



INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Project-Team TEMICS

*Traitement, Modélisation d'Images et
CommunicationS*

Rennes

THEME COG

Activity
R
Report

2006

Table of contents

1. Team	1
2. Overall Objectives	1
2.1. Overall Objectives	1
3. Scientific Foundations	3
3.1. 3D scene modelling based on projective geometry	3
3.2. Frame expansions	4
3.3. Rate-distortion theory	5
3.4. Distributed source coding and multiple access channel	5
3.4.1. Multiple access channel	5
3.5. Watermarking as a problem of communication with side information	6
4. Application Domains	7
4.1. Introduction	7
4.2. Compression with advanced functionalities	8
4.3. Multimedia communication	8
4.4. Copy protection, copyright enforcement and enriched content	8
5. Software	8
5.1. Video communication platform	8
5.2. WAVIX: Wavelet based video codec	9
5.3. 3D Model-based video codec and player	9
6. New Results	10
6.1. Analysis and modelling of video sequences	10
6.1.1. 3d modelling of urban environments from video, GPS and GIS data	10
6.1.2. 3d scene modelling for distributed video compression	10
6.1.3. Motion modelling and tracking for distributed video compression	12
6.2. Sparse signal representation and compression	14
6.2.1. Sparse signal representation	14
6.2.2. Compressed representations of the Laplacian pyramid	16
6.2.3. Multiresolution representations for compression and local description	17
6.2.4. Distributed video compression	17
6.3. Joint source-channel coding	18
6.3.1. Frames as joint source-channel codes for MIMO channels	18
6.3.2. Multiple description coding	19
6.3.3. Error-resilient source codes and soft decoding	19
6.3.4. Parameter estimation in convolutive and instantaneous mixtures	20
6.3.5. Robust estimation and detection	20
6.4. Distributed source-channel coding	21
6.4.1. Correlated sources over multiple access channels	21
6.4.2. Distributed coding for sensor networks	21
6.5. Image and video watermarking and security	22
6.5.1. Optimal watermark detection	22
6.5.2. Watermarking Security	22
6.5.3. Fingerprinting	23
7. Contracts and Grants with Industry	23
7.1. Industrial contracts	23
7.1.1. CRE with France Telecom on 3D reconstruction of urban scenes	23
7.1.2. CRE with France Telecom on distributed video compression	24
7.1.3. CIFRE contract with Thomson on the problem of spectral deconvolution for video compression and protection	24
7.2. National contracts	24

7.2.1. RIAM-COPARO	24
7.2.2. RNRT-COSINUS	25
7.2.3. RNRT de l'ANR	25
7.2.4. RIAM-ESTIVALE	25
7.3. European contracts	26
7.3.1. FP6-IST NoE SIMILAR	26
7.3.2. FP6-IST NOE NEWCOM	26
7.3.3. FP6-IST STREP DANAE	27
7.3.4. FP6-IST STREP DISCOVER	27
8. Other Grants and Activities	28
8.1. National initiatives	28
8.1.1. ACI Fabriano	28
8.1.2. ACI Masse de données - Codage	28
8.2. International initiatives	29
9. Dissemination	29
9.1. Standardization	29
9.2. Invited talks	29
9.3. Leadership within the scientific community	29
9.4. Awards	30
9.5. Invitations and visits	30
9.6. Teaching	30
10. Bibliography	31

1. Team

Head of project-team

Christine Guillemot [DR Inria, HdR]

Administrative assistant

Huguette Béchu [TR Inria, shared with VISTA and ASPI]

Inria research scientists

Teddy Furon [CR]

Gagan Rath [CR]

Aline Roumy [CR]

Cnrs research scientist

Caroline Fontaine [CR]

Faculty members

Jean-Jacques Fuchs [Professor, Ifsic, University of Rennes 1, HdR]

Luce Morin [Associate professor, Ifsic, University of Rennes 1, HdR]

CNRS technical staff

Laurent Guillo [IR]

Inria projects technical staff

Florelle Pauchet [RIAM-COPARO project, until June 30, 2006]

Adrien Schadle [RNRT-COSINUS project, until Aug., 2006]

Christel Chamaret [RNRT-COSINUS project]

Inria junior technical staff

Jérôme Allesia [since Sept. 1st, 2006]

Inria post-doctoral fellows

Fabien Galand [RIAM-ESTIVALE project, since Sept. 1st, 2006]

Jayanth Nayak [until Sept. 15, 2006]

Wenxian Yang [RNRT-COPARO Project, until June 30, 2006]

Yang Liu [since Dec. 15 2006]

Ph.D Students

Robin Chatterjee [Inria grant, IST-DANAE project]

Olivier Crave [also with ENST-PARIS, ENST grant]

Denis Kubasov [Inria grant, IST-DISCOVER project]

Khaled Lajnef [Inria grant, CRE France Télécom]

Matthieu Maître [PhD student at Urbana Champaign, ACET from May to Aug. 2006]

Simon Malinowski [Ministry of research grant]

Sébastien Maria [Ministry of research grant, until Sept. 30, 2006]

Aurélie Martin [CIFRE contract with Thomson, from Oct. 1st, 2006]

Gael Sourimant [Inria grant, CRE France Télécom]

François Tonnin [CNRS - Brittany Region grant, also with Texmex, until May 30th, 2006]

2. Overall Objectives

2.1. Overall Objectives

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

- *Analysis and modelling of video sequences.* The support of advanced interaction functionalities such as video content manipulation, or navigation requires the development of video analysis and modelling algorithms. TEMICS focuses on the design of solutions for segmenting video objects and for extracting and coding their main attributes (shape, motion, illumination, ...). In order to support navigation within video scenes, the ability to construct a 3D model of the scene is a key issue.

One specific problem addressed is the design of algorithms for 3D modelling from monocular video sequences with optimum tradeoff between model reliability and description cost (rate). Finally, the optimal support of the above functionalities in networked multimedia applications requires scalable, compact and transmission noise resilient representations of the models and of their attributes, making use of joint source-channel coding principles (see below).

- *Compression, scalable coding and distributed source coding.*

Scalable video compression is essential to allow for optimal adaptation of compressed video streams to varying network characteristics (e.g. to bandwidth variations) in various applications (e.g. in unicast streaming applications with pre-encoded streams, and in multicast applications). Frame expansions and in particular 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. 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, light-weight video compression systems (extension of MMS-based still image transmission to video) 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 theorem established in 1973. Even though first theoretical foundations date back to early 70's, it is only recently that concrete solutions, motivated by the above applications, aiming at approaching the theoretic performance bounds have been introduced.

- *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, stating that source and channel optimum performance bounds can be approached as close as desired by designing independently source and channel coding strategies, 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 a theoretical and practical framework setting the foundations for optimal design of image and video transmission systems 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 source 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 in order to address the tradeoff performance-complexity.

- *Distributed joint source-channel coding.* Distributed joint source-channel coding refers to the problem of sending correlated sources over a common noisy channel without communication between the senders. Note that cooperation among channel and source encoding of one sender is allowed but not between different senders. This problem occurs mostly in network, where the communication between the nodes is not possible or not desired due to its high energy cost (network

video camera, sensor network...). A major difference with the joint source-channel case is that the separation between source and channel coding does not always hold. This depends on the setup. If asymmetric source encoding is performed (one source can be recovered perfectly from the data sent by this source only), then separation holds. Otherwise it depends on the channel. For independent channels, source channel separation holds but for interfering channels, counterexamples can be found where joint source-channel scheme (but still distributed) performs better than the separated scheme. In this area, we design distributed source-channel schemes. The source channel encoder of each sender can either be joint or disjoint depending on the context (since separation holds in some specific cases only).

- *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, and steganography. However, data hiding technology can also be used for enhancing a signal by embedding some meta-data. The data hiding 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. Some 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. Our developments rely on scientific foundations in the areas of signal processing and information theory, such as communication with side information at the transmitter.

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 (IETF, 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 (ACI FABRIANO, ACI CODAGE, RNRT COSINUS, RIAM COPARO, RIAM ESTIVALE) and European projects (IST-DANAE, IST-SIMILAR, IST-DISCOVER and 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

Keywords: *3D reconstruction, camera models, computer vision, epipolar constraints, fundamental matrix, perspective projection, 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 C_i is modelled as a *perspective projection*, the *projection equations* are :

$$\tilde{p}_i = P_i \tilde{x}, \quad (1)$$

where \tilde{x} is a 3D point with homogeneous coordinates $\tilde{x} = (x \ y \ z \ 1)^t$ in the scene reference frame \mathcal{R}_0 , and where $\tilde{p}_i = (X_i \ Y_i \ 1)^t$ 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 \tilde{x} , given \tilde{p}_i , i.e. to solve Eqn. (1) with respect to coordinates \tilde{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

Keywords: *Wavelet transforms, error correcting codes, filter banks, multiple description coding, oversampled 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 $\Phi_F \equiv \{\phi_i\}_{i \in I}$, ϕ_i 's are the frame vectors and I is the index set. The i th frame expansion coefficient of \mathbf{x} is defined as $(F\mathbf{x})_i \equiv \langle \phi_i, \mathbf{x} \rangle$, 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 $\tilde{\Phi}_F \equiv \{(F^h F)^{-1} \phi_i\}_{i \in I}$. 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

Keywords: *OPTA limit (Optimum Performance Theoretically Attainable), Rate allocation, Rate-Distortion optimization, channel modelization, error correcting codes, joint source-channel coding multiple description coding, lossy coding, oversampled frame expansions.*

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

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

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 and multiple access channel

Keywords: *Slepian-Wolf coding, coding with side information, correlated sources, multiple access channels.*

Distributed source coding (DSC) is the separate compression of (many) correlated sources (separate in the sense that no cooperation/communication between the sources is allowed). The source coding theorem establishes that to reliably compress a source X , the compression rate R_x (in bits per source symbol) must be greater than the entropy of the source $H(X)$. Therefore to jointly compress the sources X and Y , the sum rate $R_x + R_y$ must satisfy $R_x + R_y > H(X, Y)$. On the other hand, if no cooperation is allowed between the encoders, it is clear that a sum rate of $H(X) + H(Y)$ is enough. A very surprising result due to Slepian and Wolf shows that the sufficient rate for separate encoding (but joint decoding) of two correlated sources is the joint entropy $H(X, Y)$. Other saying adding the constraint of encoding separation does not incur any loss in terms of compression rate.

