix of a block code in the real or complex field. A parity check matrix for this code can be obtained from the singular value decomposition of the frame operator, and therefore the standard syndrome decoding algorithms can be utilized to correct coefficient errors. The structure of the parity check matrix, for example the BCH structure, can be used to characterize discrete frames. In the case of oversampled filter banks, the frame expansions can be looked upon as convolutional codes.

3.3. Rate-distortion theory

Coding and joint source channel coding rely on fundamental concepts of information theory, such as notions of entropy, memoryless or correlated sources, of channel capacity, or on rate-distortion performance bounds. Compression algorithms are defined to be as close as possible to the optimal rate-distortion bound, $R(D)$, for a given signal. The source coding theorem establishes performance bounds for lossless and lossy coding. In lossless coding, the lower rate bound is given by the entropy of the source. In lossy coding, the bound is given by the rate-distortion function $R(D)$. This function $R(D)$ gives the minimum quantity of information needed to represent a given signal under the constraint of a given distortion. The rate-distortion bound is usually called OPTA (*Optimum Performance Theoretically Attainable*). It is usually difficult to find close-form expressions for the function $R(D)$, except for specific cases such as Gaussian sources. For real signals, this function is defined as the convex-hull of all feasible (rate, distortion) points. The problem of finding the rate-distortion function on this convex hull then becomes a rate-distortion minimization problem which, by using a Lagrangian formulation, can be expressed as

$$\frac{\partial J}{\partial Q} = 0 \quad \text{where} \quad J = R + \lambda D \quad \text{with} \quad \lambda > 0.$$

The Lagrangian cost function J is derivated with respect to the different optimisation parameters, e.g. with respect to coding parameters such as quantization factors. The parameter λ is then tuned in order to find the targeted rate-distortion point. When the problem is to optimise the end-to-end Quality of Service (QoS) of a communication system, the rate-distortion metrics must in addition take into account channel properties and channel coding. Joint source-channel coding optimisation allows to improve the tradeoff between compression efficiency and robustness to channel noise.

3.4. Distributed source coding

Distributed source coding (DSC) has emerged as an enabling technology for sensor networks. It refers to the compression of correlated signals captured by different sensors which do not communicate between themselves. All the signals captured are compressed independently and transmitted to a central base station which has the capability to decode them jointly. DSC finds its foundation in the seminal Slepian-Wolf (SW) and Wyner-Ziv (WZ) theorems. Let us consider two binary correlated sources X and Y. If the two coders communicate, it is well known from Shannon's theory that the minimum lossless rate for X and Y is given by the joint entropy $H(X,Y)$. Slepian and Wolf have established in 1973 that this lossless compression rate bound can be approached with a vanishing error probability for long sequences, even if the two sources are coded separately, provided that they are decoded jointly and that their correlation is known to both the encoder and the decoder. The achievable rate region is thus defined by $R_X \geq H(X|Y)$, $R_Y \geq H(Y|X)$ and $R_X + R_Y \geq H(X,Y)$, where $H(X|Y)$ and $H(Y|X)$ denote the conditional entropies between the two sources.

In 1976, Wyner and Ziv considered the problem of coding of two correlated sources X and Y, with respect to a fidelity criterion. They have established the rate-distortion function $R*_{X|Y}(D)$ for the case where the side information Y is perfectly known to the decoder only. For a given target distortion D, $R*_{X|Y}(D)$ in general verifies $R_{X|Y}(D) \leq R*_{X|Y}(D) \leq R_X(D)$, where $R_{X|Y}(D)$ is the rate required to encode X if Y is available to both the encoder and the decoder, and R_X is the minimal rate for encoding X without SI. Wyner and Ziv have shown that, for correlated Gaussian sources and a mean square error distortion measure, there is no rate loss with respect to joint coding and joint decoding of the two sources, i.e., $R*_{X|Y}(D) = R_{X|Y}(D)$.

3.5. Watermarking as a problem of communication with side information

Digital watermarking aims at hiding discrete messages into multimedia content. The watermark must not spoil the regular use of the content, i.e., the watermark should be non perceptible. Hence, the embedding is usually done in a transformed domain where a human perception model is exploited to assess the non perceptibility criterion. The watermarking problem can be regarded as a problem of creating a communication channel within the content. This channel must be secure and robust to usual content manipulations like lossy compression, filtering, geometrical transformations for images and video. When designing a watermarking system, the first issue to be addressed is the choice of the transform domain, i.e., the choice of the signal components that will *host* the watermark data. An extraction function $E(.)$ going from the content space $\mathcal{C}$ to the components space, isomorphic to $\mathbf{R}^N$, must then first be defined.

The embedding process actually transforms a host vector $\mathbf{V}$ into a watermarked vector $\mathbf{V_w}$. The perceptual impact of the watermark embedding in this domain must be quantified and constrained to remain below a certain level. The measure of perceptual distortion is usually defined as a cost function $d(\mathbf{V_w} - \mathbf{V})$ in $\mathbf{R}^N$ constrained to be lower than a given distortion bound d_w. Attack noise will be added to the watermark vector. In order to evaluate the robustness of the watermarking system and design counter-attack strategies, the noise induced by the different types of attack (e.g. compression, filtering, geometrical transformations, ...) must be modelled. The distortion induced by the attack must also remain below a distortion bound $d(\mathbf{V_a} - \mathbf{V}) < d_a$. Beyond this distortion bound, the content is considered to be non usable any more. Watermark detection and extraction techniques will then exploit the knowledge of the statistical distribution of the vectors $\mathbf{V}$. Given the above mathematical model, one has then to design a suitable communication scheme. Direct sequence spread spectrum techniques are often used. The chip rate sets the trade-off between robustness and capacity for a given embedding distortion. This can be seen as a labelling process $S(.)$ mapping a discrete message $m \in \mathcal{M}$ onto a signal in $\mathbf{R}^N$.

The decoding function $S^{-1}(.)$ is then applied to the received signal $\mathbf{V_a}$ in which the watermark interferes with two sources of noise: the original host signal ($\mathbf{V}$) and the attack ($\mathbf{A}$). The problem is then to find the pair of functions $\{S(.), S^{-1}(.)\}$ that will allow to optimise the communication channel under the distortion constraints $\{d_t, d_a\}$. This amounts to maximizing the probability to decode correctly the hidden message:

$$\max \text{Prob}[S^{-1}(S(m) + \mathbf{V} + \mathbf{A}) = m] \quad \text{under constraints } d_t, d_a$$

A new paradigm stating that the original host signal $\mathbf{V}$ shall be considered as a *channel state* only known at the embedding side rather than a source of noise, appeared recently. The watermark signal thus depends on the channel state: $\mathbf{S} = S(m, \mathbf{V})$. This new paradigm known as communication with side information, sets the theoretic foundations for the design of new communication schemes with increased capacity.

4. Application Domains

4.1. Introduction

The application domains addressed by the project are networked multimedia applications via their various needs in terms of image and 2D and 3D video compression, network adaptation (e.g., resilience to channel noise), or in terms of advanced functionalities such as navigation, copy and copyright protection, or tracing of illegal content usage.

4.2. Compression with advanced functionalities

Compression of images and of 2D video (including High Definition and Ultra High Definition) remains a widely-sought capability for a large number of applications. The continuous increase of access network bandwidth leads to increasing numbers of networked digital content users and consumers which in turn triggers needs for higher core bandwidth and higher compression efficiencies. This is particularly true for mobile applications, as the need for wireless transmission capacity will significantly increase during the years to come. Hence, efficient compression tools are required to satisfy the trend towards mobile access to larger image resolutions and higher quality. A new impulse to research in video compression is also brought by the emergence of new formats beyond High Definition TV (HDTV) towards high dynamic range (higher bit depth, extended colorimetric space), super-resolution, formats for immersive displays allowing panoramic viewing and 3DTV.

Different video data formats and technologies are envisaged for interactive and immersive 3D video applications using omni-directional videos, stereoscopic or multi-view videos. The "omni-directional video" set-up refers to 360-degree view from one single viewpoint or spherical video. Stereoscopic video is composed of two-view videos, the right and left images of the scene which, when combined, can recreate the depth aspect of the scene. A multti-view video refers to multiple video sequences captured by multiple video cameras and possibly by depth cameras. Associated with a view synthesis method, a multi-view video allows the generation of virtual views of the scene from any viewpoint. This property can be used in a large diversity of applications, including Three-Dimensional TV (3DTV), and Free Viewpoint Video (FTV). The notion of "free viewpoint video" refers to the possibility for the user to choose an arbitrary viewpoint and/or view direction within a visual scene, creating an immersive environment. Multi-view video generates a huge amount of redundant data which need to be compressed for storage and transmission. In parallel, the advent of a variety of heterogeneous delivery infrastructures has given momentum to extensive work on optimizing the end-to-end delivery QoS (Quality of Service). This encompasses compression capability but also capability for adapting the compressed streams to varying network conditions. The scalability of the video content compressed representation, its robustness to transmission impairments, are thus important features for seamless adaptation to varying network conditions and to terminal capabilities.

In medical imaging, the large increase of medical analysis using various image sources for clinical purposes and the necessity to transmit or store these image data with improved performances related to transmission delay or storage capacities, command to develop new coding algorithms with lossless compression algorithms or *almost* lossless compression characteristics with respect to the medical diagnosis.

4.3. Multimedia communication

Networked multimedia is expected to play a key role in the development of 3G and beyond 3G (i.e. all IP-based) networks, by leveraging higher bandwidth, IP-based ubiquitous service provisioning across heterogeneous infrastructures, and capabilities of rich-featured terminal devices. However, networked multimedia presents a number of challenges beyond existing networking and source coding capabilities. Among the problems to be addressed is the transmission of large quantities of information with delay constraints on heterogeneous, time-varying communication environments with non-guaranteed quality of service (QoS). It is now a common understanding that QoS provisioning for multimedia applications such as video or audio does require a loosening and a re-thinking of the end-to-end and layer separation principle. In that context, the joint source-channel coding and the cross-layer paradigms set the foundations for the design of efficient solutions to the above challenges.

In parallel, emerging multimedia communication applications such as wireless video (e.g. mobile cameras), multi-sensors, multi-camera vision systems, surveillance systems are placing additionnal constraints on compression solutions, such as limited power consumption due to limited handheld battery power. The traditional balance of complex encoder and simple decoder may need to be reversed for these particular applications. In addition, wireless camera sensors capture and need to large volume of redundant data wighout information exhange between the sensors. The redundancy and correlation between the captured data can then only be removed on the receiving end. Distributed source coding is a recent research area which aims at addressing these needs.

