INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Project-Team TEMICS

TraitEment, Modélisation d'Images et CommunicationS

Rennes

THEME COG

Activity Report

2005

Table of contents

1. Team

Head of project-team
Christine Guillemot [DR Inria]

Administrative assistant
Huguette Béchu [TR Inria, shared with VISTA]

Inria research scientists
Teddy Furon [CR]
Henri Nicolas [CR until August 31st, 2005]
Gagan Rath [CR]

Cnrs research scientist
Caroline Fontaine [CR, since Sept. 1st, 2005]

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

CNRS technical staff
Laurent Guillo [IR]

Inria projects technical staff
Vincent Bottreau [IST-DANAE project, until August 31st, 2005]
Florelle Pauchet [COPARO project, since April 1st, 2005]
Adrien Schadle [COSINUS project, since May 1st, 2005]
Christel Chamaret [COSINUS project, since Sept. 1st, 2005]

Inria junior technical staff
Cécile Marc [until April 30th, 2005]

Inria post-doctoral fellows
François Cayre [IST-DANAE project, until August 31st, 2005]
Raja Durai [CRE France Télécom, since May 1st, 2005]
Ceilidh Hoffman [50% TEMICS, 50% PLANETE Sophia-Antipolis, until August 31st, 2005]
Marion Jeanne [COPARO Project, until April 31st, 2005]
Jayanth Nayak [since sept. 1st, 2005]
Wenxian Yang [COPARO Project, since sept. 1st, 2005]

Ph.D Students
Raphaèle Balter [Cifre grant, France Télécom, until March 31st, 2005]
Vivien Chappelier [Ministry of research grant]
Robin Chatterjee [Inria grant, IST-DANAE project]
Olivier Crave [also with ENST-PARIS, ENST grant, from Oct. 1st, 2005]
Hervé Jégou [Ens-Cachan grant]
Denis Kubasov [[Inria grant, IST-DISCOVER project, from Oct. 1st, 2005]
Khaled Lajnef [Inria grant, CRE France Télécom]
Simon Malinowski [Ministry of research grant, from Oct. 1st, 2005]
Gael Sourimant [Inria grant, CRE France Télécom]
François Tonnin [Inria- Brittany Region grant, also with Texmex]

2. Overall Objectives

2.1. Introduction

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).
- *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.
- *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.

- *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) and European projects (IST-DANAE, IST-SIMILAR and IST-DISCOVER). 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 $\mathcal{C}_i$ is modelled as a *perspective projection*, the *projection equations* are :

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

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 a 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's are the frame vectors and I is the index set. The ith frame expansion coefficient of $\mathbf{x}$ is defined as $(F\mathbf{x})_i \equiv \langle\phi_i, \mathbf{x}\rangle$, 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 $\tilde{\Phi}_F \equiv \{(F^hF)^{-1}\phi_i\}_{i\in I}$. 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. 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 $E(.)$ be the extraction function going from the content space $\mathcal{C}$ to the components space, isomorphic to $\mathbf{R}^N$

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

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 $S(.)$ mapping a discrete message $m \in \mathcal{M}$ onto a signal in $\mathbf{R}^N$:

$$S(.): \quad \begin{array}{ccc} \mathcal{M} & \mapsto & \mathbf{R}^N \\ m & \rightarrow & \mathbf{S} = S(m) \end{array}$$

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

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

A new paradigm stating that the original host signal $\mathbf{V}$ shall be considered as a *channel state* only known at the embedding side rather than a source of noise, 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.

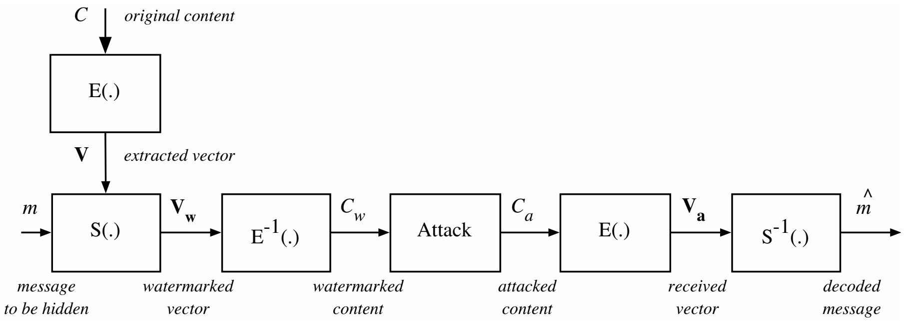

Figure 1. Classical watermarking scheme

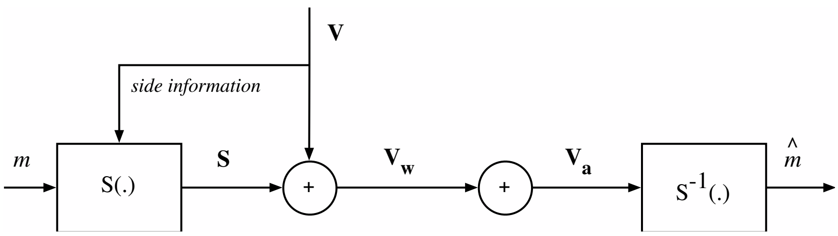

Figure 2. Watermarking as a problem of communication with side information.

4. Application Domains

4.1. Scope

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 preferrably 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 additionnal 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: Laurent Guillo [contact person], Cécile Marc, 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 to study and to assess, in a realistic way, new algorithms implementing joint source channel coding, video modelling or video coding. The platform is still under development. It includes:

- the software library MOVIQS (*module pour de la vidéo sur Internet avec qualité de service*) (registered at the Agency for the Protection of Programmes (APP) under the number IDDN.FR.001.030031.000.S.P.2003.000.10200). It is a dynamic link library used by a video streaming server and the related clients. They can take advantage of three of its main mechanisms: video transport in both unicast and multicast mode, congestion control and rate regulation, and loss control.
- WULL, a windows UDP-Lite library, that implements the UDP-Lite protocol according to the very new RFC 3828. UDP-Lite is a new transport protocol recommended for applications which would rather receive damaged data than having corrupted data discarded by the network. UDP-Lite is especially useful for applications that are error tolerant, and in particular to video codecs such as WAVIX developped by TEMICS. Using UDP-Lite as transport protocol in our video communication platform allows the transport layer to signal to the application layer (e.g., the video decoder) that it is going to process corrupted data. Taking into account this information can significantly improve the decoding speed. Wull provides developers with specific UDP-Lite sockets they can use in the same way they use classic UDP socket. It can be used with both IPv4 and IPv6. As partners in the IST-DANAE project, France Telecom and T-systems are using WULL. WULL 1.0 has been registered at the Agency for the Protection of Programmes (APP) under the number IDDN.FR.001.270018.000.S.P.2004.000.10000.
- A ROHC compression engine: ROHC stands for RObust Header Compression protocol. ROHC provides a compression scheme that can drastically reduce the size of packet headers. Due to limited bandwidth, IP/UDP/RTP/TCP packets sent over cellular links benefit considerably from header compression. Several profiles are described in the RFC 3095, such as IP/UDP or IP/UDP/RTP for both IPv4 and IPv6 and have been implemented by the ARMOR team and the ENST-Bretagne. New profiles have been added via a collaboration between the ARMOR and the TEMICS teams. They involve the new protocol UPD-Lite and have been validated with WULL. These new profiles are IP/UDP-Lite and IP/UDP-Lite/RTP for both IPv4 and IPv6. These extensions are currently being registered at the Agency for the Protection of Programmes (APP).

In 2005, wa have started the integration of the above modules and libraries into a streaming server. This streaming server must be able to efficiently stream video over wireless links. To do so, it will be able to 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. Algorithms implemented in MOVIQS will be used. Then, the server can take advantage of a scalable video encoded by the codec WAVIX to regulate its sending rate. A wireless link can be noisy and receivers might get erroneous data or entire packets might be lost. To cope with such problems, and in partcular to optimize the throughput seen by the receiver when a video is streamed over the wireless link, bit error and packet loss resilience techniques can be used in conjunction with UDP-lite.