3.4.1. Multiple access channel

The multiple access channel (MAC) models the case where many independent senders transmit data over a common noisy channel. The MAC can therefore be seen as a generalization of the one sender case, whose properties are studied in the channel coding theorem. This theorem a.k.a. second Shannon's theorem considers the transmission of one sender $X \sim p(x)$ over a noisy channel with output Y and transition probability $p(y|x)$. This theorem shows that the maximal achievable rate (in bits per channel use) of the sender is the capacity of the channel defined as:

$$\max_{p(x)} I(X; Y)$$

where $I(X;Y)$ is the mutual information (MI) between the r.v. X and Y . The mutual information is a function of the input density $p(x)$ and of the channel transition probability $p(y|x)$. It measures the amount of information shared by the 2 r.v. X and Y . The set of achievable rates can therefore be rewritten as:

$$\bigcup_{X \sim p(x)} \{R \text{ s.t. } R \leq I(X;Y)\}$$

The 2-user MAC models the transmission of 2 independent senders $(X_1, X_2) \sim p(x_1)p(x_2)$ over the same channel with output Y and transition probability $p(y|x_1, x_2)$. The transmission rates are denoted R_1 and R_2 . The capacity region is the closure of the set of achievable (R_1, R_2) rate pairs and it is shown to be the closure of the following set:

$$\bigcup_{(X_1, X_2) \sim p(x_1)p(x_2)} \left\{ (R_1, R_2) \text{ s.t. } \begin{array}{l} R_1 \leq I(X_1; Y|X_2) \\ R_2 \leq I(X_2; Y|X_1) \\ R_1 + R_2 \leq I(X_1, X_2; Y) \end{array} \right\}$$

This result can be extended to any number of senders. Closed form expression of the capacity region exists in different examples, for instance the Gaussian MAC.

3.5. Watermarking as a problem of communication with side information

Keywords: *capacity, discrimination, information theory, side information, watermarking.*

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. Let $E(\cdot)$ be the extraction function going from the content space \mathcal{C} to the components space, isomorphic to \mathbf{R}^N

$$\begin{array}{l} E(\cdot) : \mathcal{C} \mapsto \mathbf{R}^N \\ C \rightarrow \mathbf{V} = E(C) \end{array}$$

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, also sketched in Fig. 1, 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 $S(\cdot)$ mapping a discrete message $m \in \mathcal{M}$ onto a signal in \mathbf{R}^N :

$$\begin{array}{l} S(\cdot) : \mathcal{M} \mapsto \mathbf{R}^N \\ m \rightarrow \mathbf{S} = S(m) \end{array}$$

The decoding function $S^{-1}(\cdot)$ 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 $\{S(\cdot), S^{-1}(\cdot)\}$ 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, as sketched in Fig. 2, 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.

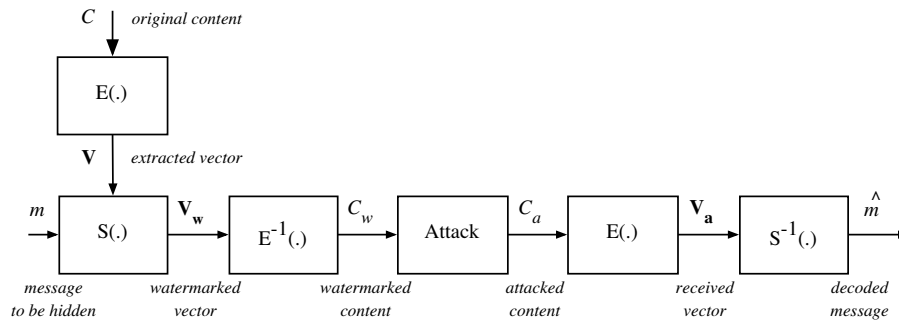


Figure 1. Classical watermarking scheme

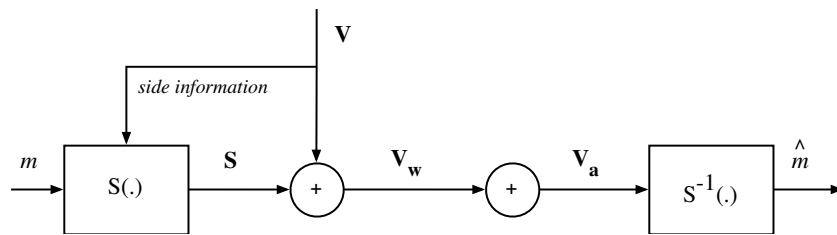


Figure 2. Watermarking as a problem of communication with side information.

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 video compression, network adaptation (e.g., resilience to channel noise), or in terms of advanced functionalities such as navigation, content copy and copyright protection, or authentication.

4.2. Compression with advanced functionalities

Notwithstanding the already large number of solutions, compression remains a widely-sought capability especially for audiovisual communications over wired or wireless IP networks, often characterized by limited bandwidth. The advent of these delivery infrastructures has given momentum to extensive work aiming at optimized end-to-end QoS (Quality of Service). This encompasses low rate compression capability but also capability for adapting the compressed streams to varying network conditions. Scalable coding solutions making use of mesh-representations and/or spatio-temporal frame expansions are developed for that purpose. At the same time, emerging interactive audiovisual applications show a growing interest for 3-D scene navigation, for creating intermediate camera viewpoints, for integrating information of different nature, (e.g. in augmented and virtual reality applications). Interaction and navigation within the video content requires extracting appropriate models, such as regions, objects, 3-D models, mosaics, shots... The signal representation space used for compression should also be preferably amenable to signal feature and descriptor extraction for fast and easy data base access purposes.

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 paradigm sets the foundations for the design of efficient solutions to the above challenges. Distributed source coding is driven by a set of emerging applications such as wireless video (e.g. mobile cameras) and sensor networks. Such applications are indeed placing additional constraints on compression solutions, such as limited power consumption due to limited handheld battery power. Distributed source coding is a radical departure from the conventional compression paradigm. The source statistics being exploited in the decoder, the traditional balance of complex encoder and simple decoder is reversed.

4.4. Copy protection, copyright enforcement and enriched content

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 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 content enhancement application, meta-data are embedded in the host signal to prevent their unintentional removal when submitted to transformations. The content becomes self-contained, the created meta-data transmission channel traveling with the content itself. The embedded data must be non perceptible, and possibly robust to a very limited number of classical content processing (e.g., compression, transcoding, postproduction treatments). This application requires a high embedding rate, but no security is needed. Other potential applications are copyright enforcement, authentication, tracing, fingerprinting, and steganography.

5. Software

5.1. Video communication platform

Participants: Christel Chamaret, Laurent Guillo [contact person], Adrien Schaddle.

With the support of the contract RNRT-COSINUS, TEMICS pursues the development of a video communication platform. 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 "Criqs":

- The video streaming server is able to take into account information from the receiver about the perceived quality and also from the link layer. 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. These mechanisms are being assessed in collaboration with Thales Communication which has provided some unequal error protection techniques (UEP).
- The streaming client, "Criqs", can execute 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 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 RFC 3550. These two libraries are respectively registered 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.

5.2. WAVIX: Wavelet based video codec

Participants: Christel Chamaret, Laurent Guillo [contact person].

WAVIX (Wavelet based Video Coder with Scalability) is a low rate fine grain scalable video codec based on a motion compensated t+2D wavelet analysis. Wavix supports three forms of scalability: temporal via motion-compensated temporal wavelet transforms, spatial scalability enabled by a spatial wavelet transforms and SNR scalability enabled by a bit-plane encoding technique. The produced bitstream embeds the different levels of temporal and spatial resolutions as well as quality. 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. In 2006, new techniques have been incorporated in the codec in order to increase its rate-distortion performance as well as its resilience to packet losses. A new motion estimator improving the quality of motion fields and the reliability of Motion Compensated Temporal Filtering (MCTF) has been implemented. A technique for bit budget repartition between motion fields and texture adapted to the frame characteristics (scene with camera motion...) has been investigated. This has allowed improving the PSNR values as well as the visual quality especially of the reconstructed signal, especially at low bit rates. The packet loss problem led to the implementation of a redundant temporal signal expansion. Tests with the video streaming platform have shown a performance gain over a large range of loss rates (from 5% to 20%). The Wavix codec is used in the RNRT COSINUS project, as well as in the context of a collaboration with Thalès, to demonstrate the control of perceived quality while the nature and the quality of wireless network (UMTS) vary.

5.3. 3D Model-based video codec and player

Participants: Jérôme Allasia, Laurent Guillo, Luce Morin [contact person].

The Temics team has in the past years developed a software for 3D modelling of video sequences which allows interactive navigation and view point modification during visualization on a terminal. From a video sequence of a static scene viewed by a monocular moving camera, this software 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. The encoder supports scalability for both geometry and texture information. The codec is being integrated in the TEMICS video communication platform described above. The 3D sequence will then be accessed from a distant terminal and streamed by the server which will dynamically adapt the transmitted bitstream to both network and terminal capabilities.

6. New Results

6.1. Analysis and modelling of video sequences

Keywords: *2D and 3D meshes, 3D modelling, motion modelling, multimodal data fusion, projective geometry, registration, tracking, triangulation.*

6.1.1. 3d modelling of urban environments from video, GPS and GIS data

Participants: Luce Morin, Gael Sourimant.

This work is done in collaboration with the SIAMES project-team (Kadi Bouatouch). From a video sequence of a static scene viewed by a monocular moving camera, we have studied methods to automatically construct a representation of a video as a stream of textured 3D models. 3D models are extracted using stereovision and dense matching maps estimation techniques. However, this approach presents some limitations, in particular the presence of drift in the location and orientation of the 3D models. Drift is due to accumulation of uncertainties in the 3D estimation process along time. This is a strong limitation for virtual reality applications such as insertion of synthetic objects in natural environments. In addition, the model is limited to the areas captured by the video camera. In the case of urban environments, GIS (Geographic Information Systems) provide a complete, geo-referenced modelling of city buildings. However, they are far less realistic than video captures, due to artificial textures and lack of geometric details. Video and GIS data thus give complementary information: video provides photorealism, geometrical details, precision in the fronto-parallel axes; GIS provides a "clean" and complete geometry of the scene, structured into individual buildings. We have also added a GPS acquisition synchronized with video acquisition. In order to combine these three types of data, the first step is to register the data in the same coordinate system. This is the point we have focused on in 2006. The registration step is indeed a bottleneck in the whole modeling system, as it requires geometric correspondences between non similar data contents. GPS data only provide a rough approximation of the camera position (but not its orientation) with regards to the GIS database, so GIS/Video registration can be seen as a two-steps procedure. First, a semi-interactive registration procedure is applied on the first video frame using Dementhon's pose algorithm, given the rough initial estimate provided by GPS data. In a second step, the pose is tracked for each frame, using a visual servoing approach based on the registration of 2D interest points extracted from the images, and 3D points that correspond to these feature points projection onto the building model provided by the GIS database. As a result, we get the pose of the camera, e.g. its position and orientation with regards to the GIS database, for each frame of the video sequence. A robust estimation scheme allows to cope with matching and tracking errors. Registration results obtained on real video sequence of several hundred frames showed a very good back-projection of GIS onto video frames, stable camera path (no jitter) and marginal drift. Using this computed camera pose, the next step will be to address 3D models geometry refinement and extraction of the buildings texture from the video.

