# Activity Report 2011

# **Project-Team TEMICS**

# Digital image processing, modeling and communication

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

# Table of contents

# Project-Team TEMICS

**Keywords:** Image Processing, Signal Processing, Video

# 1. Members

**Research Scientists**

Christine Guillemot [Team Leader, DR, Inria, HdR]
Claude Labit [DR, INRIA, 20%, HdR]
Aline Roumy [CR, INRIA]

**Faculty Members**

Jean-Jacques Fuchs [Professor, SPM, University of Rennes 1, HdR]
Olivier Le Meur [MdC, IFSIC, University of Rennes 1]

**Technical Staff**

Laurent Guillo [IR, CNRS]

**PhD Students**

Ana Charpentier [INRIA grant, MEDIEVALS project, until April 2011]
Marco Bevilacqua [INRIA grant, Joint INRIA/Alcatel Lucent lab.]
Jeremy Aghaei [INRIA grant, contract with Astrium, since 3rd Oct. 2011]
Safa Cherigui [CIFRE contract with Technicolor]
Mouid Keskes [CIFRE contract with Orange, since 1st Dec. 2011]
Daria Khaustova [CIFRE contract with Orange, since 1st Dec. 2011]
Mounira Ebdelli [INRIA grant, ANR-ARSSO project]
Josselin Gauthier [Ministry of research grant]
Vincent Jantet [ENS grant, until 30th Sept. 2011]
Jonathan Taquet [Inria-Region ARED grant, until 31st Oct. 2011]
Mehmet Turkan [INRIA grant, CORDIS, until 31st Oct. 2011]

**Post-Doctoral Fellow**

Raul Martinez [INRIA, since 1st July 2011]

**Administrative Assistant**

Huguette Béchu [TR, INRIA, shared with Fluminance and Serpico]

**Others**

Ronan Boitard [INRIA, until 15th Sept. 2011]
Thomas Guionnet [INRIA, since 3rd Oct. 2011]
Velotiaray Toto-Zarasoa [INRIA, until 31st May 2011]

# 2. Overall Objectives

## 2.1. Introduction

The goal of the TEMICS project-team has been over its past 12 years of existence the design and development of algorithms and practical solutions in the areas of analysis, modelling, coding, communication and watermarking of images and video signals. In 2011, the watermarking activities have been stopped in order to re-focus on visual data analysis, modeling, representation, compression and communication. These areas of research will indeed be the focus of the successor project-team called SIROCCO. 2011 being a transition period between TEMICS and SIROCCO, the project-team activities have thus been structured and organized around the following inter-dependent research axes:

- Analysis and modeling for compact representation and navigation [1] in large volumes of visual data[2].
- Representation and compression of visual data
- Distributed processing and robust communication of visual data

Given the strong impact of standardization in the sector of networked multimedia, TEMICS, in partnership with industrial companies, seeks to promote its results in standardization (JPEG, MPEG). While aiming at generic approaches, some of the solutions developed are applied to practical problems in partnership with industry (Alcatel Lucent, Astrium, Orange labs., Technicolor) or in the framework of national projects (ANR-ARSSO, ANR-PERSEE). The application domains addressed by the project are networked visual applications via their various requirements and needs in terms of compression, of resilience to channel noise, or of advanced functionalities such as navigation.

## 2.2. Analysis and modeling for compact representation

Analysis and modeling of the visual data are crucial steps for a number of video processing problems: navigation in the 3D scene, compression, loss concealment, denoising, inpainting, editing, content summarization and navigation. The focus is on the extraction of different cues such as scene geometry, edge, texture and motion, on the extraction of high-level features (GIST-like or epitomes), and on the study of computational models of visual attention, useful for different visual processing tasks.

In relation to the above problems, the project-team places a special focus on 3D and multi-view content. 3D displays, going from stereo solutions with glasses to multi-view auto-stereoscopic displays which do not require glasses, are starting to appear for home environments. Depth perception on such displays requires transmitting two or more views, i.e. very large volumes of data. The TEMICS project-team focuses on several algorithmic problems to analyze, represent, compress and render multi-view video content. The team first addresses the problem of depth information extraction. The depth information is associated with each view as a depth map, and transmitted in order to perform virtual view generation. The design of algorithmic solutions for the end-to-end capturing, compression, transmission and rendering of 3D and multi-view content actually raises a number of unresolved theoretical and practical issues related to depth or scene geometry estimation, analysis of multi-view video, efficient compression of video plus depth data, and high quality rendering (view generation process). Given one view with its depth information, depth image-based rendering techniques have the ability to render views in any other spatial positions. However, the issue of intermediate view reconstruction remains a difficult ill-posed problem. Most errors in the view synthesis are caused by incorrect geometry information, inaccurate camera parameters, and occlusions/disocclusions. efficient inpainting techniques are necessary to restore disocclusions areas.