4.4. Copy protection, copyright enforcement and traitor tracing

Data hiding has gained attention as a potential solution for a wide range of applications placing various constraints on the design of watermarking schemes in terms of embedding rate, robustness, invisibility, security, complexity. Here are two examples to illustrate this diversity. In copy protection, the watermark is just a flag warning compliant consumer electronic devices that a pirated piece of content is indeed a copyrighted content whose cryptographic protection has been broken. The priorities are a high invisibility, an excellent robustness, and a very low complexity at the watermark detector side. The security level must be fair, and the payload is reduced to its minimum (this is known as zero-bit watermarking scheme). In the fingerprinting (or traitor tracing) application, user identifying codes are embedded in the host signal to dissuade dishonest users to illegally give away the copyrighted contents they bought. The embedded data must be non perceptible not to spoil the entertainment of the content, and robust to a collusion attack where several dishonest users mix their copies in order to forge an untraceable content. This application requires a high embedding rate as anti-collusion codes are very long and a great robustness, however embedding and decoding can be done off-line affording for huge complexity.

5. Software

5.1. New software and former software still in use or evolving

5.1.1. Libit

Participant: Laurent Guillo [contact person].

Libit is a C library initially developed by Vivien Chappelier and Hervé Hégou former Ph.D students in the TEMICS project-team. It extends the C language with vector, matrix, complex and function types, and provides some common source coding, channel coding and signal processing tools. The goal of libit is to provide easy to use yet efficient tools commonly used tools to build a communication chain, from signal processing and source coding to channel coding and transmission. It is mainly targeted at researchers and developpers in the fields of compression and communication. The syntax is purposedly close to that of other tools commonly used in these fields, such as MATLAB, octave, or IT++. Therefore, experiments and applications can be developped, ported and modified simply. As examples and to ensure the correctness of the algorithms with respect to published results, some test programs are also provided. (http://libit.sourceforge.net). The library is made available under the GNU Library General Public Licence and is widely used in the software developments of the project-team.

5.1.2. Soft VLC decoding library

Participant: Simon Malinowski [contact person].

This library contains a set of robust decoding tools for variable length codes (VLC) and for quasi-arithmetic codes. It contains tools for soft decoding with reduced complexity with aggregated state models for both types of codes. It also includes soft decoding tools of punctured quasi-arithmetic codes with side information used for Slepian-Wolf coding of correlated sources. This software requires the Libit library (see above) and the GMP (GNU Multiple Precision) library. The tools are currently being extended and further developed for resilient video transmission over erasure channels in the context of the collaboration with Alcatel-Lucent (see Section 7.1.3).

5.1.3. Oriented wavelet based image codec

Participant: Christine Guillemot [contact person].

This still image codec is based on oriented wavelet transforms developed in the team. The transform is based on wavelet lifting locally oriented according to multiresolution image geometry information. The lifting steps of a 1D wavelet are applied along a discrete set of local orientations defined on a quincunx sampling grid. To maximize energy compaction, the orientation minimizing the prediction error is chosen adaptively. This image codec outperforms JPEG-2000 for lossy compression. Extensions for lossless compression are being studied. This software has been registered at the APP (Agence de Protection des Programmes) under the number IDDN.FR.001.260024.000.S.P.2008.000.21000.

5.1.4. Video communication platform

Participant: Laurent Guillo [contact person].

The TEMICS project-team pursues the development of a video communication platform, called VISIUM. This platform provides a test bed allowing the study and the assessment, in a realistic way, of joint source channel coding, video modelling or video coding algorithms. It is composed of a video streaming server "Protée", of a network emulator based on NistNet and of a streaming client "Pharos":

- The streaming server allows for the streaming of different types of content: video streams encoded with the WAVIX coder as well as streams encoded with the 3D-model based coder. The video streaming server is able to take into account information from the receiver about the perceived quality. This information is used by the server to estimate the bandwidth available and the protection required against bit errors or packet losses. The server can also take advantage of scalable video streams representations to regulate the sending rate.
- The streaming client, "Pharos", built upon a preliminary version called “Criqs”, can interact with the server by executing scripts of RSTP commands. They can gather specific commands such as "play", "forward", "rewind", "pause", establish RTP/RTCP connections with the server and compute QoS information (jitter, packet loss rate,...). The client enables the plug-in of different players and decoders (video and 3D).

The server "Protée" and the client "Criqs" are respectively registered at the Agency for the Protection of Programmes (APP) under the number IDDN.FR.001.320004.000.S.P.2006.000.10200 and the number IDDN.FR.001.320005.000.S.P.2006.000.10800. This platform makes use of two libraries integrated in both the server and the client. The first one "Wull6" is an extension to IPv6 of the "Wull" library implementing the transport protocol UDP-Lite base on the RFC 3828. The second one "bRTP" implements a subset of the RTP/RTCP protocols based on the RFC3550. These two librairies are respectively registred at the Agency for the Protection of Programmes (APP) under the number IDDN.FR.001.270018.001.S.A.2004.000.10200 and the number IDDN.FR.001.320003.000.S.P.2006.000.10200.

5.1.5. Distributed video codec

Participant: Velotiaray Toto-Zaratosoa [contact person].

A distributed video coding software has been developed within the DISCOVER European research project (http://www.discoverdvc.org) with contributions of the TEMICS project-team. In particular, the TEMICS project-team has contributed to the following modules of the DISCOVER codec: side information extraction with the mesh-based and hybrid block/mesh based motion estimation techniques, Rate control, Optimal MMSE reconstruction at the decoder, and on the exploitation of the source memory with a distributed DPCM solution. The DISCOVER codec is one of the most (or even the most) efficient distributed video coding solutions nowadays. Its executable files, along with sample configuration can be downloaded from http://www.discoverdvc.org/. The results of a comprehensive performance evaluation of the DISCOVER codec can be found on the web-page http://www.img.lx.it.pt/~discover/home.html. In 2009, the codec has evolved by enhancing the Slepian-Wolf decoder in order to account for the memory present in the input sources. A two-state Gilbert-Elliott finite state process is used for modelling the sources. The model parameters are estimated with an Expectation-Maximization algorithm and the LDPC decoder has been modified in order to take into account the Gilbert-Eliott source model.

5.1.6. M3DPlayer: 3D video player

Participant: Gael Sourimant [contact person].

A 3D player - named M3DPlayer - supporting rendering of a 3D scene and navigation within the scene has been developed. It integrates as a plug-in the 3D model-based video codec of the team. From a video sequence of a static scene viewed by a monocular moving camera, the 3D model-based video codec allows the automatic construction of a representation of a video sequence as a stream of textured 3D models. 3D models are extracted using stereovision and dense matching maps estimation techniques. A virtual sequence is reconstructed by projecting the textured 3D models on image planes. This representation enables 3D functionalities such as synthetic objects insertion, lightning modification, stereoscopic visualization or interactive navigation. The codec allows compression at very low bit-rates (16 to 256 kb/s in 25Hz CIF format) with a satisfactory visual quality. It also supports scalable coding of both geometry and texture information. The first version of the software was registered at the Agency for the Protection of Programmes (APP) under the number IDDN.FR.001.130017.000S.P.2003.000.41200.

The 3D player integrated as a component to the VISIUM communication platform (see Section 5.1.4) allows remote access to a 3D sequence. In this context the video streaming server "Protée" has been modified to support 3D streams. In the near future, we would like to introduce scalability capabilities of the coded representation of the 3D models and texture, in order to enable dynamic adaptation of the transmitted bitstream to both network and terminal capabilities. The player is registered at the APP (Agence de Protection des Programmes) under the number IDDN.FR.001.090023.000.S.P.2008.000.21000.

In 2009, we focused on improving the rendering engine, based on recent OpenGL extensions, to be able to render the viewed scenes on an auto-stereoscopic display with low-end graphic cards. In our case, auto-stereoscopic display requires the rendering of eight 1920x1200 frames instead of just one for a standard display. We also plan to integrate the visualization of LDI (Layered Depth Images) and LDV (Layered Depth Videos) into the M3DPlayer.

5.1.7. Depth maps extractor

Participant: Gael Sourimant [contact person].

This software aims at estimating depth maps from multi-view videos, to provide Multi-View plus Depth (MVD) videos. MVD videos can be used to synthesize virtual views of the scene, or to render a new multi-view video with a different number of views than the original video, for instance in an auto-stereoscopic display setup. This software has been developed in the context of the DGE/Region research project Futurim@ges. In the near future, we plan to compare this software to the Depth Estimation Reference Software from the MPEG 3DV group, in terms of virtual views synthesis quality.

5.1.8. WSVC: Wavelet-based Scalable Video Codec

Participants: Ronan Boitard, Christine Guillemot, Laurent Guillo [contact person].

A wavelet based video codec, called WSVC, based on a motion compensated t+2D wavelet analysis, has been developed ans is in the process of being filed at the Agency for the Protection of Programmes (APP). the codec is based on a former version called WAVIX which was registered at the Agency for the Protection of Programmes (APP) under the number IDDN.FR.001.160015.000.S.P.2003.000.20100. As Wavix, the codec supports three forms of scalability: temporal via motion-compensated temporal wavelet transforms, spatial scalability enabled by a spatial wavelet transform and SNR scalability enabled by a bit-plane encoding technique. A so-called /extractor/ allows the extraction of a portion of the bitstream to suit a particular receiver temporal and spatial resolution or the network bandwidth. the codec has significantly evolved by re-writing and enhancing a number of functions, by adding new temporal transforms and motion fields coding tools. This codec will be used in the context of the collaboration with Alcatel-Lucent (see Section 7.1.3).

5.1.9. Watermarking platform

Participants: Mathieu Desoubeaux, Caroline Fontaine [contact person].