6.1.2. 3d scene modelling for distributed video compression

Participants: Christine Guillemot, Matthieu Maître, Luce Morin.



(a) Before

(b) During

(c) After

Figure 3. Successive steps of the initial Gis-Video registration. 3A rough relative position is provided by Gps. 3Automatically extracted building corners are manually registered in the image. 3Once correspondances are given, pose is computed automatically

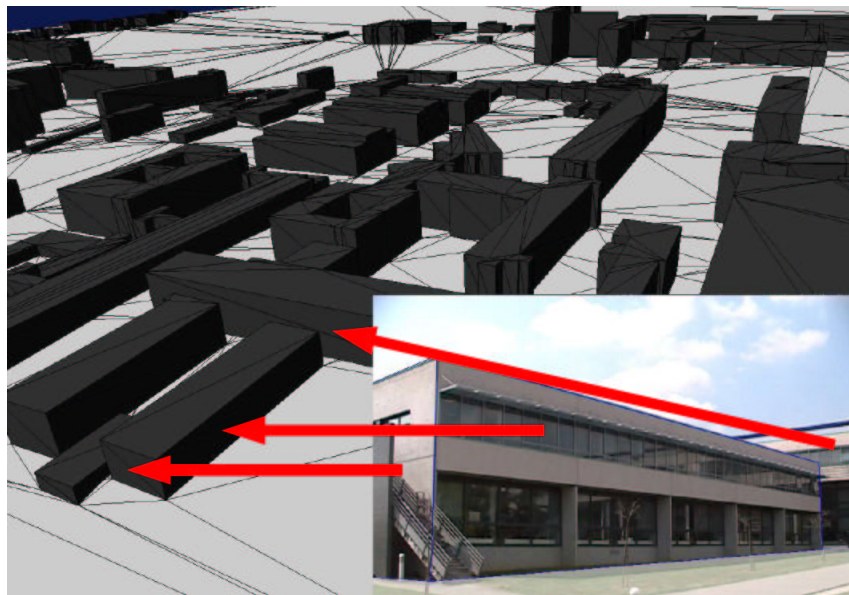


Figure 4. A perspective vrml view of the Gis database. The red arrows show the correspondance between video building facades and their corresponding 3D model

Distributed Video Coding (DVC) has emerged as an alternative to classical motion-compensated codecs (e.g. MPEGx/H.26x). It offers improved robustness, scalability and low complexity at the coder. A typical application is video recording using mobile cameras with transmission over wireless networks (e.g. cell-phone cameras). A drawback of such an algorithm is its reduced rate-distortion performances. This becomes clear by looking at the outline of the algorithm: a video sequence is split into a series of Group of Frames (GOF) whose first frames called “keyframes” are coded using intra coding (e.g. JPEG) and whose other frames are reconstructed at the decoder by first applying block-based motion compensation (BB-MC) between consecutive keyframes and correcting this prediction using some information called “parity bits” coming from the coder. BB-MC being a crude motion model, the keyframe frequency must be very high (usually every other frame) to maintain an acceptable PSNR, causing a major performance hit. We have focused on the design of a new motion model based on 3D mesh modeling. Assuming that the scene being recorded is static, the geometry called “epipolar geometry” can be estimated by detecting and matching corners on pairs or triplets of keyframes. Given any point in a keyframe and this geometry, the corresponding point in the other keyframe is known to lie over a line, thus reducing the search for correspondences from 2D to 1D. The scene is then modeled by a mesh whose control point depths are estimated from the matched corners. This mesh defines a motion field over the intermediate frames, which allows their prediction from the keyframes. Experimental results have shown that such a 3D-DVC scheme allows keyframes to be separated by more than 10 frames. Moreover, the above results led us to consider techniques beyond strict DSC, which would keep its valuable characteristics. We have proposed two different schemes using interest points extracted and tracked for each intermediate frame. The first scheme (*3D-DVC-TD*) extracts points at the encoder and performs tracking at the decoder. Such a scheme enables to greatly improve side-information estimation by reducing misalignments (see figure 6). In the second scheme (*3D-qDVC-TE*) points are tracked between consecutive frames at the encoder thus introducing a limited temporal dependency between a keyframe and its following frames in the GOF. With this “quasi-DVC” scheme, a better estimation of camera pose for intermediate frames provides better SI prediction and rate-distortion performances. It also enables to adapt the keyframe rate to video contents. The performance of the three proposed schemes has been assessed by comparing their rate-distortion performances with the standard H.264 codec (in Intra and IPPP modes), and with state-of-the art 2D-DVC codecs, on video sequences of static scenes. As shown on figure 5, 3D-DVC-TD and 3D-qDVC-TE outperform H.264 Intra and classical 2D-DVC [20].

6.1.3. Motion modelling and tracking for distributed video compression

Participants: Christine Guillemot, Matthieu Maître, Luce Morin.

The 3D coders described above make the assumption that the video sequence comes from a unique camera moving in a static 3D environment with Lambertian surfaces. The corresponding motion models exhibit strong geometrical properties, which allow their parameters to be robustly estimated. They are of interest to specialized applications such as augmented reality, remote-controlled robots operating in hazardous environments or remote exploration by drones. We have also investigated several approaches with the aim of relaxing these constraints and considering generic 2D motions. The 3D coders gave two important insights into frame interpolation for distributed video compression: first, motion fields between keyframes need to be close to the groundtruth, and second, small but precise motion adjustments are required at intermediate frame to align the intermediate motion fields with the groundtruth. Both of these properties are more difficult to attain with 2D motions than with 3D motions due to the absence of epipolar geometry and of global motion model at intermediate frames (the intermediate projection matrices). Like the 3D codec, the 2D codec encodes the keyframes independently using H.264 intra. Sparse correspondences between keyframes are then detected at the decoder, which serve as an initialization for dense motion estimation. In the 2D case, correspondences are found between triplets of keyframes. Feature points are first detected independently on each keyframe using the Harris corner detector and the Laplacian-based blob detector at multiple scales. They are then matched to obtain correspondences using constraints on motion size, normalized cross-correlation between blurred-image descriptors, sampling-independent sum of absolute differences, acceleration, and unicity. Correspondences are propagated between triplets of keyframes to make them denser. Propagation is performed by alternatively removing correspondences with either poor motion smoothness (intra- and inter-scale) or with motion intersections, and matching feature points using constraints with weaker thresholds along with an additional

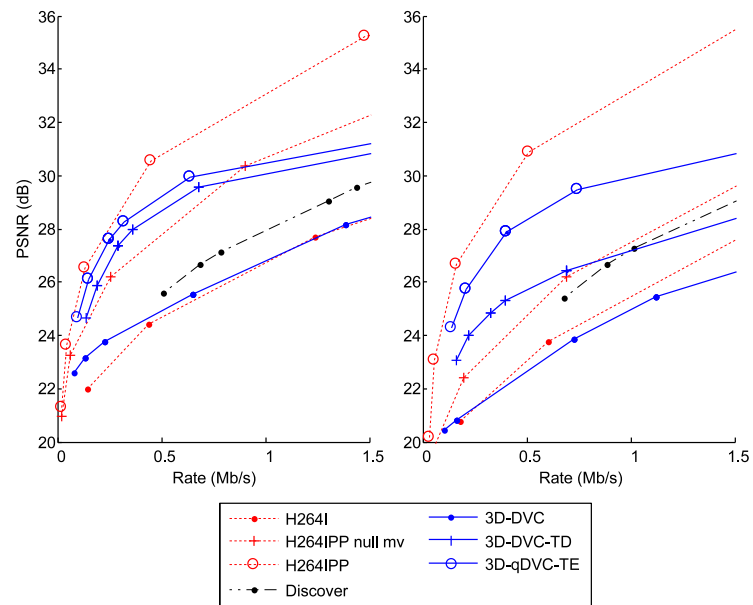
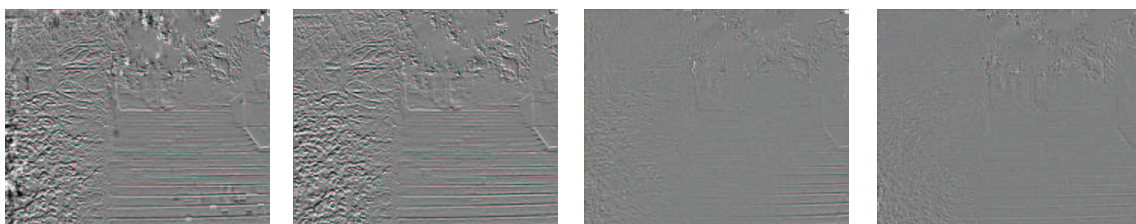


Figure 5. Rate-distortion curves for different H.264 modes, 2D-DVC and the three proposed 3D codecs (left: street sequence, right: stairway sequence).



(a) 2D-DVC

(b) 3D-DVC

(c) 3D-qDVC-TE

(d) 3D-DVC-TD

Figure 6. Correlation noise on the stairway sequence for different frame interpolation schemes.

constraint on motion smoothness. Correspondences are finally propagated once more, this time between pairs of keyframes. Dense motion estimation is cast into a maximum-a-posteriori problem where motion vectors are hidden variables forming a hidden Markov field with jointly-Gaussian probabilistic distributions. The noisy observed variables are the motions obtained by block-based subpixel matching using Lucas-Kanade optimization with Levenberg-Marquardt iterations. These observed variables are also jointly Gaussian, with covariance matrices equal to the Hessians of the block textures. Motion adjustment at intermediate frames is obtained by detecting edges at the encoder and tracking these edges at the decoder. Edges are found using the Canny edge detector. They are coded by edge linking and edge-chain downsampling followed by DPCM direction-coding with variable length codes. Edges are tracked using dense motion estimation between the decoded edge-images with the motion field between keyframes as prior. Preliminary experimental results, shown in part in Figures 7 and 8, indicate that the proposed frame interpolation (OF) outperforms the classical Block-Based Motion Interpolation (BBMI) in terms of side-information correlation and rate-distortion. However, the PSNR improvements brought by edge-based alignment are not yet sufficient to overcome the bitrate overhead they generate.