## 2.3. Representation and compression of visual data

The objective is to develop algorithmic tools for constructing low-dimensional representations of multi-view video plus depth data, of 2D image and video data, of visual features and of their descriptors. Our approach goes from the design of specific algorithmic tools to the development of complete compression algorithms. The algorithmic problems that we address include data dimensionality reduction, the design of compact representations for multi-view plus depth video content which allow high quality 3D rendering, the design of sparse representation methods and of dictionary learning techniques. The sparsity of the representation indeed depends on how well the dictionary is adapted to the data at hand. The problem of dictionary learning for data-adaptive representations, that goes beyond the concatenation of a few traditional bases, has thus become a key issue for further progress in this area.

---

[1] By navigation we refer here to scene navigation by virtual view rendering, and to navigation across slices in volumic medical images
[2] By visual data we refer to natural and medical images, videos, multi-view sequences as well as to visual cues or features extracted from video content

Developing complete compression algorithms necessarily requires tackling visual processing topics beyond the issues of sparse data representation and dimensionality reduction. For example, problems of scalable, perceptual, and metadata-aided coding of 2D and 3D visual data, as well as of near lossless compression of medical image modalities (CT, MRI, virtual microscopy imaging) are tackled. Finally, methods for constructing rate-efficient feature digests allowing processing in lower-dimensional spaces, e.g. under stringent bandwidth constraints, also falls within the scope of this research axis.

## 2.4. Distributed processing and robust communication

The goal is to develop theoretical and practical solutions for robust image and video transmission over heterogeneous and time-varying networks. The first objective is to construct coding tools that can adapt to heterogeneous networks. This includes the design of (i) sensing modules to measure network characteristics, of (ii) robust coding techniques and of (iii) error concealment methods for compensating for missing data at the decoder when erasures occur during the transmission. The first objective is thus to develop sensing and modeling methods which can recognize, model and predict the packets loss/delay end-to-end behaviour. Given the estimated and predicted network conditions, the objective is then to adapt the data coding, protection and transmission scheme. Classical protection methods use Forward Error Correction (FEC). The code rate is then adapted to the visual data priority. However, the reliability of the estimated PER, impacts the performance of FEC schemes. This is the first problem we propose to investigate focusing on the problem of constructing codes which, together with a scalable source representation, would be robust to channel uncertainty, i.e. which would perform well not only on a specific channel but also "universally", hence reducing the need for a feedback channel. This would be a significant advantage compared with rateless codes such as fountain codes which require a feedback channel. Another problem which we address is the *cliff* effect from which suffer classical FEC schemes when the loss rate exceeds the error correction capacity of the code. The followed direction is based on Wyner-Ziv coding, used as a tool for lossy systematic error correction. The other problem addressed concerns error concealment. This refers to the problem of estimating lost symbols from the received ones by exploiting spatial and/or temporal correlation within the video signal. Classical approaches are based on spatial and/or spatio-temporal interpolation. We investigate new methods relying on video models (based on sparsity, epitomes, ...).

The availability of wireless camera sensors has also been spurring interest for a variety of applications ranging from scene interpretation, object tracking and security environment monitoring. In such camera sensor networks, communication energy and bandwidth are scarce resources, motivating the search for new distributed image processing and coding (Distributed Source Coding) solutions suitable for band and energy limited networking environments. In the past years, the team has developed a recognized expertise in the area of distributed source coding, which in theory allows for each sensor node to communicate losslessly at its conditional entropy rate without information exchange between the sensor nodes. However, distributed source coding (DSC) is still at the level of the proof of concept and many issues remain unresolved. The goal is thus to further address theoretical issues as the problem of modeling the correlation channel between sources, to further study the practicality of DSC in image coding and communication problems.

## 2.5. Highlights

The paper "Online dictionaries for image prediction" [41] by Mehmet Turkan and Christine Guillemot has been among the 8 nominated for the best student paper award at IEEE-ICIP, Sept. 2011. The paper "Lossy compression of distributed sparse sources: a practical scheme" by G. Coluccia, E. Magli, A. Roumy and V. Toto-Zarasoa [26] has received the Best Paper Award "Francesco Carassa" 2011 awarded by GTTI (Gruppo Italiano Telecomunicazioni e Teoria dell'Informazione).