This software platform aims at integrating as plug-ins a set of functions, watermark detection and robust watermarking embedding and extraction, for different applications (fingerprinting and copyright protection). These plug-ins include the Broken Arrows software and the Chimark2 software. The Broken Arrows software has been developed in collaboration with the CNRS-Gipsa-lab in Grenoble, in the context of the international challenge BOWS-2 (Break Our Watermarking System - 2nd Edition). The source code has been registered at the Agency for the Protection of Programmes (APP) under the number IDDN.FR.001.170012.000.S.P.2008.000.41100. The software is available as an open-source code distributed under the INRIA-CNRS license CECILL (see http://bows2.gipsa-lab.inpg.fr). The

Chimark2 software developed in the context of the RNRT-Diphonet project is a robust image watermarking tool and registered at the Agency for the Protection of Programmes (APP) under the number IDDN.FR.001.480027.001.S.A.2002.000.41100. The robust watermarking tool has then be extended in the context of the IST-Busman project for marking video signals. The platform also aims at supporting different types of attacks (collusion, compression, crop, de-synchronization, ...) and at demonstrating the performance of the watermarking tools in presence of these attacks.

This software platform has continued to evolve in 2009. The platform can now be built with the CMake cross platform "make", and we can provide autonomous installers to install the platform on new computers. To enable the portability under Linux OS and Mac OS, new OpenCV functionnalities have been developed, based on the use of ffmpeg. Functionalities have been extended and a set of benchmarking scripts have been written in Python, to provide a benchmarking platform for today and future watermarking techniques. The embedding and traitor tracing techniques we already implemented last year, based on an adaptation of the Broken Arrows software to video coupled with Tardos anti-collusion codes, have been improved. Our software now offers a multi-tab player which is able to open AVI, CIF and QCIF movies, but also many image formats as bmp, pgm, or jpeg.

5.2. ADT PICOVIN

Participants: Ronan Boitard, Laurent Guillo [contact person], Gabrielle Ombrouck, Tangui Poirier, Yann Serrand.

The ADT Picovin is a technological development action, which works closely with the project-team TEMICS. This new development structure gives its support to the project-team to carry out a new and efficient video codec. This support is twofold. The project-team TEMICS provides the ADT with innovative algorithms that are first adapted and integrated in the current codec. Then the resulting codec is evaluated thanks to a large base of videos in order to measure how significant these new contributions are and to detect as soon as possible unexpected behaviour. Hence, step by step, the emerging new codec performs better and better. Throughout this process, it is compared with the still evolving competing video codecs, for instance JM and KTA.

In April 2009, the ISO/MPEG working group issued a call for evidence with the expectation that responses would report the existence of technologies that would be able to efficiently take into account new challenges in video coding such as ultra high definition (e.g. 4Kx2K displays). Responses were reviewed during the 89th MPEG meeting in London in July 2009. Significant improvements were achieved and a call for proposal will be probably issued in October 2009 during the 90th MPEG meeting in Xian, China.

The ADT Picovin aims at providing a video codec that meets the requirements, or a significant subset, of the coming call for proposal. In 2009, we focused on the integration of algorithms, which significantly improved spatial and temporal predictions, and the development of an evaluation platform, which can cope with very demanding test conditions. This new development structure started in October 2008 and will last three years. Three junior engineers and one permanent engineer from the SED Rennes (development and experimentation department of INRIA Rennes) take part to the ADT. It is supported by the technological development department of INRIA.

6. New Results

6.1. 3D representation and compression of video sequences

3DTV and Free Viewpoint Video (FVV) are now emerging video formats expected to offer an anhanced user experience. The 3D experience consists either in 3D relief rendering called 3DTV, or interactive navigation inside the content called FTV (Free viewpoint TV). 3D information can be easily computed for synthetic movies. However, so far, no professional nor consumer video cameras capable of capturing the 3D structure of the scene are available (except of course for Z-cameras prototypes). As a consequence, 3D information representing real content have to be estimated from acquired videos, using computer vision-based algorithms.

This is the scope of the first research axis described below which focuses on depth maps extraction. Once the depth information has been extracted, the resulting videos with the associated depth maps must be coded to be stored or transmitted to the rendering device. 3D representations of the scene as well as associated compression techniques must be developed. The choice of the representation of the 3D is of central importance. On one hand, it sets the requirements for acquisition and signal processing. On the other hand, it determines the rendering algorithms, degree and mode of interactivity, as well as the need and means for compression and transmission. This is the scope of the next two research axis below.

6.1.1. Depth map extraction for multi-view sequences

Participants: Claude Labit, Gael Sourimant.

In this study - supported by the Futurim@ges project - we focus both on extending our 3D model-based video codec to multi-view image sequences, and on post-processing this estimated 3D information to provide attractive videos purified from usual 3D artifacts (badly modelled occlusions, texture stretching, etc.).

For general multi-view image sequences, the scene structure can be estimated not only over time if the camera is moving, but also over space. At time t, the N acquired views can be used to compute the 3D structure of the scene more easily since the camera bank calibration is supposed to be known. We are now designing algorithms to integrate this space-based reconstruction to the classical time-based reconstruction, in order to produce higher quality 3D models.

Moreover, in the highly constrained context of registered fronto-parallel cameras, we proposed a depth map extraction algorithm based on the state-of-the-art studies in Stereo Matching (depth map estimation from a couple of stereoscopic images). Here the camera poses do not need to be estimated. The depth of each pixel only depends on the horizontal measured motion field between neighbouring images (in the space sense). Our algorithms seeks to find this motion field by local matching using gradient and intensities differences between images. This initial flow is then globally refined using a loopy Belief Propagation optimization and robust plane fitting, which are constrained by a Mean-Shift segmentation of the input images to preserve the depth discontinuities.

Once the 3D structure of the scene has been estimated (and stored as a pool of depth maps), it is post-precessed using our Java-based software *M3dEncoder2*. One goal in this post-precessing steps is to generate auto-stereoscopic-ready videos, so that they can be displayed on available auto-stereoscopic screens (at present time we aim the Newsight™ and the Philips™ 3D screens).

6.1.2. 3d representation for multi-view video sequences based on a soup of polygons

Participants: Thomas Colleu, Claude Labit.

This study is carried out in collaboration with France Telecom under a CIFRE contract. The aim is to study a new data representation for multi-view sequences. The data representation must allow real-time rendering of good quality images for different viewpoints, possibly virtual (ie non acquired) viewpoints. Moreover, the representation should be compact and efficiently compressed so as to limit the data overload compared with encoding each of the N video sequences using traditional 2D video codec. In 2008, a new representation that takes as an input multi-view video plus depth (MVD), and results in a polygon soup was proposed. A polygon soup is a set of polygons that are not necessarily connected to each other. Each polygon is defined with image plus depth data and rendered as a 3D primitive thanks to a graphics processing unit (GPU). The polygons are actually quadrilaterals (aka quads) and they are extracted with a quadTree decomposition of the depth maps. The advantages of using polygonal primitives instead of point primitives traditionally used in other representations were shown. Also many redundancies across the viewpoints were reduced so as to obtain a compact representation.

In 2009, first, the construction of the representation has been improved in order to reduce the number of quads and eliminate corona artifacts around depth discontinuities [20],[41]. The main idea of this improvement is to discard as much small quads as possible from all viewpoints since these quads considerably increase the data load and potentially contain corona artifacts. Second, the rendering step was improved. Indeed, when several quads overlap, one should define a strategy to determine which colour should be drawn. It was decided to adaptively merge the overlapping quads according to the distance between the desired viewpoint and the one associated to the quad being merged. This kind of merging is usually referred as view-dependent texture mapping. Finally, a coding scheme was proposed for encoding the extracted polygon soup representation. This coding scheme takes advantage of the quadTree structure; performs a predictive coding of the depth information; and encodes all information using Context Adaptive Binary Arithmetic Coding (CABAC). The performance of the coding scheme was compared with the encoding of the depth maps using JPEG2000.

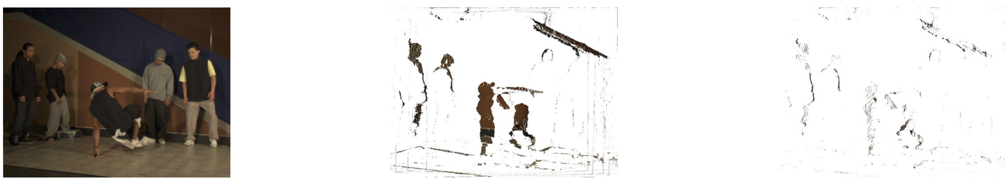
The results show that three viewpoints can be represented with an increase of only 10

6.1.3. 3d representation for multi-view video sequences based LDI representations

Participants: Christine Guillemot, Vincent Jantet.

This study is carried out in collaboration with INSA/IETR (Luce Morin). A multi-view video is a collection of video sequences captured for the same scene, synchronously by many cameras at different locations. Associated with a view synthesis method, a multi-view video allows the generation of virtual views of the scene from any viewpoint. This property can be used in a large diversity of applications, including Three-Dimensional TV (3DTV), Free Viewpoint Video (FTV), security monitoring, tracking and 3D reconstruction. The huge amount of data contained in a multi-view sequence motivates the design of efficient compression algorithms.

The compression algorithm strongly depends on the data representation, which in turn very much depends on the view synthesis methods. View synthesis approaches can be classified in two classes: Geometry-based rendering (GBR) approaches and Image-based rendering (IBR) approaches. GBR methods use a detailed 3D model of the scene. These methods are useful with synthetic video data but they become inadequate with real multi-view videos, where 3D models are difficult to estimate. IBR approaches are an attractive alternative to GBR. They allow the generation of photo-realistic virtual views. The Layer Depth Image (LDI) representation is one of these IBR approaches. In this representation, pixels are no more composed by a single color and a single depth value, but can contain several colors and associated depth values. This representation reduces efficiently the multi-view video size, and offers a fast photo-realistic rendering, even with complex scene geometry. Various approaches to LDI compression have been proposed based on classical LDI's layers constructions. The problem is that layers generated are still correlated, and some pixels are redundant between layers. In 2009, we have developed an Incremental LDI construction (I-LDI) method to reduce the inter-layer correlation [28]. The number of layers is significantly reduced for an equivalent final rendering quality.