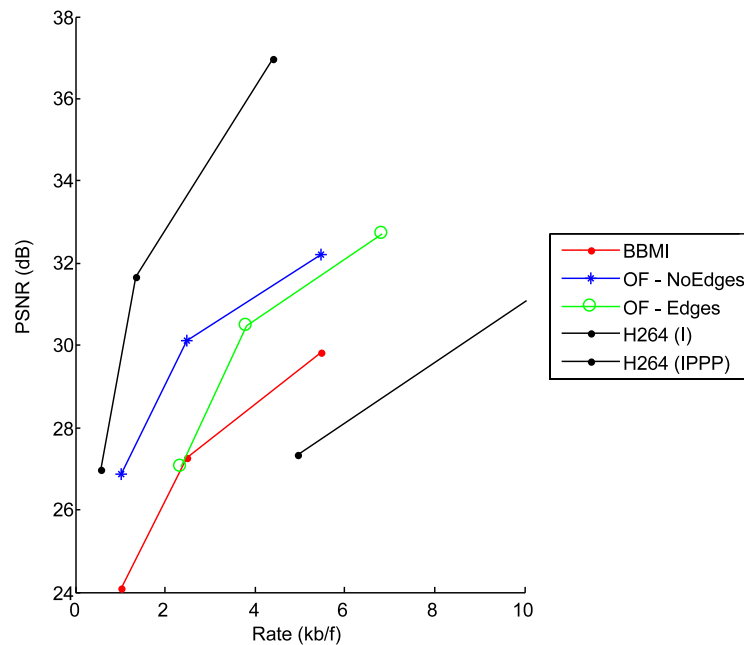


Figure 7. Rate-distortion curves from the foreman sequence for different H.264 modes, BBMI and OF with and without edge-alignment.

6.2. Sparse signal representation and compression

Keywords: *compression, frame expansions, rate-distortion theory, scalability, wavelet transforms.*

6.2.1. Sparse signal representation

Participant: Jean-Jacques Fuchs.



(a) BBMI



(b) OF

Figure 8. Correlation noise from the foreman sequence for different frame interpolation schemes.

In statistical modeling of signals or systems a lot of effort is in general devoted to first identify the structure of the model that best fits the observations. Recent studies have lead to an alternative approach in which sparsity in modeling is achieved without the need for a preliminary structure estimation procedure that is always difficult to achieve. It can be seen as a Bayesian or inverse-problem approach in which the classical maximum likelihood criterion is replaced by a compound criterion that combines fit of the model to the observations with prior information or sparsity requirements. We have mainly been considering a criterion that combines an ℓ_2 and ℓ_1 norms, where the ℓ_2 -part measures the fit of the model to the observations (e.g., the maximum likelihood criterion in the presence of Gaussian noise) and the ℓ_1 -part ensures parsimony of the representation. In case of linear parametrizations the criterion remains convex and problems with moderate to high number of unknowns are reliably solved with standard programs, such linear or quadratic programming, from well established scientific program libraries. Recently dedicated algorithms have also been developed that allow to handle far larger models. Using such an algorithm we have applied the Global Matched Filter (GMF) to Space Time Adaptive Processing (STAP). This allowed us to develop new techniques that are efficient even in case the environment is heterogeneous or target-rich [11], [22], [21], [23], [14]. Indeed, while the algorithmic part is easy, the analysis of the performance of parsimony based approaches is in general quite difficult. This means that while there are many applications with remarkable results, theoretical results are .. sparse and only trivial problems are amenable to solutions. This is a quite active domain of research among a quite small community however and the aim is of course to extend the few results obtained so far to more difficult scenarios or to different criteria. In [15] we propose to replace the usual ℓ_2 - ℓ_1 norms criterion by an ℓ_1 - ℓ_1 or ℓ_∞ - ℓ_1 norms criterion. Other extensions of these techniques are presented in [16], [17].

6.2.2. *Compressed representations of the Laplacian pyramid*

Participants: Christine Guillemot, Gagan Rath, Wenxian Yang.

Scalable representation of visual signals such as image and video is highly important for modern multimedia communications. In a heterogeneous communication network with diverse receiving devices, scalability allows to adapt the bit rate of the transmitted data to the network bandwidth, and/or the resolution of the transmitted data to the rendering capability of the receiving device. Scalability in spatio-temporal domain and/or in SNR is generally implemented by representing the original video signal in the form of a base layer signal together with several enhancement layer signals. The base layer signal provides the signal representation for a minimum level of reconstruction features whereas the enhancement layers upgrade or refine the base layer signal to the upper scalability levels. Spatial scalability, as the name suggests, provides layers with different spatial resolutions. This is usually achieved as a low resolution coarse signal together with several higher resolution enhancement layers. The lower base layer signal together with the higher enhancement layer signals constitutes the well-known Laplacian pyramid (LP) representation. In the context of scalable video coding (SVC), the compression of the spatial enhancement layers is an important issue. In the current SVC standard software JSVM, the enhancement layers are first transform coded using a 4x4 integer transform, and then the quantized transform coefficients are encoded using the context adaptive variable length coding (CAVLC). This coding procedure does not take into account the fact that the LP has a redundant construction. Another approach will be to convert it to a critical representation scheme through transforms which also compact the energy of the transform coefficients efficiently. When the LP has an open-loop configuration, that is, when the quantization of the coarse signal is performed outside the prediction loop, the transforms can be derived from the downsampling and the upsampling filters through Singular Value Decomposition (SVD) and QR factorization. Depending on the relationship between the filters, these transforms can be applied to the enhancement layers differently such that the resulting coefficients make a critical representation. When the enhancement layers of an LP are obtained through the closed-loop prediction, that is, having the decoded base layer for prediction, the conversion to a critical representation is not an efficient approach. The introduction of the quantizer in the prediction loop breaks the dependency structure in the coefficients. However, the insight obtained from the open-loop transforms can help to enhance the prediction and the application of the transforms. Improved prediction tends to lower the energy of the enhancement layer and the subsequent application of transforms can result in better energy compaction which finally can result in improved compression efficiency. Since the application of the improved prediction and the transforms do not always guarantee better rate-distortion performance over the current coding scheme, they can be incorporated

as optional modes in JSVM software. Experimental results with some standard test sequences demonstrate coding gains up to 1 dB for I pictures, and up to 0.7 dB for both I and P pictures.

6.2.3. *Multiresolution representations for compression and local description*

Participants: Christine Guillemot, François Tonnin.

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

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

The final problem is rarely addressed in the literature and consists in the proper quantization of transformed coefficients for both good image reconstruction and preservation of description quality. As the transform is redundant, one image has many possible representations. We first use POCS (Projection onto Convex Sets) to find a sparser representation and we adapt the classical technique in order to preserve the content of the neighborhoods of extracted points. Then, we design a compression scheme allowing the reconstruction of steerable coefficients from the information required by description, which is reduced to an energy and an orientation coefficient for each spatial point. The final step is the quantization of these coefficients. This compression scheme allows to detect illegal copies in image bases compressed at one bit per pixel. Videos appear to be the next challenge. On the one hand, HDTV represents sets of images event bigger that what is present in most still image collections. On the other hand, some functionalities are required by the professionals of the domain in order to develop there services: scalable coding to allow an easy distribution of the content on many platforms, copy detection as a complementary tool to DRM. These aspects form the core of the ICOS-HD project that will begin in 2007.

6.2.4. *Distributed video compression*

Participants: Christine Guillemot, Denis Kubasov, Khaled Lajnef, Jayanth Nayak, Gagan Rath.

In 2006, we have developed several approaches to increase the R-D performance of distributed video compression. We have developed a technique for side information extraction based on mesh-based motion compensated interpolation as well as for rate control based on an estimation of the noise of the correlation channel. A distributed predictive coding has also been developed to account for sources with memory. Predictive coding has been successfully employed in compressing the DC component in transform based video

coders. We have looked at extending this technique to the distributed video coding scenario. In conventional prediction, by making a set of simplifying assumptions, the rate-distortion function of the source being compressed can be shown to be equivalent to the memoryless rate distortion function of the residual. Therefore, for Gaussian sources, minimizing the variance of the residual becomes the goal of prediction filter design. In distributed source coding, we are given a pair of correlated sources, possibly with memory. Distributed prediction involves applying a pair of filters independently on the two source sequences. The goal of prediction filter design is not the same as above. In addition to the variance of the residuals, coding efficiency also depends on the correlation between the residuals at any given instant. In the multiterminal coding case, where both sources are compressed by separate encoders and there is a single decoder, a rate-distortion theoretic analysis led us to derive a cost function for prediction filter design, based on the determinant of the correlation matrix of the prediction residual at any given instant. A cost function has been derived similarly in the Wyner-Ziv coding case, where only one of the sources is encoded while the other source functions as side-information at the decoder. Unlike in conventional prediction, the algorithms that we have discovered for solving the two optimization problems are of an iterative nature and we do not know if they lead to the optimal solution. However, in all the cases that we tested, these algorithms have converged with reasonable results. As can be expected, different cost functions lead to different filters. For example, the Wyner-Ziv cost of a filter designed using the conventional cost function was up to 6dB more than the Wyner-Ziv cost of a filter designed to minimize the Wyner-Ziv cost itself, in the cases that we tested. We also studied the performance of the various predictors in a distributed coding system with uniform quantization and turbo code based Slepian-Wolf coding. With data that has a correlation structure of the DC component of a video frame, we observed that Wyner-Ziv prediction yields around 1.0 dB rate-distortion gain over the no prediction case, which compares favorably with the theoretically expected gain of 1.35 dB for this data. Distributed predictive coding is currently being incorporated into the Discover video coder.

6.3. Joint source-channel coding

Keywords: *error and erasure resilient coding and decoding, generalized likelihood ratio test, information theory, maximum likelihood estimation, multiple description coding, robust detection, stochastic modelling.*

6.3.1. Frames as joint source-channel codes for MIMO channels

Participants: Robin Chatterjee, Gagan Rath.