# 3. Scientific Foundations

## 3.1. Introduction

The research activities on analysis, compression and communication of visual data mostly rely on tools and formalisms from the areas of statistical image modelling, of signal processing, of coding and information theory. However, the objective of better exploiting the HVS properties in the above goals also pertains to the areas of perceptual modelling and cognitive science. Some of the proposed research axes are also based on scientific foundations of computer vision (e.g. multi-view modelling and coding). We have limited this section to some tools which are central to the proposed research axes, but the design of complete compression and communication solutions obviously rely on a large number of other results in the areas of motion analysis, transform design, entropy code design, etc which cannot be all described here.

## 3.2. Parameter estimation and inference

Parameter estimation is at the core of the processing tools studied and developed in the team. Applications range from the prediction of missing data or future data, to extracting some information about the data in order to perform efficient compression. More precisely, the data are assumed to be generated by a given stochastic data model, which is partially known. The set of possible models translates the a priori knowledge we have on the data and the best model has to be selected in this set. When the set of models or equivalently the set of probability laws is indexed by a parameter (scalar or vectorial), the model is said parametric and the model selection resorts to estimating the parameter. Estimation algorithms are therefore widely used at the encoder in order to analyze the data. In order to achieve high compression rates, the parameters are usually not sent and the decoder has to jointly select the model (i.e. estimate the parameters) and extract the information of interest.

## 3.3. Data Dimensionality Reduction

A fundamental problem in many data processing tasks (compression, classification, indexing) is to find a suitable representation of the data. It often aims at reducing the dimensionality of the input data so that tractable processing methods can then be applied. Well-known methods for data dimensionality reduction include the principal component analysis (PCA) and independent component analysis (ICA). The methodologies which will be central to several proposed research problems will instead be based on sparse representations and on the so-called "non negative matrix factorization" framework.

The objective of *sparse representations* is to find a sparse approximation of a given input data. In theory, given $A \in \mathbb{R}^{m \times n}$, $m < n$, and $\mathbf{b} \in \mathbb{R}^m$ with $m << n$ and $A$ is of full rank, one seeks the solution of $\min\{\|\mathbf{x}\|_0 \ : \ A\mathbf{x} = \mathbf{b}\}$, where $\|\mathbf{x}\|_0$ denotes the $L_0$ norm of $x$, i.e. the number of non-zero components in $z$. There are many solution $x$ to $Ax = b$ and the problem is to find the sparsest, the one for which $x$ has the fewest non zero components. In practice, one actually seeks an approximate and thus even sparser solution which satisfies $\min\{\|\mathbf{x}\|_0 \ : \ \|A\mathbf{x} - \mathbf{b}\|_p \leq \rho\}$, for some $\rho \geq 0$, characterizing an admissible reconstruction error. The norm $p$ is usually 2, but could be 1 or $\infty$ as well. Except for the exhaustive combinatorial approach, there is no known method to find the exact solution under general conditions on the dictionary $A$. Searching for this sparsest representation is hence unfeasible and both problems are computationally intractable. Pursuit algorithms have been introduced as heuristic methods which aim at finding approximate solutions to the above problem with tractable complexity.

*Non negative matrix factorization* (NMF) is a non-negative approximate data representation [3]. NMF aims at finding an approximate factorization of a non-negative input data matrix V into non-negative matrices W and H, where the columns of $W$ can be seen as *basis vectors* and those of $H$ as coefficients of the linear approximation of the input data. Unlike other linear representations like principal component analysis (PCA) and independent component analysis (ICA), the non-negativity constraint makes the representation purely additive. Classical data representation methods like PCA or VQ can be placed in an NMF framework, the differences arising from different constraints being placed on the $W$ and $H$ matrices. In VQ, each column of $H$ is constrained to be unary with only one non-zero coefficient which is equal to 1. In PCA, the columns of $W$ are constrained to be orthonormal and the rows of $H$ to be orthogonal to each other. These methods

[3] D.D. Lee and H.S. Seung, "Algorithms for non-negative matrix factorization", Nature 401, 6755, (Oct. 1999), pp. 788-791

of data-dependent dimensionality reduction will be at the core of our visual data analysis and compression activities.

## 3.4. Perceptual Modelling

The human visual system is not able to process all visual information of our visual field at once. To cope with this problem, our visual system must filter out the irrelevant information and reduce redundant information. This feature of our visual system is driven by a selective sensing and analysis. The former would be related to the intrinsic conception of our biological system. For instance, it is well known that the greatest visual acuity is provided by the fovea (center of the retina). Beyond this area, the acuity drops down with the eccentricity. Another example concerns the light that impinges on our retina. Only the visible light spectrum lying between 380 nm (violet) and 760 nm (red) is processed. To conclude on the selective sensing, it is important to mention that our sensitivity depends on a number of factors such as the spatial frequency, the orientation or the depth. These properties are modeled by sensitivity function (for instance, CSF standing for Contrast Sensitivity Function). The latter point dealing with our capacity of analysis is related to the visual attention.