5.2. WAVIX: Wavelet based video codec

Participants: Christine Guillemot, Laurent Guillo [contact person], Cécile Marc.

The software WAVIX (Wavelet based Video Coder with Scalability) is a low rate fine grain scalable video codec based on a motion compensated 2D+t wavelet analysis. In order to code the spatio-temporal subbands, the first release used the EBCOT algorithm provided by the Verification Model JPEG2000. That release 1.0 has been registered at the Agency for the Protection of Programmes (APP) under the number IDDN.FR.001.160015.000.S.P.2003.000.20100 and then used by Thomson as part of a partnership. Wavix supports three forms of scalability: temporal via motion-compensated temporal wavelet transforms, spatial scalability supported via a spatial wavelet transforms and SNR scalability supported via a bit-plane encoding technique. The produced bitstream embedds the different levels of temporal and spatial resolutions as well as of quality. A so-called *extractor* allows to extract the portion of the bitstream to suit a particular receiver temporal and spatial resolution or to adapt the transmitted rate to the network characteristics. In 2005, a more elaborate motion encoding approach has been integrated in the codec. Error-resilient motion and texture decoding tools have been developped, integrated and validated in presence of different types of errors (random, error traces of 802.11 and 3GPP wireless and mobile networks). The benefit of these solutions used jointly with UDP-Lite has been assessed in terms of end-to-end throughput and SNR of the received and reconstructed signals. The decoder takes also advantage of information provided by the lower layers and in particular by UDP-Lite to increase the speed of the decoding process.

5.3. 3D Model-based video codec

Participants: Raphaèle Balter, Luce Morin [contact person].

From a video sequence of a static scene viewed by a monocular moving camera, this software allows 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. 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 software includes a coder and decoder of the scene representation. The codec allows to compress at low and very low bit-rates (16 to 256 kb/s in 25Hz CIF format) with a satisfactory visual quality. The new version developped in 2005 provides scalability for both geometry and texture information, by mean of second generation wavelets decomposition. Real-time visualization for interactive navigation including smooth transition between 3D models is also possible. The software has been registered at the Agency for the Protection of Programmes (APP) under the number IDDN.FR.001.13017.000.S.P.2003.000.41200.

5.4. Libit

Participants: Vivien Chappelier, Hervé Jégou [contact person].

Libit is a C library for information theory and signal processing. It extends the C language with vector, matrix, complex and function types, and provides some common source coding, channel coding and signal processing tools. The goal of libit is to provide easy to use yet efficient tools, and is mainly targeted at researchers and developpers in the fields of compression and communication. The syntax is purposedly close to that of other tools commonly used in these fields, such as MATLAB, octave, or IT++. Therefore, experiments and applications can be developped, ported and modified simply. Additional goals of the library include portability to many platforms and architecture, and ease of installation. Rather than trying to provide the very latest state-of-the-art techniques or a large panel of specific methods, this library aims at providing the most general and commonly used tools to build a communication chain, from signal processing and source coding to channel coding and transmission. Among these tools are some common source and channel models, modulation and quantization techniques, wavelet analysis, entropy coding, etc... As examples and to ensure the correctness of the algorithms with respect to published results, some test programs are also provided. (URL: http://libit.sourceforge.net).

5.5. Muxcodes library

Participant: Hervé Jégou [contact person].

This software is related to error-resilient entropy codes (so-called multiplexed codes) developed by the TEMICS project team. The software is composed of a set of functions for designing multiplexed codes according to a given probability distribution, encoding and decoding functions and of a test program for several realizations of encoded sources and of noisy channels. This software requires the Libit library (see above) and the GMP (GNU Multiple Precision) library. The library has been registered at the Agency for the Protection of Programmes (APP).

6. New Results

6.1. Analysis and modelling of video sequences

Keywords: *2D and 3D meshes*, *active contours*, *active meshes*, *augmented reality*, *dense fields*, *depth maps*, *disparity*, *hierarchical meshes*, *illumination models*, *mosaicking*, *motion*, *projective geometry*, *segmentation*, *tracking*, *triangulation*.

6.1.1. 3d Scene modelling from monocular video sequences

Participants: Raphaèle Balter, Luce Morin.

This research activity aims at developping a scalable 3D model-based video codec. Scalability is a key feature enabling content adaptation to varying network and terminals capabilities. In the approach developed, the video sequence is represented by a hierarchical and evolving 3D model leading to a stream (or sequence) of consistent 3D models. The sequence of 3D models is then analyzed with wavelet transforms of second generation. The consistency of the models is ensured by maintaining the conectivity constant but with a geometry evolving in time. The single connectivity mesh (SCM) allows a coherent wavelet decomposition of the sequence of models. The evolving model is also preserved at the different levels of the multiresolution representation of the geometry of the scene. To obtain an efficient and scalable compression of this scene representation, dependencies between texture, geometry and camera positions are exploited. Bitrate allocation to the mesh information (i.e. to the geometry of the scene) and to the wavelet or texture coefficients is optimized in a rate-distortion sense. The performance obtained in terms of PSNR versus bitrate as well as in terms of visual quality is higher than the one achieved with the H.264 standard. The coherence maintained between the successive models allows to improve the visual quality when reconstructing intermediate views along virtual navigation paths (see Fig.3). In addition, the approach has the advantage of being scalable.

6.1.2. 3d Scene modelling for distributed video compression

Participants: Christine Guillemot, Mathieu Maître, Luce Morin.

Distributed Source Coding (DSC) 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. cellphone 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 perfomance hit.

This research activity focusses 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

a

b

Figure 3. Reconstruction on the virtual path with (a)- a uniform grid and fading and with (b)- a non uniform evolving mesh.

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 modelled 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 scheme allows keyframes to be separated by more than 20 frames.

Moreover, such results led us to consider techniques beyond strict DSC, which would keep its valuable characteristics. By allowing points to be tracked between consecutive frames at the coder – thus introducing a limited temporal dependency between a keyframe and its following frames in the GOF – the keyframe frequency can be adapted to the video motion content. Indeed, keyframes only need to be sent when point track lengths exceed a given threshold, or when a given number of tracks have been lost. Moreover, these tracks allow the estimation of accurate camera 3D locations for all frames – not only the keyframes – which greatly improves the intermediate frame prediction. Finally, the coder complexity and added bitrate were kept in mind by limiting the number of tracked points to a minimum and using a naive tracking algorithm.

6.1.3. 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). The approach presented in the previous section 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, georeferenced modelling of city buildings. However, they are far less realistic than video captures, due to articificial textures and lack of geometric details.

The goal of this project is to use the complementarity of video and GIS data to provide a model with both realism, geometric details and global geometry consistency. The first steps of this study have been to acquire synchronized video and GPS data, and to use the GPS information as an initial estimate for camera localization in the GIS. A GIS data base for the University of Rennes 1 Campus has been provided by FT R&D, as a set of polygons specifying ground location of buildings, associated with an elevation value. The GPS data have been transformed from universal geographic coordinates into the local reference frame of the GIS. A graphic interface was designed to visualize the 3 types of acquired data (see figure 4).

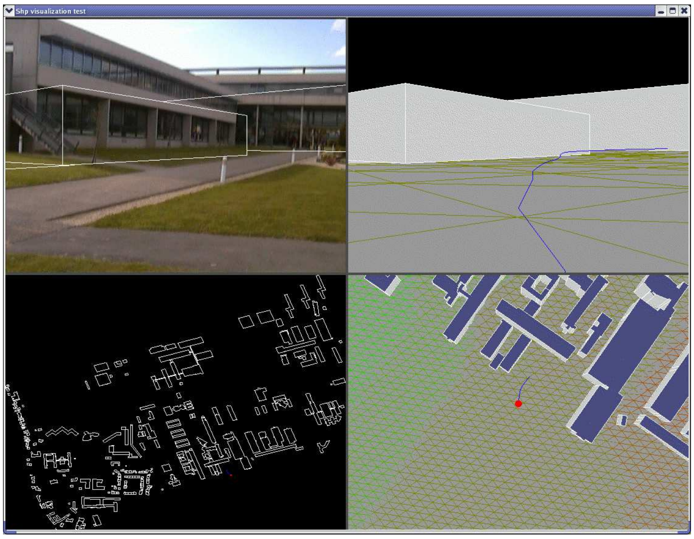

Figure 4. (a)- acquired video with superimposed GIS data; (b)- rendered GIS data with camera path in blue ; (c)- global map of GIS data, with camera path in blue and current camera position in red ; (d)- zoom on the GIS data and camera path around the current camera position.