(a) 1st layer (b) 2nd layer (c) 3rd layer

Figure 1. First layers of an I-LDI *frame. 8 inputs Views+Depth are used for the generation.*

Techniques have also been designed to overcome visual artifacts, like sampling holes resulting from disocclusion, cracking effects and ghosting artifacts. The LDI representation itself is one solution against disocclusion. Information about occluded texture is stored in extra layers, and used during the rendering stage to fill disocclusion holes. Many cracking effects can be observed in the rendered texture, due to sampling and pixelizing. The implemented solution makes use of interpolation (inpainting) techniques. Detected cracks are filled with the median color estimated on a sliding window. Ghosting effects may appear around depth discontinuity, because pixels along object boundaries receive information from both foreground and background colors. An

Edge detector performed on the depth map, followed by a local foreground / background classification, permits to isolate potentially blended pixels, and reduce the ghosting effects. Tests on the Breakdancers and Ballet MVD data sets show that extra layers in I-LDI contain only 10% of first layer pixels, compared to 50% for LDI. I-LDI Layers are also more compact, with a less spread pixel distribution, and thus easier to compress than LDI (see figure above). Visual rendering is of similar quality with I-LDI and LDI.

6.2. Sparse representation, compression and interaction with indexing

6.2.1. Sparse signal representation algorithms

Participants: Jean-Jacques Fuchs, Cedric Herzet, Gagan Rath.

Sparse representations and compressed sensing continue to be fashionable domains in the signal processing community and numerous pre-existing areas are now labeled this way. Sparse approximation methods aim at finding representations of a signal with a small number of components taken from an overcomplete dictionary of elementary functions. The problem basically involves solving an under-determined system of equations with the constraint that the solution vector has the minimum number of non-zero elements. Except for the exhaustive combinatorial approach, there is no known method to find the exact solution under general conditions on the dictionary. Among the various algorithms that find approximate solutions, pursuit algorithms (matching pursuit, orthogonal matching pursuit or basis pursuit) are the most well-known.

In 2009, we have continued to apply our research in this area to the estimation-detection context [25] and further develop fast-algorithms for criterion [24], [42] that are distinct from the ubiquitous ℓ_2 penalized ℓ_1 criterium. Using the second order cone programming algorithm, we apply sparse representation techniques to complex data in the array pattern synthesis problem. It allows to extend min-max techniques to arbitrary array geometries as well as the sparsity of the number of radiating elements required to achieve a given pattern. This last feature is often of utmost importance and is generally handled in a combinatorial way.

We have also focussed on the application of statistical tools to the resolution of the sparse representation problem. First, we have shown that the standard sparse representation problem can be regarded as the limit case of a MAP problem involving Bernoulli-Gaussian (BG) variables. This connexion gives new insights into existing Bayesian algorithms and paves the way for the design of new ones. In particular, we have proposed new sparse representation algorithms based on a mean-field relaxation of the BG MAP problem. These algorithms are shown to give the best performance among several algorithms available in the literature while having the same complexity order.

6.2.2. Anisotropic basis for image compression

Participants: Angélique Drémeau, Jean-Jacques Fuchs, Christine Guillemot, Cedric Herzet.

Closely related to the sparse representation problem is the design of dictionaries adapted to “sparse” problems. The sparsity of the signal representation indeed depends on how well the bases match with the local signal characteristics. The adaptation of the transform to the image characteristics can mainly be made at two levels: i) in the spatial domain by adapting the support of the transform; ii) in the transformed domain by adapting the atoms of the projection basis to the signal characteristics. In 2009, a first image compression algorithm has been developed based on anisotropic directional DCT (DDCT) bases. A set of image bases is constructed as the concatenation of local anisotropic rectangular DDCT bases. the approach extends the DDCT concept to rectangular bases for which the support is defined using a bintree segmentation. Dynamic programming is then used to select a basis from this set according to a rate-distortion criterion. The bintree segmentation which locally adapts the support and the direction of the transform increases the number of possible image bases, which are then more likely to catch the local properties of the image. In a second approach, the transform basis is selected (in a rate-distortion sense) in a set of bases made up of the concatenation of local multi-scale anisotropic (rectangular) bases. The sets of local bases are optimized in a sparsity-distortion sense. The resulting vectors are then quantized and entropy coded in a jpeg-like manner to assess the rate-distortion performance of the basis.

The problem of dictionnary learning of sets of bases has also been studied. This problem has already been largely addressed in the literature. An existing technique to construct a dictionary made up of the union of orthonormal bases, and based on a classification of the training data into P classes has been revisited. The method has in particular been placed in a probabilistic framework by considering the training data as realizations of a mixture of Gaussians. The learning task is thus reformulated as a MAP estimation problem which is then solved by an EM-algorithm procedure. So far, the training algorithm has been tested only on synthetic data generated according to the same model of mixture of Gaussians as considered in the training algorithm. The validation of the algorithm by taking real images as training data is the next step.

6.2.3. Texture synthesis and prediction based on sparse approximations

Participants: Jean-Jacques Fuchs, Christine Guillemot, Aurélie Martin, Mehmet Turkan.

The problem of texture prediction can be regarded as a problem of texture synthesis. Methods based on sparse approximations, and using orthogonal matching pursuit and basis pursuit algorithms, have been investigated for this texture synthesis and prediction problem. The problem is looked at as a problem of texture synthesis (or inpainting) from noisy observations taken from a causal neighborhood. The goal of sparse approximation techniques is to look for a linear expansion approximating the analyzed signal in terms of functions chosen from a large and redundant set (dictionnary). In the methods developed, the sparse signal approximation is run in a way that allows for the same operation to be done at the decoder, i.e. by taking the previously decoded neighborhood as the known support. The sparse signal approximation is thus run with a set of *masked* basis functions, the masked samples corresponding to the location of the pixels to be predicted. The decoder proceeds in a similar manner by running the algorithm with the *masked* basis functions and by taking the previously decoded neighborhood as the known support. In a first step, we have considered dictionaries based on classical waveforms such as DCT and DFT. The approach integrated in an H.264 based encoder has shown gains around 7

Another method of dictionary construction based on texture patches taken in a causal neighborhood of the region to be approximated has been developed. This approach can be regarded as an extension of template matching widely used for image inpainting. Significant spatial prediction gains have been shown compared to static DCT or DFT dictionnaries. Similarly, high temporal prediction gains have been obtained compared to classical block matching techniques. The method has been further improved by locally - and in a rate-distortion sense- adapting the approximation support [38]. These prediction and dictionary construction methods have been validated both for spatial and temporal prediction in the context of the ADT Picovin. Local texture analysis and classification techniques have been developed and assessed in order to better take into account the presence of discontinuities, of edges, and more generally the local texture characteristics. Methods of residue coding based on adaptive rate-distortion optimized directional transforms have also been developed.

6.2.4. Perceptual 2D and 3D video coding

Participants: Josselin Gauthier, Christine Guillemot, Olivier Le Meur.

In collaboration with Zhenzhong Chen from the National Technical University of Singapore, we have pursued a study on video coding exploiting perception and foveation models developed in 2008. Due to the spatial and temporal masking effects, the human visual system has the limitation on the perceptibility of certain levels of noise. Since the human visual system is space-invariant where the fovea has the highest density of sensor cells on the retina, the visual acuity decreases with increased eccentricity relative to the fovea. We have thus worked on the design of a foveated just-noticeable-distortion (JND) model. In contrast to traditional JND methods which exploit the visibility of the minimally perceptible distortion but assume the visual acuity to be consistent over the image, a foveation model is incorporated in the spatial and temporal JND models. The foveated JND model is developed by combining the spatial JND as a function of luminance contrast and spatial masking effect, the temporal JND to model the temporal masking effect, and a foveation model to describe the relationship between the visibility threshold and eccentricity relative to the fovea. Since the perceptual acuity decreases with increased eccentricity, the visibility threshold of the pixel of the image increases when the distance between the pixel and the fixation point increases. The spatio-temporal JND model can thus be improved by accounting for the relationship between visibility and eccentricity. Associated with the proposed foveated JND model, more imperceptible distortion can be tolerated in the contaminated image.

The foveated JND model has been used for H.264/AVC video coding. Bit allocation and rate-distortion optimization are performed according to the foveated JND profile. The regions with higher visibility thresholds are coded with larger quantizers since these regions can tolerant higher distortion. The saved bit rate can be used to improve the quality of the regions which cannot tolerant high distortion. Therefore, the subjective quality of the whole image is improved. The performance of the foveated JND model has been assessed with subjective tests following the corresponding protocols in Rec. ITU-R BT.500. The subjective tests have demonstrated the validity of the FJND model which leads to better perceptual quality of the reconstructed video for the same rate constraints. This study will be pursued and extended to 3D video coding and rendering. For both previously mentionned subjects, it might be useful to consider new low-level and high-level visual features in order to significantly improve the performance. Among these new features, we can mention three of them that it would be interesting to consider first: the depth information, the type of the scene and the visual interest of the different areas of the video (saliency map).

6.2.5. Lossless coding for medical images

Participants: Claude Labit, Jonathan Taquet.

Usual techniques, such as lossless JPEG, LS-JPEG or lossless JPEG2000, are currently proposed and have been compared in a preliminary study (with ETIAM as industrial partner). The first obtained results show the opportunity to launch a new prospective research study in order to

- increase the present performances of traditional lossless compression schemes which are quite low (the average compression ratio of 3.8:1 obtained over a large medical imaging database is usually mentioned in tutorial books);
- extend these techniques, which are essentially based on block decomposition and spatio-frequential transforms, to oriented wavelet approaches (such as oriented wavelet transforms and quincunx sampling *Lifting algorithm*) and their adaptation to telemedicine and medical imaging storage applications.

Afterwards, during the project (J. Taquet's thesis), we propose to take into account coding algorithms commonly used in multimedia domain and so, adapt these techniques formerly developped in the lossy compression framework:

- explore, taking out the too restrictive constraint of exact digital reconstruction, *almost* lossless compression schemes i.e compression techniques which generate no irreversible degradations for the medical diagnosis; these techniques have to be selectively evaluated with respect to the medical imaging sources processed;
- propose new functionalities such as Region-Of-Interest analysis and coding (ROI-Based compression) which enable a simultaneous optimization of several rate-distortion functions adapted to the a-priori defined ROIs.

A first step of this study explores the improvements that might be expected for volumetric medical images, like computed tomography (CT) or magnetic resonance imaging (IRM). These images are composed by a succession of slices regularly sampled in the 3D space and are stored in sequences of 2D images with 12 to 16 bits per pixel. For diagnosis purpose, the compression has to be performed without or with controlled digital losses. To improve distant image restitution, the data flows should also allow a progressive representation and random access.