Multiple-input multiple-output (MIMO) wireless systems promise increased capacity due to parallel channels. One of the ways to improve the reliability in such systems is to use diversity in space and time. Diversity modulation in the form of sending redundant data from multiple antennas is an effective way of combating the malign effects of a fading channel such as amplitude and phase distortions, additive channel noise, etc. Space-time coding is a structured way of generating such redundant data from the symbol stream at the output of the channel modulator. A space-time block code, for instance, encodes a block of symbols to several blocks of symbols, called a space-time codeword, which are then mapped to different transmitting antennas. The structure in the code helps to decode the transmitted symbol vector more reliably in a fading environment. Another way of generating redundant data will be to expand the incoming signal through a frame operator at the source level. Frames provide a redundant set of spanning vectors for the Hilbert space. The expansion of the continuous-valued input signal vector by a frame thus provides a redundant set of continuous-valued coefficients representing the input vector. These redundant symbols can then be formatted to N substreams, where N denotes the number of transmitting antennas. Each stream then can follow the standard source coding and channel coding before being transmitted from different antennas. Such a system, though apparently simple, has many technical challenges. The frame expansion basically leads to an equivalent symbol coded system which is analogous to a space-time code, with the difference that the former accepts a continuous-valued vector instead of a modulation symbol-vector as input. For a given quantizer and subsequent index assignment, the symbol coded system is a function of the frame expansion operator. The structure of the resulting codebook decides the fundamental limits of the system such as the maximum achievable mutual information and the decoding error probability in a fading environment. By choosing the quantization parameter suitably, the maximum achievable mutual information can be improved over a space-time coded system with equivalent

quantization parameter at the source level. This, however, comes at the cost of increased probability of symbol error across the channel. Despite this fact, the frame operator at the source level has error correcting abilities. The frame operator is equivalent to the generator matrix of a block code in the real field. A simple projection of the decoded frame coefficient vector onto the signal space through the pseudo-inverse of the frame operator eliminates the noise component lying on the null-space of the code. More advanced error correction techniques can be performed at the decoder if the signal expansion at the encoder uses error correcting frames. Since the equivalent symbol coded system is a function of the frame expansion operator, the performance of this joint source-channel coded system can be optimized through the design of the frame operator. Depending on the application requirement, the optimization could be in terms of the maximum achievable mutual information or the minimum probability of error.

6.3.2. Multiple description coding

Participants: Olivier Crave, Christine Guillemot.

This study is carried out in collaboration with ENST-Paris (Béatrice Pesquet-Popescu and Christophe Tillier). Multiple description coding has been introduced as a generalization of source coding subject to a fidelity criterion for communication systems that use diversity to overcome channel impairments. Several correlated coded representations of the signal are created and transmitted on different channels. The design goals are therefore to achieve the best average rate-distortion performance when all the channels work, subject to constraints on the average distortion when only a subset of the channels is received correctly. Distributed source coding is related to the problem of separate encoding and joint decoding of correlated sources. This paradigm naturally imparts resilience to transmission noise. The duality between the two problems, that is multiple description coding (MDC) and distributed source coding (DSC), is being explored in order to design loss resilient video compression solutions. A first algorithm based on overcomplete temporal signal expansions and the Wyner-Ziv coding principles has been designed.

6.3.3. Error-resilient source codes and soft decoding

Participants: Christine Guillemot, Simon Malinowski.

In 2005, we have introduced a new set of state models to be used in soft-decision (or trellis) decoding of variable length codes. So far, two types of trellises have been considered to estimate the sequence of emitted symbols from the received noisy bitstream: the bit-level trellis proposed by Balakirsky and the bit-symbol trellis. The bit-level trellis leads to decoders of low complexity, however does not allow to exploit symbol *a priori* information (e.g., termination constraint), hence suffers from some sub-optimality. In contrast, the bit-symbol trellis allows to exploit *a priori* information on the sequence of symbols and, coupled with the BCJR algorithm, it allows to obtain sequence estimates minimizing the Bit Error Rate (BER) and the Symbol Error Rate (SER). However, the number of states of the bit/symbol trellis is a quadratic function of the sequence length, leading to a complexity not tractable for realistic applications. We have thus developed a novel set of state models and the corresponding trellises for the estimation of the Hidden Markov chain. The state model is defined by both the internal state of the VLC decoder (i.e., the internal node of the VLC codetree) and the rest of the Euclidean division of the symbol clock by a fixed parameter T . Therefore, the approach consists in aggregating states of the bit/symbol trellis which are distant of T instants of the symbol clock. If $T = 1$, the resulting trellis is equivalent to the usual bit-level trellis proposed by Balakirski. If T is greater or equal than the symbol sequence length $L(S)$, the trellis is equivalent to the bit/symbol trellis. The intermediate values of this parameter allow to gracefully trade complexity against the estimation accuracy. The state aggregation leads to close-to-optimum estimations with significantly reduced complexity.

In 2006, the state models have been extended for quasi-arithmetic codes. The choice of the parameter T is related to the capability of the codes to resynchronize. We have thus studied the error recovery capability of quasi-arithmetic codes with the help of a so-called error state diagram. Transfer functions defined on this error state diagram allow us to estimate the probability that the number of symbols in the transmitted and decoded sequences differ by a given amount ΔS . The entropy of this quantity gives the maximum amount of information that the soft decoder augmented with a length constraint will be able to exploit. We have shown that the probability that the quasi-arithmetic decoder does not re-synchronize in a strict sense (or

equivalently $P(\Delta S = 0)$) and the entropy of the termination constraint are not significantly altered by the state aggregation. This proves that the performances of a Viterbi decoder run on the aggregated trellis can be optimal for a significantly reduced complexity in comparison with the bit/symbol trellis. The complexity has also been compared with the one of the sub-optimal M-stack algorithm.

6.3.4. *Parameter estimation in convolutive and instantaneous mixtures*

Participant: Aline Roumy.

Many receivers need to first estimate some parameters before processing the received data. These parameters can be related to the sources (correlation between sources) or to the transmission (SNR, multipath coefficients). We have developed generic tools in order to evaluate the performance degradation due to channel imperfect knowledge at the receiver. We consider the impact of imperfect knowledge in 2 different contexts: (i) a convolutive mixture (ii) an instantaneous mixture. **Channel coefficients in a convolutive mixture:** In wireless communications, the transmission channel introduces time-varying multipath fading to the transmitted signal and hence, a convolutive mixture of the transmit data is built where the convolution coefficients are the multipath coefficients of the channel. Therefore an equalizer is needed to recover the transmitted data at the receiver. The optimal equalizer to be used is based on maximum *a posteriori* (MAP) detection and depends on the transmission channel, which is a priori unknown. Therefore, the receiver contains a channel estimation algorithm to estimate a proper channel parameter set. Moreover efficient equalizers has been proposed to take into account that the data are coded: the turbo-equalizer. It contains a MAP equalizer fed with *a priori* information on the transmitted data and provided by another module in the receiver, for instance the decoder.

This has motivated our study on the impact of channel estimation and *a priori* information in a maximum a posteriori (MAP) equalizer. More precisely we have first considered the case where the MAP equalizer is fed with a priori information on the transmitted data and studied analytically their impact on the MAP equalizer performance. Then we have assumed that the channel is not perfectly estimated and shown that the use of both the a priori information and the channel estimate is equivalent to a shift in terms of the signal-to-noise ratio (SNR) for which we have provided an analytical expression [13]. Then we have studied analytically the behavior of the whole turbo-equalizer (with perfect channel knowledge) [25]. The aim of this study was to perform in the future the analytical convergence analysis of turbo equalizers using MAP equalization and to derive statistical description of estimates resulting from iterative code-aided estimation algorithms. This work has been performed in collaboration with N. Sellami (ISECS, Sfax, Tunisia) in the Inria-DGRSRT project framework. This study of the robustness of the MAP equalizer naturally led us to the design of equalizers robust to channel estimation errors. Finally, in collaboration with V. Ramon, C. Herzet, L. Vandendorpe (Université Catholique de Louvain la Neuve, UCL, Belgium) under the umbrella of the NewCom project, we have extended the analysis to other equalizers as the Minimum Mean Square Error / Interference Cancellation equalizer. Sensitivity of this equalizer to an imperfect knowledge of the channel taps and signal-to-noise ratio (or, equivalently, noise variance) is analyzed [24]. We have also compared the robustness of all these equalizers to different parameter estimation errors. **SNR estimation in an instantaneous mixture:** A way to solve the MAC problem, is DS-CDMA (Direct Sequence Code Division Multiple Access). If synchronous and random spreading sequences are considered, an instantaneous mixture of all the senders is received. The receiver that minimizes the per-user bit error rate (BER) is the symbol Maximum a posteriori (MAP) detector. This receiver is derived under the hypothesis of perfect channel state information at the receiver. In [19] we consider the case where the channel noise variance is estimated and analyze the effect of this mismatch. We show that the Bit Error Rate (is piecewise monotonic wrt. the estimated noise variance, reaching its minimum for the true channel variance. We also provide an upper bound of the individually optimum receiver performance under noise variance mismatch. Thus we give a theoretical justification for the usual bias towards noise variance underestimation adopted by the community.

6.3.5. *Robust estimation and detection*

Participants: Jean-Jacques Fuchs, Sébastien Maria.

It often happens in communication systems that the information is carried by a signature whose exact value is generally assumed to be known and necessary to recover the transmitted information. If the signature is not precisely known robust procedures have to be implemented. Depending upon the type of communications system that is considered, the uncertainty on the signature may be due to many different reasons such as perturbations in the transmission channel, bad calibrations, diffraction, deformations or measurement noise. The objective is to develop detection schemes that are robust to a large class of modelling errors while requiring a small amount of prior information. We consider the case where in a communication system it is desired to recover a signal of interest in the presence of interferences and noise using an array of L sensors. The problem amounts to estimating the temporal waveform s_k in the model

$$y_k = as_k + i_k + e_k, \quad i_k = Au_k$$

where $y \in C^L$ is the observation at time k , $a \in C^L$ is the signature or steering vector, i_k denotes the interference and e_k the broadband noise. The interference is assumed to lie in a known linear subspace spanned by the columns of the known $L \times P$ matrix A and e_k is generally modelled as complex Gaussian noise whose covariance matrix is noise up to a multiplicative constant. In most situations, exact knowledge of a is difficult to obtain due to bad calibration of the sensors, uncertainties about the propagation, local scattering, etc. Different techniques have already been proposed to mitigate the effects of uncertainties on a . We propose a new statistical model for the uncertainty that leads quite naturally to an optimal detector. It consists in assuming that one has an observation say $b = a + \alpha n$ with n a perturbation vector, i.e. before the transmission starts one observes or knows a noisy version of the signature that will be in use. The techniques that are involved in the analysis of the test rely on optimization and perturbation analysis that are similar to the one that have been used in [11], [12], [18].

6.4. Distributed source-channel coding

Keywords: *MAC channels, Slepian-Wolf coding, Wyner-Ziv coding, information theory.*

In 2006, we have started a few theoretical studies in the area of distributed source coding and on the problem of transmission of correlated sources over multiple access channels.

6.4.1. Correlated sources over multiple access channels

Participants: Aline Roumy, Christine Guillemot.

In collaboration with O. Dabeer (Tata Institute of Fundamental Research, India), we study the case of transmission of correlated sources over the MAC channel. In this context, the source-channel separation theorem does not hold. More precisely, some counterexamples can be built where joint schemes outperform the separated scheme. We therefore focus on joint schemes but restrict ourself to the class of linear processes in order to propose low processing complexity algorithms. For correlated sources over the MAC, we derive the linear transmit filters which minimize the mean square error under some power constraint. This work is submitted to the National Conference on Communications (NCC-2007). We also investigate the case of sources with memory in a submission to the International Conference on Acoustics, Speech, and Signal Processing (ICASSP-2007). We further plan to derive rate achievability properties of the class of linear processes in the context of transmission of correlated sources over the MAC.