6.2. Compression and distributed source coding

Keywords: *Bayesian estimation, Internet, compression, congestion control, error and erasure resilient coding and decoding, information theory, multiple description coding, overcomplete frame expansions, probabilistic inference, quality of service, rate-distortion theory, scalability, stochastic modelling, turbo principle, wavelet transforms, wireless communication.*

6.2.1. Scalable video compression

Participants: Vincent Bottreau, Christine Guillemot, Gagan Rath, Wenxian Yang.

Scalable video compression is essential for a range of emerging Multimedia applications. The ISO/MPEG standardization group has therefore issued a call for proposal initiating the specification phase of a scalable video coding standard (in MPEG-21/SVC). The objectives are to specify a coding/decoding solution allowing the generation of a unique bitstream embedding resolutions and rates going from QCIF at 7.5 Hz and 48 Kbits/s to standard definition at 30Hz and a few Mbits/s, with a number of intermediate operating points in terms of spatial and temporal resolutions and in terms of bit rates. In that context, we have worked on two specific problems:

- the study of spatial transforms based on the frame theory allowing to improve the rate-distortion performance of the multi-resolution or hierarchical representation of the signal;
- a bit-plane coding strategy to achive an efficient SNR scalable representation of the signal.

Multi-resolution or hierarchical representation of visual data such as image and video is an important concept. Its significance from the point of compression and transmission of image and video data over various communication media is well recognized. The central idea underlying such a representation is to build the signal gradually by adding higher resolution detail signals to a coarse low resolution signal in a hierarchical manner. Multi-resolution data representation is highly efficient for applications having varying bandwidth or scalability constraints. One of the examples of hierarchical image representation is the Laplacian pyramid.

The coarse resolution data is obtained by filtering and downsampling the upper layer high resolution signal. Then the higher resolution detail signal is obtained by taking the difference of the upper layer signal and the upsampled and filtered coarse signal used as a prediction. Unlike the well-known wavelet representation, the Laplacian pyramid is over-complete. Because of this property, the Laplacian pyramid has been of less interest to the compression community when wavelets were introduced. Recently they have recaptured the interest because of comparable rate-distortion performance with respect to wavelets in the context of scalable image and video compression.

The Laplacian pyramid is overcomplete. The redundancy is contained in the higher resolution detail layers. Therefore, the higher the hierarchy of the pyramid, the larger the redundancy. Because of the overcompleteness, the performance of the Laplacian pyramid can be analyzed using the frame theory. The construction of the pyramid can be formulated in terms of a frame operator constructed from the downsampling and upsampling filter coefficients. Therefore, under white quantization noise model, the reconstruction based on the dual frame operator should lead to minimum error. The frame operator, however, is easy to construct only if the filters satisfy at least the biorthogonality property. Further, the application of the dual frame operator is limited only to the decoding operation, and thus it does not change the bandwidth requirements of the pyramid.

To reduce the bandwidth requirements, we have developed several methods to eliminate partially or completely the redundancy in the signal representation generated by the LP. The corresponding signal reconstruction methods to be used on the decoder side have been designed. Redundant samples in the detail signals generated by the LP can indeed be dropped, yet be perfectly reconstructed at the decoder from the received samples. The singular value decomposition of the operator of the Laplacian Pyramid (LP) reveals a transform so that when applied on the LP subands the resulting detail layers contain exactly a critical number of nonzero samples, and thus we can dispense with the transmission of zero samples. Instead of eliminating the redundancy completely, we can also control it at different base layers. For example, we can allow the full redundancy at lower resolution, but eliminate it gradually as we go up the hierarchy. The goal is to achieve a better rate-distortion performance compared to that of the original Laplacian pyramid.

In the emerging standard, Fine Grain SNR (Quality) Scalability (FGS) is supported by encoding a refinement signal that corresponds to a bisection of the quantization step size. Transform coefficients are this way progressively refined by repeatedly decreasing the quantization step size and applying a modified CABAC entropy coding process akin to sub-bit-plane coding. We have developed an alternative approach making use of binarization codes adapted to the spatial transform coefficient statistical distributions. The scanning order of the binarisation tree is optimized in a rate-distortion sense. Having estimated for each transform coefficient subband the parameters (α, β) of the modeled generalized Gaussian distribution (GGD), one can design binarization codes that optimally, in a rate-distortion sense, map coefficient symbols into bin strings or codewords. To put it in a nutshell, such codes are built in such a way that every new transmitted bit gives the highest MSE decrease for a given rate constraint. The resulting codes have to be such that the symbol energy is mainly concentrated on the first bits of the symbol representation (i.e. on the first bit transitions of the corresponding codetree). A consequence of this is that the order according to which transform coefficient codewords are further parsed and processed is not necessarily the "classical" one, i.e. bit-plane by bit-plane for each coefficient at the same time, from the MSB to the LSB. These codes can be best decoded using an expectation-based decoding technique. The method avoids exhaustive quantization step search.

6.2.2. Oriented wavelet transforms for compression and de-noising

Participants: Vivien Chappelier, Christine Guillemot.

Wavelets are well-known mathematical tools for representing 1-D signals with a finite number of discontinuities with a small number of coefficients. However, for images modelled as homogeneous regions delimitated by contours, curve discontinuities are not fully captured by separable wavelets. In image compression applications, high energy coefficients cluster around the edges and most of the bitrate is spent to code the contours. Thus, new transforms (e.g., curvelets and contourlets) have been designed to better take into account - and capture - geometrical patterns present in images. Curvelets and contourlets are implemented with filter banks with directional selectivity in the high frequencies, so that the resulting coefficients represent oriented portions of edges instead of points. Their main advantage is that they do not require a geometric model of the image. The counterpart is that discrete implementations of curvelet transforms are currently highly redundant, which limits their interest for compression applications. The bandlets follow a different approach as they use a geometric model to describe the discontinuities of the image (parametrized curves or regularity flows) and wrap wavelets along these discontinuities. Though theoretically more efficient than curvelets for compression purposes, this approach is computationally intensive and its main problem lies in the optimization of the bitrate allocation between the image geometry description and the wavelet coefficients.

We have designed a new critically sampled transform based on wavelet lifting locally oriented according to multiresolution image geometry information. The lifting steps of a 1D wavelet are applied along a discrete set of local orientations defined on a quincunx sampling grid. To maximize energy compaction, the orientation minimizing the prediction error is chosen adaptively. The orientation is restricted to a binary information per wavelet coefficient (horizontal/vertical for even levels, diagonal/antidiagonal for odd levels) so as to minimize the orientation map coding cost. Perfect reconstruction is ensured by the reversibility of the lifting scheme. Each level of decomposition consists in splitting the sampling grid in two complementary quincunx cosets and applying the lifting steps along the chosen orientations. This oriented wavelet decomposition provides one high-pass subband and one low-pass subband on which the decomposition is iterated to obtain a multiscale analysis.

In a coding application, the multiresolution orientation map must be transmitted to the decoder so that the lifting steps can be inverted. In a first approach Markov random fields have been used to regularize the map and further reduce its coding cost. However, this model is very dense, hence its coding cost remains too high. A quad-tree structure has then been used to describe the geometry of the image leading to an efficient representation and a simple rate-distortion optimization which allows to balance the rate allocation between the map and the wavelet coefficients. As long as the map is sufficiently homogeneous, interesting properties of the original wavelet are preserved, such as regularity and quasi-orthogonality. The mutual information between the wavelet coefficients has been studied and compared to the one observed with a separable wavelet transform, evidencing better properties of signal energy compaction of the oriented transform. In order to assess the efficiency of this new transform in a real compression system, a context-based arithmetic encoder with contexts adapted to the transform properties has been developed. Gains in PSNR of up to 1 dB have been obtained compared to wavelets, for a range of bit rates. The performance of the transform has also been assessed for image de-noising. For image denoising, a Markov model is used to extract the orientations from the noisy image. A simple hard thresholding of the wavelet coefficiets can then be applied for denoising. The de-noising performance is illustrated in Fig.5.