Based to a consistent database of several biomedical images, which combines CT and MRI images of various origins, we compare usual lossless (or near lossless) 2D algorithms (CALIC, JPEG-LS, JPEG 2000, SPIHT) with their extensions to 3D sources, in order to evaluate the coding gain by using the third dimension, and the enhancement of sophisticated 3D algorithms in comparison to basic methods. It shows that results were quite variable, depending on how the original image was acquired. The impacts of the noise and of the sampling distance in the third dimension were noticeable. The most noisy images were better compressed with 2D predictive algorithms (CALIC's files size are 3.7

This Ph-D thesis, partially supported by a research grant of Bretagne Council, will take also benefit of the IHE-Europe technical coordination (« Integrating the Heath Care Enterprise ») hosted at Irisa/INRIA Rennes Bretagne-Atlantique research center and of the presence, at Rennes, of several SMEs as industrial partners, such as ETIAM, SME leader in Europe developing innovative tools for multimedia connectivity and medical imaging communication.

6.2.6. Feature extraction for joint compression and local description

Participants: Christine Guillemot, Fethi Smach, Joaquin Zepeda.

The objective of the study initiated in 2007, in collaboration with TEXMEX, is to design signal representation and approximation methods amenable to both image compression (that is with sparseness properties), feature extraction and description. Feature extraction requires the image representation to be covariant under a set of admissible transformations. The Gaussian scale space is thus often used for description, however it is not amenable to compression. One robust (non-sparse) descriptor based on the Gaussian scale space is the SIFT descriptor, usually computed in affinely normalized selected regions. A recent approach referred to as *Video Google* tackles the high dimensionality problem of this descriptor by forming a single sparse descriptor obtained from multiple input non-sparse SIFT descriptors. The approach consists in vector quantizing each SIFT descriptor on codewords called visual words, and then taking a (weighted) histogram of codeword indices. It allows using the principles of *inverted files*. Inverted file indices provide a solution to the indexation of high dimensional data (specifically, textual documents) by representing the data as sparse vectors. Document similarity calculations are thus carried out efficiently using the scalar products between these sparse vectors.

We have developed a related approach which constructs a sparse descriptor (called visual sentences), by using a sparse approximation of each SIFT descriptor rather then using a simple vector quantization [39]. The aim is to tackle the problem of SIFT descriptor high dimensionality, while retaining the local property of the input descriptors. The obtained descriptors retain the local characteristic of the input descriptors rather than forming a single global descriptor, while still enabling the use of inverted file type indices which provide a solution to the indexation of high dimensional data by taking advantage of sparse vectors properties. Indeed, document similarity calculations are thus carried out efficiently using the scalar products between these sparse vectors. The approach has been assessed in the context of local querying to the *Video Google* one, where multiple input SIFT descriptors are agregated into a single sparse descriptor, resulting in the loss of description locality.

However, using sparse vectors instead of original vectors in the computation of the similarity score required in an image retrieval system or in nearest neighbor (NN) search raises a new problem. The residual transformations following the geometrical region normalization cause instabilities in the support (positions of nonzero coefficients) of the sparse vector. The instability problems can severely impact the similarity score between regions and therefore the ranking performance of the NN search task, especially when using the inner-product or correlation as the similarity measure. Inner-products between query and data base sparse vectors may significantly differ from the correlation between the original signal vectors, which we consider as a reference measure in our study. A new method has thus been introduced that makes use of sparse image representations to search for approximate nearest neighbors (ANN) under the normalized inner-product distance. The approach relies on the construction of a new sparse vector designed to approximate the normalized inner-product between underlying signal vectors. The resulting ANN search algorithm shows significant improvement compared to querying with the original sparse vectors, approach considered in the literature for content-based image search. A transform has then been introduced in order to uniformly distribute the input dataset on the unit sphere while preserving relative angular distances. This transformation has been shown to further improve the performance and complexity of the ANN search task.

The problem of dictionary design has then been addressed by developing a novel approach to construct dictonaries in a way that a different dictionary is used at each iteration of the decomposition, and that these *iteration tuned dictionaries* satisfy some desirable properties. This method and associated algorithm gave promising results and should be validated next year for compression and indexing purposes.

6.2.7. Denoising in the presence of oversampling

Participant: Jean-Jacques Fuchs.

Removing noise from a signal is feasible if some prior information is available. Oversampling is one instance where this is the case. For the noise-free signal, the information contained in the samples is then redundant, there exist implicit relations between the samples. One can thus remove part of the noise by imposing these relations on the noisy samples. Sparsity is another such instance. If one knows that the noise-free signal is sparse in a given non-redundant or redundant basis, then seeking a sparse representation for the signal amounts to remove part of the noise. In case of over-sampling, one can, for instance, project the observation onto the subspace containing "all" the band-limited signals sampled at the same sampling instants, which is lower dimensional only in case of oversampling. This projection matrix is difficult to build and to define, we seek a simple way to obtain it. The research is performed in an underwater-acoustics domain [22] but it is also of interest in a detection context [23] and of course in image processing where it is well documented.

Let $x = s + e$ be the noisy sampled observation vector of dimension N, with s the band-limited sampled signal and e the white noise vector. If P denotes the matrix associated with the aforementioned projection, then $Ps = s$ by definition. The problem to solve is

$$\min_P E\{e^T Pe\} \quad \text{under} \quad Ps = s \quad \text{with} \quad P \text{ in the set of projection matrices on } R^N,$$

and since the expectation of $e^T Pe$ is $r\sigma_e 2$ with r the rank of P, one actually seeks the projection with minimal rank that satisfies $Ps = s$. The projector can be seen as achieving a low-pass filtering and we show how to use the discrete time Fourier transform to perform it.

6.3. Joint source-channel and distributed coding

6.3.1. Joint source-channel coding based arithmetic codes with erasures

Participants: Simon Bos, Christine Guillemot, Simon Malinowski, Aline Roumy.

In the context of image or video transmission over noisy channels, the compression stage makes the transmitted message very sensitive to channel noise. Two kinds of codes are widely used in order to compress the signal : variable length and quasi-arithmetic codes. In the TEMICS project-team, a state model to be used in soft-decision (or trellis) decoding of variable length codes and quasi-arithmetic codes has been developed. This state model, compared to the optimal model, presents the advantage of achieving close to optimal decoding performance for a much lower complexity. This model is also well suited for the introduction and exploitation of a priori information (or side information) on the source in order to favor the selection of synchronous and correct paths in the soft decoding process. Two main joint source-channel coding strategies have thus been developed in the team. The first one consists in adding synchronisation markers at some instants of the decoding process. The second one is based on adding constraints on the number of decoded symbols at several instants of the decoding process. These approaches turn out to outperform widely used techniques, such as the popular approach based on the introduction of a forbidden symbol to quasi-arithmetic codes. These results have been obtained for AWGN channels, which represents transmission over mobile channels for example.

When dealing with image or video transmission over a packet based network (internet for instance), the issues are not the same. When a message (i.e., a bitstream) is sent over an Internet channel, it is subject to packet losses. This kind of channel can be modeled by an erasure channel (BEC) characterized by an probability of erasure. Techniques that are mainly used in practical schemes concerning the protection of video streams for transmission over packet-based network rely on forward error correcting (FEC) codes (Raptor codes, LDPC-Staircase...). Residual losses at the output of the FEC decoder then correspond to regions - which may be large - in images which have to be concealed. Concealment methods often make use of spatial and/or temporal interpolation techniques. In the context of the joint laboratory between INRIA and Alcatel Lucent, we are investigating methods of robust decoding of arithmetic codes in presence of erasures. The aim is to further improve the robustness of the video compression algorithms by confining the effects of losses to reduced areas, thus reducing the visual artefacts which result from concealment applied on large image areas. Methods of inpainting for recovering erased symbols are also investigated.

6.3.2. Slepian-Wolf coding of memory sources

Participants: Christine Guillemot, Aline Roumy, Velotiaray Toto-Zarasoa.

The problem of asymmetric and symmetric distributed coding for memoryless correlated sources has been addressed in 2008. Practical schemes based on LDPC codes have been designed. The minimum achievable rates depend on the correlation between the sources, but also on their respective distributions. Knowing the non-uniform source distribution, one can achieve lower rates using the same code. Syndrome-based turbo and LDPC decoders have been adapted so as to take into account the non-uniformity of the sources. The parameters of the source distributions are estimated with an Expectation-Maximization algorithm which is run jointly with the syndrome-based turbo and LDPC decoders. The case of Finite Memory Sources (FMS) has then been addressed where the sequence of symbols is assumed to be generated by an ergodic probabilistic finite state process. The distribution of the source is assumed to be dependent only on the current state of the model. A two-state Gilbert-Elliott finite state process has been considered. The Gilbert-Elliott model parameters are estimated with an Expectation-Maximization (EM) algorithm. The sequence of states of the FMS and the model parameters are iteratively estimated by the decoder together with the compressed sequence of input symbols of the source X. Sources with memory represent real sources more accurately than non-uniform sources, and their application covers wide range of systems, including biometrics or sensor networks. The rate gains obtained by accounting for the nonuniformity and the memory of the source have first been assessed experimentally with theoretical sources. The probability estimators and the modified LDPC decoders have then been integrated in a transform-domain distributed video codec, showing improved PSNR versus rate performance.

6.4. Image and video watermarking and security

6.4.1. Watermarking Robustness and Security

Participants: Caroline Fontaine, Fuchun Xie.

We have proposed several improvements of the Broken Arrows watermarking technique, strengthened its robustness and security. This work has been motivated by the BOWS-2 contest results; A. Westfeld proposed a quite efficient attack to cleanly remove the embedded watermark. This attack was based on a robustness weakness and a security attack strategy. The first step was to use linear regression to estimate the "original" (non watermarked) image from the watermarked one. We proposed a variant of Broken Arrows called AWC that counters this kind of robustness attack. The second step of Westfeld's attack was to observe a large set of watermarked images, to estimate corresponding original (non-watermarked) images through the aforementionned linear regression, then to estimate the embedded watermarks and use a clustering algorithm to classify them into bins. To attack a watermarked image, he substracted an average of the estimated watermarks lying into the corresponding bin. Recently, another security attack strategy has been proposed in the literature, using subspace estimation. We proposed a counter-measure to make both these attack fail.

6.4.2. Fingerprinting

Participants: Ana Charpentier, Caroline Fontaine.