6.4.2. Distributed coding for sensor networks

Participant: Aline Roumy.

In collaboration with M. Debbah (Eurecom Institute, Sophia-Antipolis) A. Kherani and T. Banerjee (Dept. of Computer Science and Engineering, IIT Delhi, India), we study the optimal wireless node density of a sensor network. We impose a distributed and separated access scheme (at each sensor separate source and channel coding is performed) and investigate the tradeoff between accuracy of the field investigated (taking into account the correlation of the sources) and cost of the communication.

6.5. Image and video watermarking and security

Keywords: *Robust watermarking, detection of weak signal, side information.*

6.5.1. Optimal watermark detection

Participant: Teddy Furon.

The last annual report showed that zero-bit watermarking (also known as watermark detection) is a useful framework for applications requiring high robustness. The literature in the field shows two threads. The first group uses a classical additive spread spectrum watermarking embedder and finds the corresponding optimum detector. This is usually done by resorting to the theory of weak signal detection. It results in a one-sided test whose asymptotic optimal solution is known as a Locally Most Powerful (LMP) test. The detection function is basically the derivative of the likelihood function. The second group considers the former approach as suboptimal because the embedding does not take into account the side information composed by the host signal. This group then focuses on adapting Quantized Index Modulation (QIM) based watermarking (which is optimum for positive rate or zero rate watermarking) for zero-bit watermarking. However, so far, nobody has shown that QIM is a good method for this scenario. Our approach is mixed: from the first group, we retain the use of LMP test. Therefore, for a given embedding function, the general expression of the LMP test gives us the best detection function. From the second group, we retain the idea that the watermark signal must be very dependent on the host signal, without restricting the study to the QIM method. For a given detection function, we look for the embedding function which maximizes the asymptotic relative efficacy. According to the Pitman-Noether theorem, this solution is asymptotically the best scheme under a Neyman-Pearson test strategy. We have derived closed form equations for the optimum ends of the watermarking chain when the other end is given. It is pretty easy to globally optimize it, inserting the first equation to the other. This gives birth to a partial differential equation. The solutions are heavily dependent on the probability density function of the host signal. For Gaussian white host, optimum solutions are multivariate Hermite polynomials. For flat host, optimum solutions are sine and cosine periodical functions over a lattice. Under some other assumptions, these solutions are statistically robust under an attack channel, ie. they are still optimum when the watermarked content undergoes an attack.

6.5.2. Watermarking Security

Participants: Teddy Furon, Caroline Fontaine.

In 2004 and 2005, we have set a theoretical framework for assessing the security level of watermarking schemes. This framework is based on the measure of information about the secret key that leaks from the watermarked contents observed by an opponent. Within this framework, we have analyzed the security levels of two well-known watermarking schemes: substitutive and additive spread spectrum techniques. In 2006, we have analyzed the security levels of Quantized Index Modulation (QIM) based watermarking techniques. In QIM, a set of nested lattices is defined, each of them being associated to a symbol. The watermarked signal is (or, more usually, is attracted towards) the quantized version of the host signal onto the lattice related to the hidden symbol to be transmitted. The problem is that the optimal performances are reached for well-known lattices. Therefore, no security is provided. To introduce some security, it appears that performances are not degraded when the lattices are dithered (ie. geometrically shifted) by a secret vector, shared by the embedder and the decoder. This prevents the attacker to illegally read, write and erase the watermark message. However, the question of how long the secret dither will remain a real secret has never been investigated. We have first studied the mutual information between the dither and a set of watermarked contents (watermarked with the same secret dither). This problem is only tractable for few lattices. We have derived bounds which provides a lower (ie. pessimistic for the watermark designer) estimation of the security levels. However, information theoretic tools are useful for deriving bounds on the number of watermarked contents needed to accurately estimate the secret dither. But, they do not give any clue concerning the estimation algorithm the opponent shall run, and especially its complexity. Therefore, the second part of the framework aims at giving a proof of concept, providing at least *an* (ie., which may not be the optimum) algorithm with an affordable complexity. We have used a tool from the automatic, system identification community, so-called Set Membership Estimation (SME). Briefly, the watermarked signal can be regarded as a noisy observation

of the dither. But, the noise has a bounded support given by the lattice Voronoi cell. Thus, one observation gives a bounded feasible set of values for the dither. Observing a group of watermarked contents, the opponent estimates the dither by finding the intersection of all the feasible sets. However, this is not an easy task as the description of this intersection requires an increasing number of parameters as the number of observations grows. There exist techniques which reduces the complexity by working on estimation (ie., ellipsoid) of the intersection set. Part of the work has been supported by ACI Fabiano, and part of the work has been done in collaboration with the University of Vigo (Spain). In the area of security, we have also worked on the design of a dedicated architecture to enable a good cooperation between traditional DRM tools and digital watermarking, focusing on a new interaction between the SIM card and the watermarking schemes through cryptographic protocols; this solution enables, for the first time, the audio hardware to check the link between the real file it is playing and the rights embedded in the license and managed by the SIM card.

6.5.3. Fingerprinting

Participants: Caroline Fontaine, Fabien Galand.

Fingerprinting aims at hiding data in a robust and imperceptible way, embedding different data for each legitimate user of the content. The main goal is to enable traceability of the content and to resolve frauds. We have first focused on the choice of the embedded data that would enable best tracing capability. This choice is hard because we want to face some collusion attacks, in which colluders compare their contents in order to forge a new and untraceable one. So far, the problem has been addressed mostly by the error correcting codes' community, without any link with the signal processing community which is focusing on the embedding problem. The model of attack considered so far is not realistic at all. We have initialized a study of error correcting codes working with the Euclidean distance to see if they can be used directly with embedding techniques. At the same time, we are investigating models taking into account more realistic attacks.

7. Contracts and Grants with Industry

7.1. Industrial contracts

7.1.1. CRE with France Telecom on 3D reconstruction of urban scenes

Participants: Luce Morin, Gael Sourimant.

- Convention number : 104C1028000MPR012-ALLOC117
- Title : 3D reconstruction of urban scenes by fusion of GPS, GIS and video data.
- Research axis : § 6.1.1.
- Partners : France Télécom, Irisa/Inria-Rennes.
- Funding : France Télécom.
- Period : Oct.04-Sept.07.

This contract with France Telecom R&D (started in October 2004) aims at investigating the fusion of multi-modal data from video, GPS and GIS for 3D reconstruction of urban scenes. Video and GIS give complementary information: video provides photorealism, geometrical details, precision in the fronto-parallel axes; GIS provides a "clean" and complete geometry of the scene, structured into individual buildings. A GPS acquisition synchronized with video acquisition is added in order to provide a rough estimation of camera pose in a global coordinate system. In 2006, we have addressed the fundamental issue of video and GIS registration, which is a first required step before any combination of the data itself. The proposed approach is based on extraction and tracking of interest points, and camera pose estimation via robust visual servoing.

7.1.2. CRE with France Telecom on distributed video compression

Participants: Christine Guillemot, Khaled Lajnef, Gagan Rath.

- Convention number : ALLOC 396
- Title : Codage vidéo distribué
- Research axis : § 6.2.4.
- Partners : France Télécom, Irisa/Inria-Rennes.
- Funding : France Télécom.
- Period : Jan.04- Dec.06.

This contract with France Telecom R&D (started in November 2004) aims at investigating the distributed video compression paradigm and at assessing its potential for mobile light-weight encoding systems. In 2006, we have developed a distributed video compression algorithm. Architectural limitations present in state-of-the-art solutions have been addressed. In particular, a rate control approach based on turbo-decoding confidence measures and the achievable rate-distortion bounds has been designed. The rate-distortion performance of the system has also been improved by coupling a trellis coded quantization with the turbo-code based Slepian-Wolf coder. A rate control mechanism based on an estimation of the correlation channel statistics and on the minimum achievable Wyner-Ziv rate bound has been developed.

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

Participants: Christine Guillemot, Aurélie Martin.

- Convention number : under signature.
- Title : Déconsolution spectrale appliquée à la compression
- Research axis : § 6.2.4.
- Partners : Thomson, Irisa/Inria-Rennes.
- Funding : Thomson, ANRT.
- Period : Oct.06- Sept.09.

This contract aims at developing spectral deconvolution algorithms for video prediction and compression as well as for error concealment. The problem will be addressed for both the texture and dense motion fields.

7.2. National contracts

7.2.1. RIAM-COPARO

Participants: Christine Guillemot, Marion Jeanne, Florelle Pauchet, Wenxian Yang.

- Convention number : 504C11080031324011
- Title : Codeur H.264 sur architecture parallèle programmable.
- Research axis : § 6.3.3.
- Partners : ENVIVIO, Irisa/Inria-Rennes, Vitec.
- Funding : Ministry of industry.
- Period : Sept.04- Jun.06.

The H.264 standard has been retained as the compression format for terrestrial television. In that context, the objectives of the COPARO project are to develop

- a parallel programmable architecture for real-time H.264 based video compression (Vitec);
- an H.264 real-time video encoder (Envivio);
- solutions of error resilience for H.264 based video compression and of scalability (INRIA).

One of the key components bringing the high compression performance in the H.264 solution is a new entropy coding scheme named CABAC (Context-based adaptive binary arithmetic coding). However, the CABAC algorithm suffers from sensitivity to transmission noise. TEMICS is bringing to the project algorithmic tools, allowing to make the CABAC encoding technique (and therefore the H.264 video compression solution) resilient to transmission errors typical of wireless links. The principle of soft source decoding is to use the structure and/or the statistics related to the source, encoder and channel models in order to get an optimal, or a nearly optimal, estimation of the transmitted symbols. The techniques developed concern soft-in soft-out decoding of arithmetic codes and in-line estimation of the source statistics for variable length codes. The work includes theoretical studies as well as validation in the H264 video codec.

7.2.2. *RNRT-COSINUS*

Participants: Christel Chamaret, Laurent Guillo, Adrien Schaddle.

- Convention number : ALLOC 157;
- Title : COSINUS (COmmunications de Services temps-réel / IP dans un réseau Sans fil) / Real-time IP service communication in a wireless network
- Research axis : § 5.1.
- Partners : Alcatel CIT, Institut EURECOM, IRISA /INIRIA Rennes, France Télécom, GET/ENST Bretagne, Thales Communications.
- Funding : Ministry of industry.
- Period : Dec. 04 - Dec.06.