6.2.3. *Multiresolution representations for compression and local description of images*

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 denoising applications. Yet, these critically-sampled transforms do not allow the extraction of low level signal features (points, edges, risdges, blobs) or of local descriptors. Many visual tasks such as segmentation, motion detection, object tracking and recognition, content-based image retrieval, require prior extraction of these low level features. The Gaussian scale space is almost the unique image representation used for this detection problem. Management of large databases are therefore uneasy, as the extraction of features requires first to decompress the whole database and then convert the images in the Gaussian scale space. It is thus desirable to find representations suitable for both problems: compression and signal feature extraction. However, their design criteria are somewhat antagonist. Feature extraction requires the image representation to be covariant under a set of admissible transformations, which ideally is the set of perspective transformations. Reducing this set of transformations to the group of isometries, and adding the constraint of causality, the image representation is uniquely characterized by the Gaussian scale space. In a compression perspective, one searches to reconstruct the image from a minimal amount of information, provided by quantized transform coefficients. Thus, the

Figure 5. Illustration of de-noising performance. [Left] The lena image with additive white Gaussian noise of standard deviation $\sigma = 25$ (PSNR = 20.2dB). [Right] The reconstructed image after denoising (PSNR = 30.6dB)

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.

In collaboration with the TexMex project team (Patrick Gros), we have characterized a set of multiresolution representations from the joint perspective of feature point and descriptor extraction and of compression. This analysis 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. The feature points are defined as salient points, i.e. points where a transient phenomenon can be modelled geometrically by providing a shape of saliency (corners, blobs, T-junctions) or energetically by defining a saliency measure in the scale space. It is shown that 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. This transform allows feature point extraction with a repeatability comparable to the one obtained with traditional extractors in the Gaussian scale space. At the same time, techniques such as projection on convex sets (POCS) can be used to reduce the coding cost resulting from the over-sampled signal representation. The assessment of the impact of quantization on the repeatability of feature points and descriptors extraction is on-going.

6.2.4. Distributed Source Coding

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

Distributed source coding (DSC) is a general framework which applies to highly correlated signals that are coded separately and decoded jointly. This framework applies to sensor networks but also to video compression. In the latter application, the motivation is to reduce the complexity of the encoder, at the expense of an increase of complexity of the decoder. From a theoretical point of view, DSC finds its foundations in the Slepian-Wolf theorem established in 1973. The Slepian-Wolf theorem states that for dependent binary sources Y and Z, the error decoding probability is close to zero for rates such that $R_Y \geq H(Y|Z)$,$R_Z \geq H(Z|Y)$,$R_Y + R_Z \geq H(Y,Z)$. This theorem has been extended to continuous-valued Gaussian sources by Wyner and Ziv in 1976. They have shown that for two correlated Gaussian sources Y and Z, if Z is available at the decoder, the rate-distortion performance obtained for the encoding of Y is the same whether the encoder knows the realization of Z or not. The statistical dependence between the two sources is modelled as a virtual correlation channel analogous to binary symmetric channels or additive white Gaussian noise (AWGN) channels. The source Y (called the side information) is thus regarded as a noisy version of X (called the main signal). Hence, most practical DSC solutions are based on channel codes (block, convolutional codes, turbo codes and LDPC codes).

One key aspect in the performance of the system is the mutual information between side information and the information being Wyner-Ziv encoded. We have first considered multiple sources. The corresponding rate-distortion bounds have been derived and Slepian-Wolf coding tools based on punctured turbo codes have been developed. The problem of quantization in Wyner-Ziv coding has been studied. In particular, a trellis coded quantizer accounting for the dependencies between two sources in a distributed coding set-up has been developed and its performance has been assessed. The development of a video distributed compression scheme is under progress. The focus has in particular been on the extraction of side information with appropriate correlation and mutual information with the information to be Wyner-Ziv encoded. Feature point extraction, tracking and 3D modelling tools are being used to reconstruct intermediate views considered as side information (see Section 6.1.2). Mesh-based motion compensated interpolation is another approach being investigated. Finally, we are working on the design of a coding solution accounting for the source memory.

6.3. Joint source-channel coding

6.3.1. Overcomplete frame expansions as joint source-channel codes for MIMO wireless channels

Participants: Robin Chatterjee, Christine Guillemot, Gagan Rath.

Overcomplete frame expansions have been introduced recently as signal representations that would be resilient to erasures in communication networks. Redundant block transforms such as those obtained from DFT, DCT, and DST matrices can be seen as producing discrete frame expansions in finite dimensional real or complex vector spaces whereas oversampled filter banks (OFB) can be seen as providing frame expansions in $l_2(Z)$. An OFB can be interpreted as a convolutional code over the real or complex field. With discrete frame expansions, the associated redundant transforms or, equivalently, the frame operators, can be interpreted as the generator matrices of some real or complex block codes. Therefore such frame expansions can be characterized based on the properties of the parity check matrices of the associated codes, such as the BCH structure. Various error localization, error and erasure correction algorithms with overcomplete frame expansions associated with low-pass DFT, DCT and DST codes have been designed in the period 2002-2004. The traditional BCH decoding or syndrome decoding approach is based on the concept of an error locator polynomial to localize the errors.

However, the frame expansion coefficients are quantized and encoded before being transmitted over a digital network, the error and erasure correction efficiencies are affected by the quantization noise. The quantized frame expansions can be looked upon as joint source-channel codes. The presence of quantization noise leads to difficulties beyond traditional syndrome decoding techniques. Observing some analogy between the DOA estimation problem in array signal processing and the error localization with quantized discrete frame expansions, we have developed in 2004 new decoding schemes based on subspace projection methods. The algorithms follow the eigendecomposition of syndrome covariance matrices, estimate the eigenvectors which span the error and the noise subspaces, and then estimate the error locations from the noise subspace eigenvectors.

In 2005, this study has been pursued in the context of MIMO (multiple-input multiple-output) wireless channels. Theoretical studies show that the capacity of a wireless channel can be increased by using multiple antennas at both the transmitter and the receiver. Using such a MIMO antenna system with multiplexing of data however may not result in high reliability. One of the ways to improve the reliability is to use diversity in space and time where redundant data is transmitted from multiple antennas over time. The design of redundant data streams to be transmitted from different transmitting antennas over time is the subject of so-called space-time coding. The design of a space-time code is primarily based on the minimization of the probability of transmitting a codeword and decoding a different codeword at the receiver (also called PEP - Pairwise Error Probability).

However, diversity in MIMO channels can be used to optimize the end-to-end distortion criterion of the communication chain. Codes can thus be designed to minimize the end-to-end distortion. This problem thus can be looked upon as a joint source-channel coding problem aimed for MIMO-based communication system. We have studied overcomplete frame expansions as joint source-channel codes for such channels. Consider a MIMO wireless communication system with n_t transmitting antennas and n_r receiving antennas. A source vector $\mathbf{x}$ consisting of K real-valued components is expanded into Kn_t components using an overcomplete frame expansion as $\mathbf{y} = F\mathbf{x}$, where F is a frame operator associated with a frame having Kn_t frame vectors in the K-dimensional space. The components of $\mathbf{y}$ are split into n_t vectors each having K components, and these vectors are transmitted from n_t transmitting antennas after being quantized and modulated. In this case, frame expansion builds redundancy into the system which is exploited as spatial and time diversity. The end-to-end distortion performance has been assessed both theoretically and experimentally by designing the corresponding ML and MAP decoders and turns out to be superior to the one obtained with space-time codes and linear dispersion codes. The mutual information between the channel input and output signals is shown to be increased and better approach the channel ergodic capacity, compared with space-time and linear dispersion codes.