Fingerprinting (also known as user forensics, traitor tracing, transactional watermarking, content serialization ...) aims at hiding in a robust and imperceptible way, an identifier of the consumer. The goal is to enable traceability of the content and to find back dishonest users who have illegally redistributed the content (for instance, posting it in a P2P network). Fingerprinting has a long history in research but real applications have slowly emerged this year: DRM systems over protect the content and thus they are not user-friendly. Content distribution could get rid off DRM thanks to fingerprinting used as a dissuasive weapon. This is a hot topic in the watermarking community. Fingerprinting is a difficult problem because it is a cross-design merging two layers: the fingerprinting code (the set of identifiers) and the watermarking technique hiding the identifiers in content.

A new trend in this domain is the probabilistic fingerprinting codes: The user identifiers are random binary strings with a secret statistic structure. G. Tardos introduced this concept in 2003, but it took time for the community to recognize his work as a major breakthrough. Tardos seminal work proposes a code design and proves that his performances are optimal. However, no clue is given on why it works so good, and on the setting of the parameters. Last year, we solved this point and proposed some improvements, deriving some optimal settings when the number of colluders is known, or when their collusion strategy is known. Following the steps of this previous work, in 2009, we have extended our results as follows:

- we proposed to integrate in the accusation process a dynamic estimation of the collusion strategy,
- and we provided more interesting accusation functions in the case where the collusion strategy.

Hence, we are able to provide a more powerful and reliable tracing technique. Our solution also enables us to reduce the length of the user identifiers which have to be embedded in the videos.

7. Contracts and Grants with Industry

7.1. Industrial contracts

7.1.1. CIFRE contract with France Telecom on the problem of 3D video representation

Participants: Thomas Colleu, Claude Labit.

- Title : 3D video representation for 3DTV and Free viewpoint TV.
- Research axis : § 6.1.2.
- Partners : France Télécom, Irisa/Inria-Rennes.
- Funding : France Télécom.
- Period : Oct.07-Sept.10.

This contract with France Telecom R&D (started in October 2007) aims at investigating the data representation for 3DTV and Free Viewpoint TV applications. 3DTV applications consist in 3D relief rendering and Free Viewpoint TV applications consist in an interactive navigation inside the content. As an input, multiple color videos and depth maps are given to our system. The goal is to process this data so as to obtain a compact representation suitable for 3D TV and Free Viewpoint functionalities. In 2008, we have developed a multi-view video representation based on a global geometric model composed with a soup of polygons. In 2009, the construction of the representation has been improved in order to reduce the number of quads and eliminate artifacts around depth discontinuities as well as the corresponding rendering approach (see Section 6.1.2). A coding algorithm adapted to the polygon soup representation has been developed. The performance of the coding scheme has been compared with the encoding of the depth maps using JPEG2000.

7.1.2. CIFRE contract with Thomson on the problem of spectral deconvolution for video compression and protection

Participants: Jean-Jacques Fuchs, Christine Guillemot, Aurélie Martin.

- Title : Spectral deconcolution: application to compression
- Research axis : § 6.2.3.
- Partners : Thomson, Irisa/Inria-Rennes.
- Funding : Thomson, ANRT.
- Period : Nov.06- Oct.09.

This CIFRE contract concerns the Ph.D of Aurélie martin. The objective of the Ph.D. is to develop image spectral deconvolution methods for prediction in video compression schemes. Closed-loop spatial prediction has indeed been widely used in video compression standards (H.261/H.263, MPEG-1/2/4, H.264). In H.264 used for digital terrestrial TV, the prediction is done by simply "propagating" the pixel values along the specified direction. This approach is suitable in presence of contours, the directional mode chosen corresponds to the orientation of the contour. However, it fails in more complex textured areas. The spatial image prediction approach developed in the context of this collaboration relies on sparse approximations based on matching pursuit algorithms. In a first step dictionaries based on classical waveforms such as DCT and DFT have been considered. In 2009, a method to enhance to dictionaries taking into account a phase shift between the signal to be approximated and the waveforms taken as dictionary atoms has been developed showing further rate saving (up to 9

7.1.3. Collaboration with Alcatel on robust video compression

Participants: Simon Bos, Christine Guillemot, Laurent Guillo, Aline Roumy.

- Title: Self adaptive video codec
- Research axis: 6.3.1
- Funding: Joint research laboratory between INRIA and Alcatel
- Period: Nov. 2008 - Oct. 2011.

In the framework of the joint research lab between Alcatel-Lucent and INRIA, we participate in the ADR (action de recherche) Selfnets (or Self optimizing wireless networks). More precisely, we collaborate with the Alcatel-Lucent team on a self adaptive video codec. This collaboration concerns the Ph.D. of Simon Bos. The goal is to design a video codec, which has the intrinsic knowledge of the dynamic video quality requirements, and which is able to self-adapt to the existing underlying transport network. In this approach, the video codec has to include:

- Means to dynamically "sense" the underlying transport channel (e.g BER, PER, Markov model)
- Means at the encoder to adapt dynamically the output bitrate to the estimated channel throughput and to the effective transport QoS while maintaining the video quality requirements.
- Means at the decoder to be resilient to any remaining packet losses. enditemize

7.2. National contracts

7.2.1. DGE/Region-FUTURIMAGE

Participants: Christine Guillemot, Vincent Jantet, Claude Labit, Gael Sourimant.

- Title : 3D multi-view video transmission and restitution.
- Research axis : § 6.1.1.
- Partners : Thomson R&d France, Thomson Grass Valley, Orange Labs, Alcatel-Lucent, Irisa/Inria-Rennes, IRCCyN, Polymorph Software, BreizhTech, Artefacto, Bilboquet, France 3 Ouest, TDF.
- Funding : Region.
- Period : Oct.08-Aug.10.

The Futurim@ages project studies coding, distribution and rendering aspects of future television video formats: 3DTV, high-dynamic range videos, and full-HD TV. In this context, TEMICS focuses on compact representations and restitution of multi-view videos. Multi-view videos provide interesting 3D functionalities, such as 3DTV (visualization of 3D videos on auto-stereoscopic screen devices) or Free Viewpoint Video (FFV, i.e. the ability to change the camera point of view while the video is visualized). However, multi-view videos represent a huge amount of redundant data compared with standard videos, hence the need to develop efficient compression algorithms. Stereoscopic or auto-stereoscopic devices display very specific camera viewpoints, which should be generated even if they do not correspond to acquisition viewpoints. Artifacts such as ghosting or bad modelled occlusions must be dealt with to render high quality 3D videos.

In the Futurim@ages project, we also address the problem of depth map retrieval from multi-view and from monocular videos of static scenes. For the latter case, we rebuilt the Structure from Motion part of the 3D video codec developed in the team by integrating the lattest state-of-the-art vision algorithms. As for the multi-view case, we developed a new depth map estimation algorithm which aims at preserving depth discontinuities, which is necessary to render correct virtual views of the scene. In the near future, we plan to compare this software to the Depth Estimation Reference Software (DERS) from the MPEG 3DV group, in terms of virtual views synthesis quality.

7.2.2. RIAM-ESTIVALE

Participant: Caroline Fontaine.

- Title : Secured exchanges for video transfer, in line with legislation and economy
- Research axis : § 6.4.2.
- Partners : LIS (INPG), ADIS (Univ. Paris XI), CERDI (Univ. Parix XI), LSS (Univ. Paris XI/Supelec), Basic-Lead, Nextamp, SACD.
- Funding : ANR.
- Period : 12/12/2005-12/06/2009

ESTIVALE is a project dealing with the diffusion of video on demand in several contexts: from personal use to professionnal use. People involved in the project are from different communities: signal processing and security, economists and jurist. The goal of the project is to design technical solutions for securing this delivery, through DRM and watermarking tools, and to remain consistent with the economical and juridical studies and demands. In 2009, the TEMICS project-team has contributed on the design of a practical efficient fingerprinting scheme for video (see Section 6.4.2).

7.2.3. RIAM-MEDIEVALS

Participants: Caroline Fontaine, Ana Charpentier.

- Title : Watermarking and Visual Encryption for Video and Audio on Demand legal diffusion
- Research axis : § 6.4.2.
- Partners : MEDIALIVE, LSS (Univ. Paris XI/Supelec), GET-INT, THOMSON Software et Technologies Solutions, AMOSSYS SAS.
- Funding : ANR.
- Period : 28/12/2007-28/12/2010

MEDIEVALS is a project dealing with the diffusion of video or audio on demand. MEDIALIVE developed a software to secure this delivery through visual encryption, and the goal of the project is to add watermarking/fingerprinting in the process to improve the security of the delivered content. In 2009, the TEMICS project-team has contributed on the rationale of the new software architecture. We performed tests to study the best articulation between visual encryption, watermarking and anti-collusion codes.

7.2.4. ANR-COHDEQ

Participant: Jean-Jacques Fuchs.

The RNRT project COHDEQ 40 “COHerent DEtection for QPSK 40GHz/s systems” whose coordinator is Alcatel has started in January 2007. It extends over a 3-year period and its aim is to establish the feasibility of coherent detection in optical fibers transmission systems. As far as Irisa is concerned, the work is done by ASPI in collaboration with TEMICS. It covers all the signal processing aspects of this specific digital communication system that will be able to achieve a 100 Gbit/s channel rate.

7.2.5. ANR-TCHATER

Participant: Jean-Jacques Fuchs.

The RNRT project TCHATER “Terminal Coherent Hétérodyne Adaptatif TEmps Réél” whose coordinator is Alcatel has started in January 2008. It will run over a 3-year period and aims to fully implement coherent detection in an optical fibers transmission systems, with among others the real time implementation on dedicated FPGA's that will be taken care off by the Inria-Arenaire team. As far as Irisa is concerned, the work is done by ASPI in collaboration with TEMICS. To adapt the extremely high channel rate, 4 ADC's (analog-to-digital converters) are needed and to accommodate the FPGA's to their output rate, temporal multiplexing of order 40 is required. This structure strongly impacts the signal processing algorithms developed in the COHDEQ project that have to be rewritten and adapted to be able to handle the ensuing constraints.

7.2.6. ANR-STRADE

Participant: Jean-Jacques Fuchs.