The main objective of the COSINUS project is to demonstrate the feasibility of real time services on IPv6 wireless networks (UMTS or WLAN). It addresses the following issues: Controlling the quality as perceived by the user, accounting for the specific nature and quality of the wireless link, managing the diversity of access networks (UMTS, WLAN). In this perspective, the project partners study the following technical aspects: header compression protocols (notably ROHC), unequal error protection (UEP) techniques, audio and video source encoding that is resilient to radio errors and self-adaptive for bit-rates, perceived quality assessment methods. TEMICS contributes on the issue of video streaming with resilience and QoS support on UMTS links. TEMICS' contribution is twofold. First, the streaming server must take into account information from the receiver about the perceived quality and also from the link layer about the radio link status. Once computed, such information can, for instance, provide the server with bandwidth estimations. Then, the server can take advantage of the scalable video to regulate his sending rate. The second part of the contribution deals with resilience to transmission impairments on the UMTS channel.

7.2.3. *RNRT de l'ANR*

Participant: Jean-Jacques Fuchs.

The project COHEDQ 40 "COHerent DETection for QPSK 40GHz/s systems" whose coordinator is Alcatel has been retained by the ANR in July 2006. As far as Irisa is concerned, the work will be done by ASPI and TEMICS.

7.2.4. *RIAM-ESTIVALE*

Participants: Caroline Fontaine, Teddy Furon, Fabien Galand.

- Convention number : ANR-5A0638
- Title : Échanges Sécurisés pour le Transfert d'Informations Vidéo, en Accord avec la Législation et l'Économie
- Research axis : § 6.5.3.
- Partners : Academic partners: LIS (INPG), ADIS (Univ. Paris XI), CERDI (Univ. Paris XI), LSS (Univ. Paris XI/Supélec); Industrial partners: Basic-Lead, Nextamp, SACD.
- Funding : ANR.
- Period : 31/03/2006-31/03/2009

ESTIVALE is a project dealing with the diffusion of video on demand in several contexts: from personal use to professional use. People involved in the project are from different communities: signal processing and security, economists and jurist. Our aim 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. More precisely, TEMICS is in charge of the design of fingerprinting techniques, and involved in the review of DRM and cryptographic tools.

7.3. European contracts

7.3.1. FP6-IST NoE SIMILAR

Participants: Christine Guillemot, Khaled Lajnef, Luce Morin, Gael Sourimant.

- Convention number: 104C05310031324005
- Title: *European research taskforce creating human-machine interfaces SIMILAR to human-human communication.*
- Research axis: § 6.1.1, § 6.1.2, § 6.1.3
- Partners: around 40 partners from 16 countries.
- Funding: CEE.
- Period: Jan.04-Dec.07.

The TEMICS team is involved in the network of excellence SIMILAR federating European fundamental research on multimodal human-machine interfaces and contributes on the following aspects:

- In the context of 3D modelling of video sequences we have focused on an hybrid representation mixing 2D and 3D representations of video data. Cylindrical and spherical mosaics are used for unified coding and visualization of 2D and 3D data. Such an approach allows to make no assumption on the camera acquisition path, and still provides the benefits of 3D functionalities for virtual reality applications. We have also studied the fusion of multimodal data (e.g. real 2D video, synthetic 3d models) in the context of urban scenes automatic modelling.
- TEMICS is contributing on a distributed coding framework in a context of multimodality. We have in particular developed distributed video codecs based on 3D and 2D modelling of motion in video sequences.

7.3.2. FP6-IST NOE NEWCOM

Participant: Aline Roumy.

- Convention number: NEWCOM-N/G-520502
- Title: NEWCOM: Network of Excellence in Wireless Communication.
- Research axis: § 6.3.4
- Funding: CEE.
- Period: March 2004 - March 2007.

The NEWCOM project proposal (Network of Excellence in Wireless COMMunication) addresses the design of systems “beyond 3G”. This requires to successfully solve problems such as: the inter-technology mobility management between 3G and ad-hoc wireless LANs, the coexistence of a variety of traffic/services with different and sometimes conflicting Quality of Service (QoS) requirements, new multiple-access techniques in a hostile environment like a channel severely affected by frequency selective fading, the quest for higher data rates also in the overlay cellular system, scaling with those feasible in a wireless LAN environment, permitting seamless handover with the same degree of service to the user, the cross-layer optimisation of physical coding/modulation schemes with the medium access control (MAC) protocols to conform with fully packetised transmission as well as the TCP/IP rules of the core network, and the like. In 2006, we have, in collaboration with David Gesbert and Merouane Debbah from Eurecom, Sophia-Antipolis, studied the optimal wireless node density of a sensor network. In collaboration with Valery Ramon, Cedric Herzet, and Luc Vandendorpe from Université Catholique de Louvain la Neuve, we have studied the performance degradation of iterative schemes due to channel and SNR estimation errors.

7.3.3. FP6-IST STREP DANAÉ

Participants: Christine Guillemot, Gagan Rath.

- Convention number: 104C045731324005
- Title: *Dynamic and distributed Adaptation of scalable multimedia content in a context-Aware Environment*.
- Research axis: § 6.2.2
- Partners: ENST, France Télécom, Imperial College London (ICL), Inria, Museon, Siemens, T-systems, University of Aachen, University of Geneva, University of Klagenfurt.
- Funding: CEE.
- Period: Jan.04-June.06.

The TEMICS team is involved in the STREP DANAÉ addressing issues of dynamic and distributed adaptation of scalable multimedia content in a context-aware environment. Its objectives are to specify, develop, integrate and validate in a testbed a complete framework able to provide end-to-end quality of (multimedia) service at a minimal cost to the end-user. TEMICS contributes on the aspects of fine grain scalable video coding and on the study of new source codes for increasing the error resiliency of the scalable video coder while preserving its compression and scalable properties. In collaboration with other DANAÉ partners, TEMICS contributes to different core experiments defined in the context of MPEG-21/SVC: a core experiment on spatial transforms, on error resilience and on coding with multi-rate adaptability.

7.3.4. FP6-IST STREP DISCOVER

Participants: Christine Guillemot, Denis Kubasov, Jayanth Nayak, Gagan Rath.

- Convention number: ALLOC 1336
- Title: Distributed Coding for Video Services
- Research axis: § 6.2.4
- Universitat Politècnica de Catalunya (UPC), Instituto Superior Técnico (IST), Ecole Polytechnique Fédérale de Lausanne (EPFL), Universität Hannover (UH), Institut National de Recherche en Informatique et en Automatique (INRIA-Rennes) Università di Brescia (UB).
- Funding: CEE.
- Period: Sept.05-Aug.07.

Video coding solutions so far have been adopting a paradigm where it is the task of the encoder to explore the source statistics, leading to a complexity balance where complex encoders interact with simpler decoders. This paradigm is strongly dominated and determined by applications such as broadcasting, video on demand, and video streaming. Distributed Video Coding (DVC) adopts a completely different coding paradigm by giving the decoder the task to exploit - partly or wholly - the source statistics to achieve efficient compression. This change of paradigm also moves the encoder-decoder complexity balance, allowing the provision of efficient compression solutions with simple encoders and complex decoders. This new coding paradigm is particularly adequate to emerging applications such as wireless video cameras and wireless low-power surveillance networks, disposable video cameras, certain medical applications, sensor networks, multi-view image acquisition, networked camcorders, etc., where low complexity encoders are a must because memory, computational power, and energy are scarce. The objective of DISCOVER is to explore and to propose new video coding schemes and tools in the area of Distributed Video Coding with a strong potential for new applications, targeting new advances in coding efficiency, error resiliency, scalability, and model based-video coding. TEMICS is coordinating - and contributing to - the workpackage dealing with the development of the theoretical framework and the development of Wyner-Ziv specific tools. TEMICS also contributes to the development of algorithmic tools for the complete coding/decoding architecture and to the integration of the complete video codec.

8. Other Grants and Activities

8.1. National initiatives

8.1.1. ACI *Fabriano*

Participant: Teddy Furon.

- Convention number: 103C17280031324011
- Title : Fabriano
- Research axis : § 6.5.1, § 6.5.2.
- Partners : CERDI, INRIA (TEMICS), LIS, LSS.
- Funding : Ministry of research, CNRS, INRIA.
- Period : Mid-Dec. 03 - Sept. 06.

Fabriano is an ACI (Action Concertée Incitative) dedicated to the study of technical solutions to the problem of security based on watermarking and steganography. In particular, this action aims at developing a theoretical framework for stegano-analysis to be applied to the design of algorithms that will allow to detect the presence of a message within a signal in the respect of rights and ethical issues. TEMICS proposed a theoretical framework for security level assessment of watermarking technique. It has been applied to the most famous schemes: substitution, spread spectrum and lattice quantized index modulation.

8.1.2. ACI *Masse de données - Codage*

Participants: Olivier Crave, Christine Guillemot, Simon Malinowski.

- Convention number: 104C07410031324011
- Title : Codage de masse de données
- Research axis : § 6.3.2.
- Partners : ENST-Paris, INRIA (TEMICS), I3S Université de Nice-Sophia Antipolis.
- Funding : Ministry of research.
- Period : Mid-Dec. 03 - Dec. 06.

The objective of this project is to federate research effort in the two following areas:

1. Motion-compensated spatio-temporal wavelet (MCSTW) scalable coding: Tools for scalability available in existing standards usually lack compression efficiency, and are not flexible enough to achieve combination of different scalability dimensions (e.g. spatial, temporal, SNR, object and complexity scalability) and sufficient fine granularity. MCSTW offers the ideal framework for scalable compression of video sequences. Precise research tasks include scalable motion estimation and coding methods, non-linear adaptive wavelet decompositions, more appropriate for representing temporal residuals, techniques for progressive transmission of information (embedded coding, multiple description coding, ...).
2. Distributed source video coding: Traditional predictive coding, exploiting temporal correlations in a sequence through computational-intensive motion estimation between successive frames leads to encoders with a complexity 5 to 10 times higher than the complexity of the decoders. This is well suited to streaming or broadcasting applications, but not to a transmission from a mobile terminal to a base station or for peer-to-peer mobile communications. The project is investigating multi-terminal and distributed source coding solutions building upon dualities with multiple description coding and with channel coding with side information.

8.2. International initiatives

Participant: Aline Roumy.

- Convention number: 05 I 17
- Title : Analysis of iterative turbo-like receivers
- Research axis : § 6.3.4.
- Partners : Inria-DGRSRT/Tunisian university.
- Funding : Inria-DGRSRT/Tunisian university.
- Period : Jan. 05 - Dec. 06.

This is a collaboration with N. Sellami (ISECS, Sfax, Tunisia) and I. Fijalkow (ETIS, Cergy France). The goal of the proposed project is the analysis of turbo-like receivers and more particularly the robustness to channel estimation errors. The grant supports travel and living expenses of investigators for short visits to partner institutions abroad.