6.3.2. Error-resilient source codes

Participants: Christine Guillemot, Hervé Jégou, Simon Malinowski.

In 2002-2004, we have designed new codes, that we called *multiplexed codes*, which have the property of avoiding the dramatic desynchronization problem of Variable Length Codes (VLC) used in multimedia compression systems, while still allowing to reach the entropy of the source [19], [17]. The idea underlying *multiplexed codes* builds upon the observation that most media compression systems generate sources of different priority. The design principle consists in creating fixed length codes (FLCs) for high priority information and in using the inherent redundancy to describe low priority data, hence the name “multiplexed codes”. The redundant set of FLCs is partitioned into equivalence classes according to high priority source statistics. The cardinality of each class, according to the high priority source statistics, is a key element so that the code leads to a description length as close as possible to the source entropy. Strategies for error resilient and progressive transmission of classical VLCs have also been designed [18]. These *bitstream construction* (BC) approaches allow to improve the error-resilience of the code, even when using simple hard decoding techniques. The performance can be further improved by using MAP, MPM or MMSE estimators. In contrast with solutions proposed so far in the litterature, the solutions designed have a linear complexity. The resulting bitstream structure is amenable to progressive decoding. The design of a progressive expectation-based decoding approach led to the introduction of code properties and design criteria for improved performance in terms of resilience and in a context of progressive decoding. In 2005, we have introduced new entropy source codes based on variable length production rules. A variable length code based on a binary codetree can be seen as a set of re-writing rules of the form

$$a_i s_1..s_m \rightarrow b_1..b_n,$$

where $(s_1..s_m) \in \{0,1\}^*$. This class of codes naturally extends codes based on binary codetrees and includes codes which can be regarded as finite state automata including quasi-arithmetic codes. This extension introduces additional degrees of freedom in the index assignment, allowing to improve the decoder resynchronization properties and the code performance in a context of soft decoding.

Methods for constructing codes having smaller mean description length than Huffman codes, or which preserve the lexicographic order of the source in the bit domain with at least the same compression efficiency than Huffman codes (hence over-performing the Hu-Tucker codes) have been designed. Another method allowing to construct codes such that the marginal probability is 0.5 has been developed. This property is shown to be beneficial when the codes are being used in a source-channel iterative structure making use of the turbo principle.

A new set of state models to be used in soft-decision (or trellis) decoding of variable length codes has also been introduced. 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. The complexity can be further reduced by running separate estimations on trellises of parameters T_1 and T_2 of the sequence of emitted symbols. We have shown that if T_1 and T_2 are relatively prime, and if the two corresponding sequence estimates $\hat{S}_1$ and $\hat{S}_2$ obtained with a Viterbi algorithm are such that $\hat{S}_1 = \hat{S}_2$, then, this estimate is the same as the one obtained with the trellis of parameter $T_1 \times T_2$.

The choice of the parameter T is related to the capability of the Variable Length Codes considered to resynchronize. We have thus studied the error recovery capability of VLCs with the help of a so-called error state diagram. Transfer functions defined on this error state diagram allow 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 VLC 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. We have also shown that the codes offering the best error resilience with soft decoding with contraint length are not those having the highest resynchronization probability (that is the highest $P(\Delta S)$).

6.3.3. Joint source-channel decoding of arithmetic codes

Participants: Christine Guillemot, Marion Jeanne, Florelle Pauchet.

H264 has appeared to be the most efficient video encoding standard with respect to compression. The introduction in H264 of a new entropy coding scheme named CABAC (Context-based adaptive binary arithmetic coding) is one of the key element to get high compression rates in H264. The CABAC algorithm is also used in the scalable video coding standard under specification within the joint video team between ITU and ISO/MPEG. The CABAC is an adaptive finite precision binary arithmetic coder composed of three basic components: binarization, context modeling and binary arithmetic coding.

The first step of the encoding process consists in converting the - possibly - M-ary source into a binary source. The second step estimates at run-time, the contextual statistical distributions of the resulting binary source. Very good compression efficiency can be achieved via the use of contexts capturing the statistical dependencies of the source. A reduced complexity binary arithmetic coder is then used. The CABAC suffers from a very high sensitivity to noise: a single bit error may cause the internal decoder state to be in error, with, in turn, a catastrophic effect on the reconstructed signal. The sensitivity to bit errors, and in particular the error propagation, results from several factors: the arithmetic coder memory, the context dependencies, as well as the on-line estimation of the source statistics made by the decoder from the received noisy bitstream.

The problem of sensitivity to transmission noise has not really been addressed in the standard, assuming either that error correction and/or detection would be handled in the lower layers of the communication protocol stack, or either recommending zero-order Exp-Golomb codes or Context Adaptive Variable Length Codes (CAVLC) for mobile devices. However, the loss in compression efficiency of Exp-Golomb/CAVLC methods with respect to the CABAC is in the order of 10% to 15% when coding TV signals at the same quality. We have investigated the viability of the CABAC algorithm for erroneous environments by developing a MAP sequential CABAC decoding algorithm. The sequential decoding technique has to account for the contextual modeling and the adaptive learning of the source statistics inherent to the CABAC algorithm. The algorithm has then been augmented with two error detection methods. Pruning techniques are also introduced in order to maintain the complexity within a tractable range. The pruning takes also advantage of the error detection mechanisms. The decoding and error detection solutions lead to a significant improvement compared to a hard-decision decoder CABAC, at some expenses of controlled and dynamically tunable redundancy, depending on the transmission conditions. The fine tuning of the redundancy represents an advantage compared to the use of channel codes. One cannot indeed find efficient channel codes for all desirable code rates. The solution outperforms the decoding of Exp-Golomb codes both in terms of compression efficiency and error resilience. The approach has been validated within an H.264 video coder and decoder and has been presented to ISO. The error detection mechanisms lead to the introduction of variants of the CABAC algorithm which could be beneficial to the standard under the form of profiles dedicated to mobile and wireless applications, or to applications requiring some resilience to errors (e.g., degradations of physical storage media).

6.3.4. Cross-layer design for wireless video communication

Participant: Ceilidh Hoffman.

The purpose of this joint project between the PLANETE and the TEMICS project teams is to propose various techniques to improve the quality of video delivery over wireless links. In general, a video packet originating from a server must be relayed across several communication channels before it arrives at its destination —the wireless end user. For the delivery of certain video applications that do not require real-time communications, it is possible to set up a mirror server that is only required to deliver packets over the last mile —the wireless channel— and thus bypassing intermediate layers such as transport, network, and data link layer protocols. This is the conventional approach that is well studied in the literature in which the wireless channel is modeled as a binary or multi-level discrete input-output sets, or as a continuous additive Gaussian channel with extensions to support multipath fading. In this classical model, the goal of cross-layer optimization is to design video encoders and decoders with a specified fidelity target over a link whose long-term time-averaged information rate is bounded by its Shannon channel capacity.

In our work we emphasize the notion of a wireless “network,” which is a combination of both data link (in particular, the medium access control MAC) and physical layer entities whose functionalities must be jointly optimized within the constraints of technical standard specifications. Examples of such wireless networks are wide-area cellular systems (2G and 3G networks), local-area wireless networks (IEEE 802.11 and Wi-Fi networks) and personal area wireless networks (IEEE 802.15 and Wi-Max networks). In our case special attention is given to 802.11 networks whose MAC protocol is distributed and bottomless; that is, it imposes no restriction on the number of active numbers in a system/cell. Since all users must share a common wireless link for both up- and down-link communications, each user’s peak-to-average rate —or equivalently, its inter-packet delay— is affected by the size and duty cycle of other users. Therefore, the throughput achieved by each user is substantially different from its channel capacity. It is no longer sufficient to characterize a link quality in terms of time-averaged channel capacity; we must also measure performance in terms of delay-limited capacity and outage probability.