The project STRADE “Réseaux du Futur et Services” whose coordinator is Alcatel will start in the fall of 2009. It will run over a 3-year period and aims to investigate the potentialities of optical fibers with higher effective area than those used nowadays. The overall objective is to increase the global transmission capacity of a single fiber. As far as Irisa is concerned, the work is done by ASPI in collaboration with TEMICS and concerns the signal processing aspects. One of the points that will be investigated once the fibers will be available, is the possibility of using mode division multiplexing and fully new signal processing algorithms will then be required to, among others, - improve the separation of the different modes tentatively done by physical devices that need to be developed, - to separate the different polarization and - to achieve the equalization on the different outputs.

7.2.7. ANR-ESSOR

Participants: Christine Guillemot, Aline Roumy, Velotiaray Toto-Zaratosoa.

- Title : Distributed Video Coding
- Research axis : § 6.3.2.
- Partners : CNRS/LSS, ENST-Paris, CNRS/I3S;
- Funding : ANR.
- Period : 01/11/2006-31/10/2009

Compared with predictive coding, distributed video compression holds a number of promises for mobile applications: a more flexible coder/decoder complexity balancing, increased error resilience, and the capability to exploit inter-view correlation, with limited inter-camera communication, in multiview set-ups. However, despite the growing number of research contributions in the past, key questions remain to bring monoview and multi-view DVC to a level of maturity closer to predictive coding: estimating at encoder or decoder the *virtual* correlation channel from unknown - or only partially known - data; finding the best SI at the decoder for data not - or only partially - known. Solutions to the above questions have various implications on coder/decoder complexity balancing, on delay and communication topology, and rate-distortion performance. These questions are being addressed by the ANR-ESSOR project. The TEMICS project-team more specifically contributes on the design of Slepian-Wolf and Wyner-Ziv coding tools as well as on the design of robust and joint source-channel distributed coding strategies. More specifically, in 2009, the TEMICS project-team has pursued the development of a practical distributed video compression schemes based on LDPC codes, improving its performance by designing estimation and decoding algorithms which allow us exploiting both the non-uniformity distribution and the memory of the correlated sources present in the video signal (see Section 6.3.2).

7.2.8. ANR-ICOS-HD

Participants: Christine Guillemot, Fethi Smach, Joaquin Zepeda.

- Title : Scalable Indexing and Compression Scalable for High Definition TV;
- Research axis : § 6.2.6.
- Partners : Université de Bordeaux, CNRS/I3S;
- Funding : ANR.
- Period : 01/01/2007-31/12/2010

The objective of the project is to develop new solutions of scalable description for High Definition video content to facilitate their editing, their acces via heterogeneous infrastructures (terminals, networks). The introduction of HDTV requires adaptations at different levels of the production and delivery chain. The access to the content for editing or delivery requires associating local or global spatio-temporal descritors to the content. The TEMICS project-team contributed in particular on the study of new forms of signal representation amenable to both compression and feature extraction (see Section 6.2.6). A new concept called visual sentences which can be seen as a sparse extension of the concept of visual words has been introduced and assessed for image retrieval in large databases. Methods for computing a similarity measure between descriptors has been designed for the approximate nearest neighbor search task required in content based image retrieval systems. A novel method of dictionary construction has also been developed for this problem of joint compression/description problem (see Section 6.2.6).

7.3. European contracts

7.3.1. FP6-IST NOE NEWCOM++

Participants: Christine Guillemot, Cédric Herzet, Aline Roumy.

- Title: NEWCOM: Network of Excellence in Wireless Communication.
- Research axis: 6.3.1, 6.3.2
- Funding: CEE.
- Period: Jan. 2008 - Dec. 2009.

The NEWCOM++ project proposal (Network of Excellence in Wireless COMmunication) intends to create a trans-European virtual research centre on the topic “The Network of the Future”. It was submitted to Call 1 of the VII Framework Program under the Objective ICT-2007.1.1: The Network of the Future, mainly in its target direction “Ubiquitous network infrastructure and architectures”. We participate in the workpackage WPR7 - Joint source and channel co-decoding which we now coordinate together with the task TR7.3 Tools for multi-terminal JSCC/D. WPR7 addresses issues related to the robust transmission of multimedia, and essentially video, over wireless channels (possibly terminating a wired IP network). Such issues are : (i) solving the compatibility problem with the classical OSI layers separation (to what extent can we keep this separation ?) (ii) providing new tools (and expanding existing ones) for Joint Source and Channel Coding/decoding (JSCC/D) in classical one to one, one to many (broadcast), or distributed contexts (iii) providing new tools for analysing the efficiency of these tools (iv) working on practical, long term situations, which will be used as test-beds.

8. Dissemination

8.1. Awards

+ Gael Sourimant has received the young researcher award of the CORESA committee.

8.2. Invited talks

- C. Guillemot gave a seminar at La Havana, Cuba, in the context of a Franco-Cuban winter school (10-17 January 2009);
- C. Guillemot has given an invited talk and has participated to a panel on “perceptual coding” at PCS (Chicago, May 2009);
- O. Le Meur has co-organized a special session “Visual attention: models and applications in image and video processing” at ICIP, Cairo, Oct. 2009;

- L. Guillo presented the ADT PICOVIN at the INRIA ARC-ADT workshop held in Bordeaux, September 29-30, 2009.

8.3. Leadership within the scientific community

- C. Fontaine is associate editor of the Journal in Computer Virology (Springer-Verlag);
- C. Fontaine was a member of the program committee of the conferences CORESA (Toulouse, France, March), WCC 2009 (Ullensvang, Norway, May), SSTIC 2009 (Rennes, France, June);
- C. Fontaine was a member of the organizing committee of the conference SSTIC 2009 (Rennes, France, June);
- C. Fontaine is a member of the scientific advisory board of the Brittany competence center Diwall;
- J.J. Fuchs is a member of the technical program committees of the following conferences : Eusipco 2009 and Gretsi 2009.
- J.J. Fuchs is a member of the committee that delivers the best thesis price in Signal and Image processing (prix de thèse en Signal-Image du club EEA).
- C. Guillemot is associate editor of the journal IEEE Transactions on Signal Processing (2007-2009).
- C. Guillemot is a member of the external scientific advisory board of the IST Network of Excellence VISNET2;
- C. Guillemot is a member of the Selection and Evaluation Committee of the "Pôle de Compétitivité" Images and Networks of the Region of Ouest of France.
- C. Guillemot is a member of the best paper award committee of the Eurasip Image communication journal.
- C. Guillemot is member of the "Specif-Gilles Kahn Thesis Award" committee.
- C. Guillemot is the coordinator of the ANR ICOS-HD project.
- C. Guillemot chaired the selection committees for the recruitment of two assistant professors in image at the University of Rennes 1.
- C. Labit has served as reviewer for the technical program committees of: Int. Conf of Image Processing, ICASSP, Eusipco.
- C. Labit is member of the GRETSI association board.
- C. Labit is, for the national INRIA's research department, scientific adviser of INRIA-SUPCOR (Support services for ANR collaborative research initiatives.
- C. Labit is the Scientific Board chairman of Rennes1 University (since June 1st, 2008).
- C. Labit is president of Rennes-Atalante Science Park.
- A. Roumy was a member of the technical program committee of WiOpt 2009 (Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks).

8.4. Exhibitions and demonstrations

- The TEMICS project-team presented a demo at the exhibition held during the SPIE conference "Electronic Imaging Media Forensics and Security XI" at San Jose, USA, Jan. 2009.
- The TEMICS project-team presented a demo during the visit of M. Friedel Orange Labs Research Director and during the visit of Jeajoon Lee from Samsung Electronics.

8.5. Teaching

- Enic, Villeneuve-d'Ascq, (C. Guillemot: Video communication) ;

- Esat, Rennes, (C. Guillemot: Image and video compression) ;
- Engineer degree Diic-INC, Ifsic-Spm, university of Rennes 1 (O. Le Meur, L. Guillo : image processing, compression, communication);
- Engineer degree Diic-LSI, Ifsic-Spm, university of Rennes 1 (L. Guillo, O. Le Meur : compression, video streaming);
- Engineer degree ESIR, Université de Rennes 1: Jean Jacques Fuchs teaches several courses on basic signal processing and control ;
- Master Research 2: SISEA: Jean Jacques Fuchs teaches a course on optimization and sparse representations ; he also intervenes in the Joint International Program of the University of Rennes 1 and the SouthEast University of China (Nanjing) and teaches a course in Advanced Signal Processing in the International Master of Science in Electronics and Telecommunications.
- Master Research-2 SISEA: C. Guillemot and C. Labit teach a course on image and video compression ;
- Master Research-2 Computer Science: C. Fontaine is in charge of the "Information and Computing Infrastructure Security" track, and teaches a course on information hiding ;
- Master, Security of Information Systems, Supelec-ENSTB (C. Fontaine: information hiding) ;
- Master, Network Engineering, university of Rennes I (L. Guillo, Video streaming) ;
- DRT of Rennes1 university: C. Labit supervises a DRT cursus (K. Torres) with Thomson Grass Valley addressing the problem of rate-distortion control for video MPEG-like codecs.

9. Bibliography

Major publications by the team in recent years

[1] G. CAIRE, S. GUEMGHAR, A. ROUMY, S. VERDU. *Maximizing the spectral efficiency of coded CDMA under successive decoding*, in "IEEE Trans. on Information Theory", Jan. 2004, p. 152 - 164.

[2] F. CAYRE, C. FONTAINE, T. FURON. *Watermarking Security: Theory and Practice*, in "IEEE Transactions on Signal Processing", vol. 53, n^o^ 10, 2005, p. 3976–3987, numéro spécial "Supplement on Secure Media III".

[3] V. CHAPPELIER, C. GUILLEMOT. *Oriented wavelet transform for image compression and denoising*, in "IEEE Transactions on Image Processing", vol. 15, n^o^ 10, Oct. 2006, p. 2892-2903.

[4] J.-J. FUCHS. *On the application of the global matched filter to DOA estimation with uniform circular arrays.*, in "IEEE Trans. on Signal Processing", n^o^ 49, Apr. 2001, p. 702-709.

[5] T. FURON, P. DUHAMEL. *An Asymmetric Watermarking Method*, in "IEEE Trans. on Signal Processing", vol. 51, n^o^ 4, April 2003, p. 981-995, Special Issue on Signal Processing for Data Hiding in Digital Media and Secure Content Delivery.

[6] T. GUIONNET, C. GUILLEMOT. *Soft decoding and synchronization of arithmetic codes: Application to image transmission over noisy channels*, in "IEEE Transactions on Image Processing", vol. 12, Dec. 2003, p. 1599-1609.