9. Dissemination

9.1. Standardization

- G. Rath, W. Yang, C. Guillemot and V. Bottreau, "Improved prediction and transform for spatial scalability", ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, JVT-T082, Kalgenfurt, July 2006.

9.2. Invited talks

- T. Furon gave invited talks on watermark benchmarking at the WaCha'06 conference organized by the Network of Excellence ECRYPT, on watermarking security at the GREYC lab, Caen, and on watermarking detection at the INPG lab, Grenoble.
- T. Furon participated to a round table for the National RIAM network.
- T. Furon gave a co-invited talk at the International Workshop on Digital Watermarking, Itsu Island, South Korea.
- C. Guillemot gave an invited tutorial of 3 hours on joint source-channel coding at the international conference "Mathematical Techniques and Problems in Telecommunications", Leiria, Portugal, Sept. 2006.
- G. Rath gave an invited talk on subspace techniques for error localization and correction at EPFL, May 2006
- A. Roumy gave an invited talk on the effect of parameter estimation at Eurecom, Sophia, France.

9.3. Leadership within the scientific community

- C. Fontaine is associate editor of the Journal in Computer Virology (Springer-Verlag);
- C. Fontaine is member of program committees of the following conferences: Indocrypt 2005, SSTIC 2006;
- C. Fontaine is member of the advisory board of the PhD thesis SPECIF's price;
- C. Fontaine is 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 : SAM2006 (Sensor Array and Multichannel Signal Processing Workshop), Eusipco 2006, GretsI 2007;
- T. Furon is associate editor of the EURASIP journal on Information Security;

- T. Furon is member of program committees of the following conferences: SPIE 2006 Security, Steganography, and Watermarking of Multimedia Contents VIII, ACM MM&Sec 2006, and IWDW 2006;
- C. Guillemot is associate editor of the journal IEEE Transactions on Circuit and System for Video Technology and of the international journal "New Trends in Signal Processing".
- C. Guillemot is elected member of the IEEE IMDSP (Image and MultiDimensional Signal Processing Technical Committee) and IEEE MMSP (MultiMedia Signal Processing Technical Committee) international committees;
- C. Guillemot is member of the external scientific advisory board of the IST-FP6 Network of Excellence VISNET;
- C. Guillemot is a french representative within the management committee of COST (Action COST 292 "Sematic and multimodal analysis of digital media") ;
- C. Guillemot is member of the program committees of the following conferences: IEEE-ICIP 2006, IEEE-MMSP 2006, WIAMIS 2006, CORESA 2006;

9.4. Awards

The PhD thesis of H. Jégou has received the thesis award of the section signal and image of the club EEA.

9.5. Invitations and visits

- A. Roumy visited M. Debbah and D. Gesbert (Eurecom Institute, Sophia-Antipolis, France) in May and June 2006 (NewCom network of excellence).
- A. Roumy visited N. Sellami (ISECS, Sfax Tunisia) in November 2006. (Inria-DGRSRT project).
- N. Sellami (ISECS, Sfax, Tunisia) visited the Temics group in July 2006.
- O. Dabeer (School of Technology and Computer Science, Tata Institute of Fundamental Research, Mumbai, India) visited the TEMICS group from April to June 2006.
- C. Dikici (Ph.D student at INSA-Lyon under the supervision of Prof. Attila Baskurt) visited TEMICS between Sept and Dec. 2006 in the context of a collaboration with INSA-Lyon.

9.6. Teaching

- Enic, Villeneuve-d'Ascq, (C. Guillemot: Video communication) ;
- INSA, Lyon, (C. Guillemot: Video communication) ;
- Master, Network Engineering, university of Rennes I (L. Guillo, Video streaming) ;
- Computer science and telecommunications magistère program, Ecole Normale Supérieure de Cachan, Ker Lann campus. (A. Roumy: Information theory and communication theory) ;
- Master SIC (Systèmes Intelligents et Communicants) at ENSEA, université de Cergy Pontoise. (A. Roumy: Information theory, Modern coding theory and Multiuser detection) ;
- Master of Science in Mobile Communications at Eurecom Institute, Sophia-Antipolis. (A. Roumy: Channel coding theory) ;
- Engineer degree Diic-INC, Ifsic-Spm, university of Rennes 1 (L. Morin, C. Guillemot, L. Guillo, T. Furon, Gaël Sourimant : image processing, 3D vision, motion, coding, compression, cryptography, communication) ;
- Engineer degree Diic-LSI, Ifsic-Spm, university of Rennes 1 (L. Morin, L. Guillo, Gaël Sourimant : compression, video streaming) ;

- Engineer degree DIIC, Ifsic-Spm, Université de Rennes 1: Jean-Jacques Fuchs teaches several courses on basic signal processing and control ;
- Master Research 2: STI: Jean-jacques Fuchs teaches a course on optimization ;
- Master, Security of Information Systems, Supelec-ENSTB (C. Fontaine) ;
- Master Degree, University of Montpellier (L. Morin :3D modelling for video compression) ;
- Professional degree Tais-Cian, Breton Digital Campus (L. Morin, Gaël Sourimant : Digital Images -online course-) ;
- Supelec (T. Furon : steganography and watermarking).

10. Bibliography

Major publications by the team in recent years

- [1] T. FURON, P. DUHAMEL. *An Asymmetric Watermarking Method*, in "IEEE Trans. on Signal Processing", Special Issue on Signal Processing for Data Hiding in Digital Media and Secure Content Delivery, vol. 51, n^o 4, April 2003, p. 981-995.
- [2] C. GUILLEMOT, T. TURLETTI. *Systèmes multimédia communicants*, chap. Algorithmes de codage et de contrôle de transmission sur réseaux hétérogènes, Éditions Hermès, 2000, p. 25-82.
- [3] 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.
- [4] A. GUYADER, E. FABRE, C. GUILLEMOT, M. ROBERT. *Joint source-channel turbo decoding of entropy-coded sources*, in "IEEE Journal on Selected Areas in Communications, special issue on the turbo principle : from theory to practice", vol. 19, n^o 9, sept. 2001, p. 1680-1696.
- [5] R. MOHR, L. MORIN, E. GROSSO. *Relative positioning with uncalibrated cameras*, in "in : Geometric Invariance in Computer Vision, Editors: A. Zisserman J.L. Mundy, MIT Press", 1992, p. 440-460.
- [6] 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.
- [7] S. PATEUX, G. LE GUELVOUT. *Practical watermarking scheme based on wide spread spectrum and game theory*, in "Signal Processing : Image Communication", n^o 18, Apr. 2003, p. 283-296.
- [8] G. RATH, C. GUILLEMOT. *Performance analysis and recursive syndrome decoding of DFT codes for bursty erasure recovery*, in "IEEE Transactions on Signal Processing", vol. 51, n^o 5, May 2003, p. 1335-1350.
- [9] G. RATH, C. GUILLEMOT. *Frame-theoretic analysis of DFT codes with erasures*, in "IEEE Transactions on Signal Processing", vol. 52, n^o 2, Feb. 2004, p. 447-460.
- [10] P. RAULT, C. GUILLEMOT. *Symmetric delay factorization: generalized framework for paraunitary filter banks*, in "IEEE Trans. on signal processing", vol. 47, n^o 12, Dec. 1999, p. 3315-3325.

Year Publications

Doctoral dissertations and Habilitation theses

- [11] S. MARIA. *Modélisation parcimonieuse: application à la sélection de variables et aux données STAP*, Ph. D. Thesis, Thèse de doctorat en Traitement du Signal, Université de Rennes 1, France, Dec. 2006.

Articles in refereed journals and book chapters

- [12] J. FUCHS, S. MARIA. *A new approach to variable selection using the TLS approach.*, in "IEEE Transactions on Signal Processing, accepted", 2006.
- [13] N. SELLAMI, A. ROUMY, I. FIJALKOW. *The impact of both a priori information and channel estimation errors on the MAP equalizer performance*, in "IEEE Trans. on Signal Processing", July 2006, p. 2716 - 2724.

Publications in Conferences and Workshops

- [14] F. L. CHEVALIER, S. MARIA. *STAP processing without noise-only reference: requirements and solutions*, in "Proc.of the CIE International conference on Radar, Shanghai, China", October 2006.
- [15] J. FUCHS. *Recovery conditions of sparse representations in the presence of noise*, in "Proceedings of the IEEE Conference on Acoustic, Speech and Signal Processing, Toulouse, France", may 2006.
- [16] J. FUCHS. *Sparse representations and realization theory*, in "Proceedings of the 17th Int. Symp. on Mathematical Theory of Networks and Systems, Kyoto, Japan", july 2006.
- [17] J. FUCHS. *Towards a new matrix decomposition*, in "Proceedings of the 17th Int. Symp. on Mathematical Theory of Networks and Systems, Kyoto, Japan", july 2006.
- [18] J. FUCHS. *A robust signal detection scheme*, in "Proc.of 14th European Signal Processing conference, EUSIPCO", September 2006.
- [19] R. GAUDEL, F. BONNET, J. DOMELEVO-ENTFELLNER, A. ROUMY. *Noise variance estimation in DS-CDMA and its effects on the individually optimum receiver*, in "IEEE International Workshop on Signal Processing Advances for Wireless Communications, SPAWC 2006", June 2006.
- [20] M. MAITRE, C. GUILLEMOT, L. MORIN. *3D scene modeling for distributed video coding*, in "Proc. of IEEE Int. Conf. on Image Processing, Atlanta, USA", oct 2006.
- [21] S. MARIA, J. FUCHS, L. FERRO-FAMIL, E. POTTIER. *Application of the Global Matched Filter to Space-Time Adaptive Processing*, in "Proceedings of EUSAR'06, Dresden, Germany", may 2006.
- [22] S. MARIA, J. FUCHS. *Application of the Global Matched Filter to STAP data: an efficient algorithmic approach*, in "Proceedings of the IEEE Conference on Acoustic, Speech and Signal Processing, Toulouse, France", may 2006.
- [23] S. MARIA, J. FUCHS. *Detection performance for GMF applied to STAP data*, in "Proc.of 14th European Signal Processing conference, EUSIPCO", September 2006.

-
- [24] V. RAMON, A. ROUMY, C. HERZET, L. VANDENDORPE. *Calculating the performance degradation of a MMSE/IC turbo-equalization scheme due to SNR estimation errors*, in "IEEE International Conference on Communications (ICC 2006)", June 2006.
- [25] N. SELLAMI, A. ROUMY, I. FIJALKOW. *Convergence analysis of the MAP turbo detector: New representation and convergence to the AWGN case*, in "Proc.of 14th European Signal Processing conference, EUSIPCO", September 2006.