We show that the MAC protocol of 802.11 based on distributed medium sharing is fair in guaranteeing the same number of channel access to every user over the long term. An apparent unfair behavior arises when a user’s packet delivery success rate is less than one. We also show that the disparity of respective throughputs among users is due to rate inequality, which is a physical layer condition that is independent of the corresponding MAC protocol. By subdividing the medium access time scale into contention and capture intervals, we are able to measure the individual as well as aggregate throughput and delay statistics of the system at the medium access control, physical and overall network levels. We also describe a rate adaptation algorithm that is implemented at the physical layer to improve the packet delivery success rate of each user. Finally, we propose several external measures —cell size reduction and balancing of channel access times— to improve the throughput and delay imbalance among users. Our approach ensures a minimum level of absolute QoS guarantee for video delivery to all wireless users in the system regardless of their relative locations (with respect to the access point) and physical layer channel conditions.

6.4. Image and video watermarking: Optimal watermark detection

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

Participants: Caroline Fontaine, Teddy Furon.

The rediscovery of the paper of M. Costa gave birth in 2001 to an impressive effort on developing dirty paper codes. Let $\mathbf{x}$, a vector of length n be the host signal. These codes hide a message $m \in \mathcal{M} = \{1, ..., 2^{Rn}\}$ in the host signal to produce a watermarked signal $\mathbf{y}$, such that the embedding distortion is contrained by $E\|\mathbf{y} - \mathbf{x}\|^2 < D$. $R = \log_2(|\mathcal{M}|)/n$ is defined as the rate of the watermarking scheme. In other words, the number of messages $|\mathcal{M}|$ which can be reliably transmitted is exponential in the length of the host vector. The community knows now how to construct codes whose rate R achieve capacity asymptotically ($n \to +\infty$), notably using lattice modulo channels .

These last months have seen a change of interest of the watermarking community to zero-rate codes. The problem is now to transmit a fixed and low number of bits in long length host signals in such a way that the probability of errors exponentially vanishes when $n \to +\infty$. It seems that highly robust watermarking with small number of messages is more suitable for practical applications than the functionality of embedding an exponential amount of information. In other words, rate R tends to zero and the theory of Costa does not apply anymore. The community looks whether dirty paper schemes invented for non zero rate are still optimal for this problem.

In this new trend for zero-rate codes, we have focused our effort on detection. This issue seems to be a subclass of the above problem, but indeed, it is quite different from a theoretical point of view. Receiving a vector $\mathbf{z}$, the detector has to decide between two hypothesis: either this vector has not been watermarked (Hypothesis H_0) and the detection output should be $d = 0$, either it has been watermarked and possibly attacked (Hypothesis H_1) and the detection output should be $d = 1$. The performances of the test are measured by the probability of false alarm P_{fa}= Prob$(d = 1|H_0)$ and the power of the test P_p= Prob$(d = 1|H_1)$. The term 'watermark detection' means that no message is hidden in the host signal. Some call this a zero-bit watermarking scheme. The reader must understand here that we are only interested in hiding and detecting the presence of a watermark.

For instance, in a copy protection application, a DRM system controls the access and the copy management of digital multimedia content using cryptographic primitives (encryption, signatures,...). However, due to its multimedia aspect, a pirate can digitalize the analog rendering of the host (audio, picture, video) to have a pirated content in the clear where all the DRM cryptographic protection are gone. At this point, nothing distinguishes the pirated content from an innocuous clear content as an user's vacation movie. This problem is known as the 'analog hole'. The embedding of a watermark is a potential solution. The detection of its presence in a clear content would warn the compliant device that it is indeed a forged content. The compliant device will then refuse to play and/or record it.

6.4.1. *Theoretical bound for side informed embedding*

Some works have already been done in watermarking detection where the embedder hides a reference signal $\mathbf{w}$ in the host. The detector knows this reference signal and it tests the presence of $\mathbf{w}$ in the received signal $\mathbf{z}$.

Because the embedding distortion is very small compared to the energy of the host, this resorts to an extremely classical problem in test hypothesis known as ‘detection of weak signals’.

Inspired by the theory of Costa, we know that better performances are achieved when the embedder is side informed, ie. when the watermark signal is no more a constant reference signal but it is a function of the host: $\mathbf{w} = \mathbf{w}(\mathbf{x})$. We have first derived the theoretical bounds of watermarking detection with side information. The performances (P_{fa}, P_p) of the test are limited according to the theorem of data processing: $D_{KL}(P_{fa}; P_p) \leq D_{KL}(p(\mathbf{z}|H_0); p(\mathbf{z}|H_1))$ where $D_{KL}(.;.)$ is the Kullbach-Leibler distance (aka discrimination, relative entropy), and $p(\mathbf{z}|H_i)$ is the probability density function of vector $\mathbf{z}$ under hypothesis H_i. When the watermark signal is constant and under some simple hypothesis such as whitness and gaussianity, a classical result is that $D_{KL}(p(\mathbf{z}|H_0); p(\mathbf{z}|H_1))$ is bounded by $nP/2(Q+N)$, with P, Q and N respectively the powers of $\mathbf{w}, \mathbf{x}$ and $\mathbf{n}$. This implies poor performances as the power of the host Q is usually very big compared to P.

We have shown that when side information is enabled, ie. $\mathbf{w} = \mathbf{w}(\mathbf{x})$, then the distance $D_{KL}(p(\mathbf{z}|H_0); p(\mathbf{z}|H_1))$ is indeed bounded by $nP/2N$. The power of the host signal Q does no more limit the bound of the test. We can expect far better performances, just as if the detector were knowing $\mathbf{x}$, although this data is only known at the embedding. This amazing fact is known since M. Costa work on decoding, however, we have extended it to watermark detection. This shows that the classical tools of ‘detection of weak signals’ are not optimal, and consequently watermark detection deserves its own theory where the key idea is the side informed embedding.

6.4.2. Practical embedding / detection functions

Knowing this theoretical bound, we have now to propose some schemes whose performances are as close as possible from the bound but also which are practical for real life application. In other words, we have to give a couple of matched embedding function and detection function $(\mathbf{w}(\mathbf{x}), d(\mathbf{z}))$. Side informed watermarking has been proposed recently with embedder based on quantization. Whereas these solutions greatly perform against noise $\mathbf{n}$ addition, they quickly fail when a scaling factor ρ is considered in the attack: $\mathbf{z} = \rho(\mathbf{x} + \mathbf{w}(\mathbf{x}) + \mathbf{n})$. We prefer to consider smooth continuous embedding function $\mathbf{w}(\mathbf{x})$. On the detection side, we restrict our search to a well known class of test: Neyman-Pearson test with a locally most powerful detection function. Given an embedding function, this gives us the best detection function in this class. On the embedding side, we assume that the power of the watermark signal is very small compared to host power. This allows a first order Taylor development of the power of the test. Given a detection function $d = d(\mathbf{z})$, this yields the best embedding function maximizing this first order approximation of the test. Assembling these two last results, we get a differential equation whose solutions are optimal among the restricted class considered so far.

In summary, we have found some families of optimal pairs $(\mathbf{w}(\mathbf{x}), d(\mathbf{z}))$. Some were already heuristically discovered, we have theoretically justified their use. Others are really new and they achieve efficiency never met before in watermarking detection.

7. Contracts and Grants with Industry

7.1. Industrial contracts

TEMICS has three Cifre contracts with industrial partners:

- Cifre contract with France Telecom RD in the context of the Ph.D of Raphaele Balter in the area of 3D-model based coding of video sequences. The results achieved in 2005 are described in subsection 6.1.1.
- CRE contract with France Telecom R&D (started in October 2004) in the context of the Ph.D of Gaël Sourimant on the area of 3D reconstruction of urban scenes by fusion of GPS, GIS and video data.

- CRE contract with France Telecom R&D (started in November 2004) on the problem of distributed source coding. The objective is to investigate this new coding paradigm and assess its potential for compression with mobile light-weight encoding systems.

TEMICS also supervises DRT projects in collaboration with industrial partners:

- H. Nicolas supervises the Degree of Technological Research ("Diplôme de Recherche Technologique" (DRT)) of Lila Huguenel realized at Thomson on the subject "MPEG-4 AVC video compression based on regions of interest".
- L. Morin supervises the Degree of Technological Research ("Diplôme de Recherche Technologique" (DRT)) of Olivier Gaborieau realized at Thomson on the subject "Rate allocation algorithm based on the ρ-domain in an H.264/AVC compatible video codec".