[7] H. JÉGOU, C. GUILLEMOT. *Robust multiplexed codes for compression of heterogeneous data*, in "IEEE Trans. on Information Theory", vol. 51, n^o^ 4, Apr. 2005, p. 1393-1403.

[8] L. OISEL, E. MÉMIN, L. MORIN, F. GALPIN. *One-dimensional dense disparity estimation for three-dimensional reconstruction*, in "IEEE Transactions on Image Processing", vol. 12, n^{o} 9, Sept. 2003, p. 1107–1119.

[9] S. PATEUX, G. LE GUELVOUIT. *Practical watermarking scheme based on wide spread spectrum and game theory*, in "Signal Processing : Image Communication", n^{o} 18, Apr. 2003, p. 283–296.

[10] G. RATH, C. GUILLEMOT. *Subspace-Based Error and Erasure Correction with DFT Codes for Wireless Channels*, in "IEEE Trans on Signal Processing", vol. 52, n^{o} 11, Nov. 2004, p. 3241-3252.

Year Publications

Articles in International Peer-Reviewed Journal

[11] Z. CHEN, C. GUILLEMOT. *Perceptually-friendly H.264/AVC Video Coding based on Foveated Just-Noticeable-Distortion Model*, in "IEEE Trans. on Circuits and Systems for Video Technology", 2009, accepted CN .

[12] O. CRAVE, C. GUILLEMOT, B. PESQUET-POPESCU. *Multiple scalar quantization with side information: Practical scheme and iterative decoding*, in "EURASIP Journal on Advances in Signal Processing, JASP", vol. 2009, Jan. 2009, Article ID 170924, 12 pages.

[13] O. CRAVE, B. PESQUET-POPESCU, C. GUILLEMOT. *Robust video coding based on multiple description scalar quantization with side information*, in "IEEE Trans. on Circuits and Systems for Video Technology", 2009, accepted.

[14] S. MALINOWSKI, X. ARTIGAS, C. GUILLEMOT, L. TORRES. *Punctured Quasi-Arithmetic Codes for Distributed Source Coding", IEEE Trans on Signal Processing*, in "IEEE Trans. on Signal Processing", 2009, to appear ES .

[15] S. MALINOWSKI, H. JÉGOU, C. GUILLEMOT. *Posterior marginals computation on aggregated state models for soft source decoding*, in "IEEE Trans. on communication", vol. 57, Apr. 2009, p. 888-892.

[16] N. SELLAMI, M. SIALA, A. ROUMY, I. KAMMOUN. *MAP sequence equalization for imperfect frequency selective channel knowledge*, in "ETT European Transactions on Telecommunications", 2009, accepted.

International Peer-Reviewed Conference/Proceedings

[17] A. CHARPENTIER, F. XIE, C. FONTAINE, T. FURON. *Expectation Maximisation decoding of Tardos probabilistic fingerprinting code*, in "IS&T/SPIE International Symposium on Electronic Imaging 2009", Proceedings of the SPIE, vol. 7254, SPIE, 2009, Media Forensics and Security XI CN .

[18] Z. CHEN, C. GUILLEMOT. *Perception-oriented video coding based on a foveated JND model*, in "Picture Coding Symposium, PCS, Chicago", 2009 CN .

[19] Z. CHEN, C. GUILLEMOT. *Perceptually-friendly H.264/AVC video coding*, in "IEEE Intl. Conf. On Image Processing, ICIP, Cairo", 2009 CN .

[20] T. COLLEU, L. MORIN, C. LABIT, S. PATEUX, R. BALTER. *Compact quad-based representation for 3D video*, in "3DTV International Conference", Apr. 2009.

[21] O. CRAVE, C. GUILLEMOT, B. PESQUET-POPESCU. *Multiple description video coding and iterative decoding of LDPCA codes with side information*, in "IEEE Intl. Conference on Acoustic, Speech and Signal Processing, ICASSP, Taipei", 2009.

[22] J. FUCHS. *DOA estimation in the presence of modeling errors, the Global Matched Filter approach*, in "European Conf. On Signal Processing, EUSIPCO, Glasgow, UK", aug. 2009, p. 1963-1967.

[23] J. FUCHS. *Matched detector in the presence of interference subspace uncertainty*, in "Asilomar Conf. on Signals, Systems and Computers, Pacific Grove, CA", Nov. 2009.

[24] J. J. FUCHS. *Fast implementation of a $\ell_1 - \ell_1$ regularized sparse representation algorithm.*, in "Proc.of ICASSP, IEEE Int. Conference on Acoustics, Speech and Signal Processing, Taipei", apr 2009, p. 3329-3332.

[25] J. J. FUCHS. *Identification of real sinusoids in noise, the Global Matched Filter approach.*, in "15th Ifac-Ifors symposium on Identification and system Parameter Estimation, Saint-Malo, France", jul 2009, p. 1127-1132.

[26] M. GRANGETTO, C. GUILLEMOT, E. MAGLI, S. MALINOWSKI, G. OLMO. *On the performance of arithmetic and quasi-arithmetic codes for joint source-channel coding*, in "Newcom++ workshop, Barcelona", 2009 IT .

[27] C. HERZET, V. TOTO-ZARASOA, A. ROUMY. *Error-resilient non-asymmetric Slepian-Wolf coding*, in "IEEE Intl. Conf. on Communications, ICC", 2009 MG .

[28] V. JANTET, L. MORIN, C. GUILLEMOT. *Incremental-LDI for Multi-View Coding*, in "3DTV International Conf.", Apr. 2009.

[29] I. H. KACEM, N. SELLAMI, I. FIJALKOW, A. ROUMY. *Joint training interval length and power allocation optimization for MIMO flat fading channels*, in "IEEE International Symposium on Wireless Communication Systems 2009 (ISWCS'09), Sienna, Italy", September 2009.

[30] I. H. KACEM, N. SELLAMI, I. FIJALKOW, A. ROUMY. *Training interval length optimization for MIMO flat fading channels using decision-directed channel estimation*, in "IEEE International Conference on Wireless and Mobile Computing (WiMob'09), Morocco", October 2009.

[31] I. KAMMOUN, M. SIALA, N. SELLAMI, A. ROUMY. *Symbol MAP Detection Aided with First Channel Estimate for TDMA Systems Over a Frequency Selective Channel*, in "IEEE International Symposium on Wireless Communication Systems 2009 (ISWCS'09), Sienna, Italy", September 2009.

[32] S. MALINOWSKI, X. ARTIGAS, C. GUILLEMOT, L. TORRES. *Distributed coding using punctured quasi-arithmetic codes for memory and memoryless sources*, in "Picture Coding Symposium, PCS, Chicago", 2009 ES .

[33] O. L. MEUR, P. L. CALLET. *What we see is most likely to be what matters: visual attention and applications*, in "IEEE Intl. Conf. On Image Processing, ICIP, Cairo", 2009.

[34] G. RATH, C. GUILLEMOT. *Sparse approximation with an orthogonal complementary matching pursuit algorithm*, in "IEEE Intl. Conference on Acoustic, Speech and Signal Processing, ICASSP, Taipei", 2009 IN .

[35] G. RATH. *A Comparative Study Of Some Greedy Pursuit Algorithms for Sparse Approximation*, in "European Conf. On Signal Processing, EUSIPCO, Glasgow", 2009 IN .

[36] V. TOTO-ZARASOA, E. MAGLI, A. ROUMY, G. OLMO. *On Distributed Arithmetic Codes and Syndrome Based Turbo Codes for Slepian-Wolf Coding of Non Uniform Sources*, in "European Conf. on Signal Processing, EUSIPCO", 2009 MG .

[37] V. TOTO-ZARASOA, A. ROUMY, C. GUILLEMOT, C. HERZET. *Robust and Fast Non Asymmetric Distributed Source Coding Using Turbo Codes on the Syndrome Trellis*, in "IEEE Intl. Conf. on Acoustic, Speech and Signal Processing, ICASSP, Taipei", 2009 MG .

[38] M. TURKAN, C. GUILLEMOT. *Sparse Approximation With Adaptive Dictionary For Image Prediction*, in "IEEE Intl. Conf. On Image Processing, ICIP, Cairo", 2009 TR .

[39] J. ZEPEDA, E. KIJAK, C. GUILLEMOT. *SIFT-Based Local Image Description Using Sparse Representations*, in "IEEE workshop on Multimedia Signal Processing (MMSP), Rio de Janeiro", 2009 NI .

National Peer-Reviewed Conference/Proceedings

[40] A. CHARPENTIER, C. FONTAINE, T. FURON. *Décodage EM du code de Tardos pour le fingerprinting*, in "Proc. XXIIème colloque du GRETSI", 2009.

[41] T. COLLEU, L. MORIN, C. LABIT, S. PATEUX, R. BALTER. *Représentation compacte à base de quadrilatères pour la vidéo 3D*, in "ORASIS'09 - Congrès des jeunes chercheurs en vision par ordinateur", June 2009.

[42] J. FUCHS. *Des moindres carrés aux moindres déviations*, in "Actes du 21ème colloque GRETSI, Dijon, France", Sept. 2009.

[43] J. GAUTHIER, T. MAUGEY, B. PESQUET-POPESCU, C. GUILLEMOT. *Amélioration du modèle statistique de bruit pour le codage vidéo distribué*, in "Actes du 21ème colloque GRETSI, Dijon, France", Sept. 2009.

[44] I. H. KACEM, N. SELLAMI, I. FIJALKOW, A. ROUMY. *Optimisation conjointe du nombre des symboles pilotes et de l'allocation des puissances pour les canaux MIMO non selectifs en frequence*, in "22th Symposium on Signal and Image Processing, Gretsi, Dijon, France", September 2009.

[45] G. SOURIMANT, T. COLLEU, V. JANTET, L. MORIN. *Recalage GPS / SIG / Video, et synthèse de textures de bâtiments*, in "COmpression et REpresentation des Signaux Visuels (CORESA 2009), Toulouse, France", March 2009.

[46] F. XIE, C. FONTAINE, T. FURON. *Un schéma complet de traçage de documents multimédia reposant sur des versions améliorées des codes de Tardos et de la technique de tatouage 'Broken Arrows'*, in "Proc. XXIIème colloque du GRETSI", 2009 CN .