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 : Mar.04 - Feb.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.

7.3. European contracts

7.3.1. FP6-IST NoE SIMILAR

Participants: Christine Guillemot, Khaled Lajnef, Luce Morin, Henri Nicolas, 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, § 6.2.4
- 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.

- Shadow and light source detection and analysis techniques have been developed. They are used to artificially create cast shadows of natural objects inserted in a video sequence. These methods will be applied to the mixing of 2-D (original and mosaic images, video objects) and 3-D (synthetic objects and 3-D model) video data.
- TEMICS is contributing on a distributed coding framework in a context of multimodality and coordinating with EPFL a working group dealing with the development of an information theoretic framework for analysis and representation of multimodalities.

7.3.2. FP6-IST STREP DANAE

Participants: Vincent Bottreau, François Cayre, Christine Guillemot, Gagan Rath.

- Convention number: 104C045731324005
- Title: *Dynamic and distributed Adaptation of scalable multimedia content in a context-Aware Environment*.
- Research axis: § 6.2.1, § 6.3.2
- 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 DANAE 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 DANAE 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.

7.3.3. FP6-IST STREP DISCOVER

Participants: Christine Guillemot, Denis Kubasov, Gagan Rath.

- Convention number:
- 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

Participants: François Cayre, Teddy Furon.

- Convention number: 103C17280031324011
- Title : Fabriano
- Research axis : § 6.4.1, § 6.4.2.
- Partners : CERDI, INRIA (TEMICS), LIS, LSS.
- Funding : Ministry of research, CNRS, INRIA.
- Period : Mid-Dec. 03 - Dec. 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 substitution and spread spectrum based schemes.

8.1.2. ACI Masse de données - Codage

Participants: Vivien Chappelier, Olivier Crave, Christine Guillemot, Hervé Jégou.

- Convention number: 104C07410031324011
- Title : Codage de masse de données
- Research axis : § 6.2.2, § 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.

9. Dissemination

9.1. Patents

The following patents have been transferred to Thomson:

- T. Guionnet and C. Guillemot, "Method for robust decoding of arithmetic codes", Inria patent No. 03 03288, 2003.
- H. Jégou and C. Guillemot "Robust Source Codes ", patent INRIA/ENS, 02 09287, 2002.
- H. Jégou and C. Guillemot "Compression of digital data robust to transmission noise ", patent INRIA/ENS, 02 14964, 2002.

9.2. Standardization

- V. Bottreau, C. Guillemot, H. Jégou, JVT-O046: “Improved FGS scheme for SVC”, Joint Video Team (JVT) of ISO/IEC JTC1/SC29/WG11(MPEG) and ITU-T SG16 Q.6(VCEG), Busan, March 2005.
- M. Jeanne, C. Guillemot, F. Pauchet, V. Bottreau, “CABAC variants and options for error-resilience”, Joint Video Team (JVT) of ISO/IEC JTC1/SC29/WG11(MPEG) and ITU-T SG16 Q.6(VCEG), JVT-0080, Busan, March 2005.
- N. Cammas, S. Pateux, L. Morin, "Analysis-synthesis scheme using active meshes for wavelet video coding", MPEG Workshop on Future Directions in Video Coding, Nice, France, Oct. 2005.

9.3. Invited talks and demonstrations

- Raphaèle Balter, “Streaming de séquences reconstruites en 3D Journées d’études “Des images au 3D”, seminar organized by SEE/Club 29, in partnership with GDR ISIS and SFPT, June 16-17, 2005, Paris. (http://www.see.asso.fr/htdocs/main.php/congresJourneesPassees.php/348/ http://gdr-isis.org/rilk/gdr/ReunionListe?r=382).
- For the 30th aniversary of IRISA, the TEMICS project team shows two demos: robust and fragile watermarking of still images, 3D compression and navigation. For the watermarking part, four plug-ins for the open-source image processing software GIMP have been developed. The first plug-in inserts a robust watermark. It asks for the message to be hidden and the secret key. It shows the watermarked picture and where the watermark is located in the host (ie., in the textured areas and along the contours). The second one decodes hidden messages given a secret key. The third plug-in adds a fragile watermark to a picture. The fourth verifies the authenticity of a watermarked image. If the verification fails, it graphically shows the areas in the image suspected of forgery.

9.4. Leadership within the scientific community

- T. Furon is member of the program committees of Int. Workshop on Digital Watermarking (IWDW) 2005, Electronic Imaging Security, Steganography, and Watermarking of Multimedia Content conference 2005, and GRETSI (Colloque Traitement Signal et Image) 2005;
- T. Furon has co-chaired the special session on Watermarking security at IWDW 2005.
- C. Guillemot is associate editor of the journal IEEE Transactions on Circuit and System for Video Technology;
- C. Guillemot is elected member of the international committee IEEE IMDSP (Image and MultiDimensional Signal Processing Technical Committee);
- C. Guillemot is elected member of the international committee IEEE MMSP (MultiMedia Signal Processing Technical Committee);
- C. Guillemot is member of the external scientific advisory board of the IST-FP6 Network of Excellence VISNET;
- C. Guillemot has been nominated by the ministry of research as a french representative within the management committee of COST (Action COST 292 "Sematic and multimodal analysis of digital media" ;
- C. Guillemot is member of program committees of the following conferences: IEEE-ICIP 2005, WIAMIS 2005, CORESA 2005;
- C. Guillemot is member of the steering committee of the FP6-IST Network of Excellence SIMILAR;
- C. Guillemot has served as expert for evaluating project proposals for the national RNRT and RIAM programmes;
- H. Nicolas is member of the program committees of IEEE-ICIP 2005;
- H. Nicolas is member of the "commission de spécialistes" of the University of Rennes 1.

9.5. Invitations

- C. Guillemot has been invited at the University Catholique de Louvain (UCL);
- L. Morin has been as invited professor for a one week visit at the UNAM University of Mexico; she has given a short course (10 hours) on “3D modelling from images and application to video compression”.

9.6. Teaching

- Master of Multimedia Network Security, Telecom Paris (F. Cayre: Watermarking attacks: robustness and security);
- Supelec (T. Furon, conference on digital watermarking);
- Engineer degree Diic-INC, Ifsic, university of Rennes 1 (L. Morin, C. Guillemot, L. Guillo, H. Jegou, T. Furon : image processing, 3D vision, motion, coding, compression, cryptography, communication) ;
- Engineer degree Diic-LSI, Ifsic, university of Rennes 1 (L. Morin, C. Guillemot, Gaël Sourimant : compression) ;
- Master research STI, university of Rennes 1 (C. Labit, H. Nicolas : compression) ;
- ESIGETEL Fontainebleau, (H. Nicolas : Video compression and communication) ;
- Enic, Villeneuve-d'Ascq, (C. Guillemot: Video communication) ;
- INSA, Lyon, (C. Guillemot: Video communication) ;
- Professional degree Tais-Cian, Breton Digital Campus (F. Cayre, L. Morin : Digital Images) ;
- Master, Network Engineering, university of Rennes I (L. Guillo, Video Streaming) ;

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] 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.

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

[7] 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.

[8] G. RATH, C. GUILLEMOT. *Frame-theoretic analysis of DFT codes with erasures*, in "IEEE Transactions on Signal Processing", vol. 52, n° 2, Feb. 2004, p. 447-460.

[9] P. RAULT, C. GUILLEMOT. *Symmetric delay factorization: generalized framework for paraunitary filter banks*, in "IEEE Trans. on signal processing", vol. 47, n° 12, Dec. 1999, p. 3315-3325.

Doctoral dissertations and Habilitation theses

[10] R. BALTER. *Construction d'un maillage 3D évolutif et scalable pour le codage vidéo*, Ph. D. Thesis, Thèse de doctorat en Traitement du Signal, Université de Rennes 1, France, May 2005, http://www.irisa.fr/bibli/publi/theses/2005/balter/balter.html.

[11] V. CHAPPELIER. *Codage progressif d'images par ondelettes orientées*, Ph. D. Thesis, Thèse de doctorat en Traitement du Signal, Université de Rennes 1, France, Nov. 2005.

[12] H. JÉGOU. *Robust source and joint source/channel codes for multimedia transmission over wireless channels / Codes sources robustes et codes joints source-canal pour transmission multimédia sur réseaux sans fils*, Ph. D. Thesis, Thèse de doctorat en Traitement du Signal, Université de Rennes 1, France, Nov. 2005.

[13] L. MORIN. *Modélisation 3D pour la communication vidéo*, Ph. D. Thesis, Habilitation à diriger les recherches, Université de Rennes 1, France, Dec. 2005.

Articles in refereed journals and book chapters

[14] F. CAYRE, C. FONTAINE, T. FURON. *Watermarking Security: Theory and Practice*, in "IEEE Trans. Signal Processing", vol. to appear in Special Number on Information Forensics, October 2005, 2005.

[15] T. FURON. *Fundamentals of Digital Image Watermarking*, chap. Watermarking techniques for alternative requirements, Wiley, 2005, to appear.

[16] C. GUILLEMOT, P. SIOHAN. *Joint source-channel decoding with soft information: a survey*, in "Elsevier Journal on Applied Signal Processing, special issue on the turbo principle", vol. 6, June 2005, p. 906-927.

[17] H. JÉGOU, C. GUILLEMOT. *Error-resilient first-order multiplexed codes : performance bounds, design and decoding algorithms*, in "accepted for publication in IEEE Trans. on Signal Processing", 2005.

[18] H. JÉGOU, C. GUILLEMOT. *Progressive and error resilient transmission strategies for VLC encoded signals over noisy channels*, in "accepted for publication in Eurasip journal on applied signal processing, special issue on video analysis and coding for robust transmission (minor revisions)", 2005.

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

[20] G. LE GUELVOUIT, T. FURON. *La stéganographie moderne*, in "Multi-System & Internet Security Cookbook (MISC)", vol. 18, mar 2005, p. 26-31.

[21] M. MANSHAEI, T. TURLETTI, T. GUIONNET. *An Evaluation of Media-Oriented Rate Selection Algorithm*

for Multimedia Transmission in MANETs, in "Eurasip Journal on Applied Signal Processing, Special Issue on Ad Hoc Networks, to appear", 2005.

[22] S. MARINKOVIC, C. GUILLEMOT. *Joint source-channel coding by means of over-sampled filter bank codes*, in "Eurasip journal on applied signal processing, to appear", 2005.

[23] S. MARINKOVIC, C. GUILLEMOT. *Over-sampled filter banks for erasure resilient signal transmission*, in "Eurasip journal on applied signal processing", vol. 4, 2005, p. 510-524.

[24] H. NICOLAS, F. DENOUAL. *Semi-automatic modifications of video object trajectories for video compositing applications.*, in "To be published in Eurasip Signal Processing Journal.", 2005.

[25] H. NICOLAS. *Shadow synthesis for video post-production applications*, in "IEEE Transactions on Signal Processing Letters", vol. 12, n° 4, April 2005.

[26] H. NICOLAS, J. M. PINEL. *Joint moving cast shadows segmentation and light soource detection in video sequences*, in "To be published in Eurasip Image Communication Journal.", 2005.

[27] G. RATH, C. GUILLEMOT. *Characterization of a class of error correcting frames for robust signal transmission over wireless communication channels*, in "Eurasip journal on applied signal processing, special issue on design of cross-layer communication systems", Feb. 2005, p. 229-241.

Publications in Conferences and Workshops

[28] R. BALTER, P. GIOIA, L. MORIN. *Time evolving 3D model representation for scalable video coding*, in "Proceedings of the IEEE Conference on Image Processing, Genoa, Italy", september 2005, ftp://ftp.irisa.fr/local/temics/temics/publis/2005/balter05b.pdf.

[29] G. BOISSON, E. FRANÇOIS, C. GUILLEMOT. *TWAVIX : une solution basée-ondelettes (t+2D) pour le codage vidéo scalable*, in "Proc. CORESA", Nov. 2005.

[30] N. CAMMAS, S. PATEUX, L. MORIN. *Video coding using non-manifold mesh*, in "Proceedings of the 13th European Signal Processing Conference, Antalya, Turkey", Sept. 2005.

[31] N. CAMMAS, S. PATEUX, L. MORIN, B. LE GUEN. *Estimation de mouvement par maillages et lignes de rupture pour le codage de séquences vidéo*, in "20th GRETSI Symposium on Signal and Image Procesing, Louvain-la-Neuve, Belgique", Sept. 2005.

[32] F. CAYRE, C. FONTAINE, T. FURON. *Watermarking Security Part II: Practice*, in "Proc. of SPIE-IS&T Electronic Imaging, SPIE, San Jose, CA, USA", E. J. DELP, P. W. WONG (editors). , vol. 5681, Security, Steganography, and Watermarking of Multimedia Contents VII, jan 2005, p. 758-768.

[33] F. CAYRE, C. FONTAINE, T. FURON. *Watermarking Security Part I: Theory*, in "Proc. SPIE-IS&T Electronic Imaging, SPIE, San Jose, CA, USA", E. J. DELP, P. W. WONG (editors). , vol. 5681, Security, Steganography, and Watermarking of Multimedia Contents VII, jan 2005, p. 746-757.

[34] V. CHAPPELIER, C. GUILLEMOT. *Oriented 1D wavelet transform on a quincunx pyramid for image compression*, in "Proc. IEEE Int. conf. on Image Processing, ICIP", Sept. 2005.

[35] T. FURON. *A survey of watermarking security*, in "Proc. of Int. Work. on Digital Watermarking, Sienna, Italy", M. BARNI (editor). , Lecture Notes on Computer Science, vol. 3710, Springer-Verlag, sep 2005, p. 201-215.

[36] H. JÉGOU, C. GUILLEMOT. *Codage entropique à base de règles de ré-écriture*, in "Proc. GRETSI", Sept. 2005.

[37] H. JÉGOU, C. GUILLEMOT. *Entropy coding with variable length rewriting systems*, in "Proc. IEEE International symposium on information theory, ISIT", Sept. 2005.

[38] H. JÉGOU, S. MALINOWSKI, C. GUILLEMOT. *Trellis state aggregation for soft decoding of variable length codes*, in "Proc. IEEE workshop on signal processing systems, SIPS", Nov. 2005.

[39] H. NICOLAS. *Scalable video compression scheme for telesurveillance applications based on cast shadow detection and modelling*, in "Proceedings of ICIP, Genoa, Italy", September 2005.

[40] H. NICOLAS. *Shadow synthesis in natural sequences using 3-D scene model*, in "Proceedings of VCIP, Beijing, China", July 2005.

[41] G. SOURIMANT, L. MORIN. *Codage vidé basé sur des modèles génériques 2D et 3D*, in "Proceedings of Orasis 2005, Fournols, France", May 2005.

[42] G. SOURIMANT, L. MORIN. *Model-Based Video Coding with Generic 2D and 3D Models*, in "Proceedings of Mirage 2005 (Computer Vision / Computer Graphics Collaboration Techniques and Applications), INRIA Rocquencourt, France", March 2005, ftp://ftp.irisa.fr/local/temics/temics/publis/2005/sourimant05a.pdf.

[43] F. TONNIN, P. GROS, C. GUILLEMOT. *Analyse de représentations multirésolution pour la compression et la description locale d'images*, in "Proc. CORESA", Nov. 2005.

[44] F. TONNIN, P. GROS, C. GUILLEMOT. *Analysis of multiresolution representations for compression and local description of images*, in "Proc. 8th International Conference on VISual Information Systems", Jul. 2005.

[45] F. TONNIN, P. GROS, C. GUILLEMOT. *Les representations steerables pour la compression et l'extraction d'attributs de bas-niveau*, in "Proc. GRETSI", Sept. 2005.