Visual attention which is closely linked to eye movement (note that this attention is called overt where the covert attention does not involve eye movement) allows us to focus our biological resources on a particular area. It can be controlled by both top-down (i.e. goal-directed, intention) and bottom-up (stimulus-driven, data-dependent) sources of information [4]. This detection is also influenced by prior knowledge about the environment of the scene[5]. Implicit assumptions related to Prior knowledge or beliefs form play an important role in our perception (see the example concerning the assumption that light comes from above-left). Our perception results from the combination of prior beliefs with data we gather from the environment. A Bayesian framework is an elegant solution to model these interactions[6]. We define a vector $\overrightarrow{v}_l$ of local measurements (contrast of color, orientation, etc.) and vector $\overrightarrow{v}_c$ of global and contextual features (global features, prior locations, type of the scene, etc.). The salient locations $S$ for a spatial position $\overrightarrow{x}$ are then given by:

$$S(\overrightarrow{x}) = \frac{1}{p(\overrightarrow{v}_l \,|\, \overrightarrow{v}_c)} \times p(s, \overrightarrow{x} \,|\, \overrightarrow{v}_c) \tag{1}$$

The first term represents the bottom-up salience. It is based on a kind of contrast detection, following the assumption that rare image features are more salient than frequent ones. Most of existing computational models of visual attention rely on this term. However, different approaches exist to extract the local visual features as well as the global ones. The second term is the contextual priors. For instance, given a scene, it indicates which parts of the scene are likely the most salient.

## 3.5. Coding theory

Source coding and channel coding theory [7] is central to our compression and communication activities, in particular to the design of entropy codes and of error correcting codes. Another field in coding theory which has emerged in the context of sensor networks is Distributed source coding (DSC). It refers to the compression of correlated signals captured by different sensors which do not communicate between themselves. All the signals captured are compressed independently and transmitted to a central base station which has the capability to decode them jointly. DSC finds its foundation in the seminal Slepian-Wolf[8] (SW) and Wyner-Ziv

---

[4]L. Itti and C. Koch. Computational Modelling of Visual Attention. Nature Reviews Neuroscience, Vol. 2, No. 3, pp. 194-203, 2001.

[5]J. Henderson. Regarding scenes Current. Directions in Psychological Science,vol. 16, pp. 219-222, 2007

[6]L. Zhang, M. Tong, T. Marks, H. Shan, H. and G.W. Cottrell. SUN: a Bayesian framework for saliency using natural statistics.Journal of Vision, vol. 8, pp. 1-20, 2008

[7]T. M. Cover, J. A. Thomas, Elements of Information Theory, Second Edition, July 2006.

[8]D. Slepian and J. K. Wolf. "Noiseless coding of correlated information sources." IEEE Transactions on Information Theory, 19(4), pp. 471-480, July 1973.

[9] (WZ) theorems. Let us consider two binary correlated sources $X$ and $Y$. If the two coders communicate, it is well known from Shannon's theory that the minimum lossless rate for $X$ and $Y$ is given by the joint entropy $H(X, Y)$. Slepian and Wolf have established in 1973 that this lossless compression rate bound can be approached with a vanishing error probability for long sequences, even if the two sources are coded separately, provided that they are decoded jointly and that their correlation is known to both the encoder and the decoder.

In 1976, Wyner and Ziv considered the problem of coding of two correlated sources $X$ and $Y$, with respect to a fidelity criterion. They have established the rate-distortion function $R*_{X|Y}(D)$ for the case where the side information $Y$ is perfectly known to the decoder only. For a given target distortion $D$, $R*_{X|Y}(D)$ in general verifies $R_{X|Y}(D) \leq R*_{X|Y}(D) \leq R_X(D)$, where $R_{X|Y}(D)$ is the rate required to encode $X$ if $Y$ is available to both the encoder and the decoder, and $R_X$ is the minimal rate for encoding $X$ without SI. These results give achievable rate bounds, however the design of codes and practical solutions for compression and communication applications remain a widely open issue.

# 4. Application Domains

## 4.1. Introduction

The application domains addressed by the project are:

- Compression with advanced functionalities of various image modalities (including multi-view, medical images such as MRI, CT, WSI, or satellite images
- Networked multimedia applications via their various needs in terms of image and 2D and 3D video compression, or in terms of network adaptation (e.g., resilience to channel noise)
- Content editing and post-production

## 4.2. Compression with advanced functionalities

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

Different video data formats and technologies are envisaged for interactive and immersive 3D video applications using omni-directional videos, stereoscopic or multi-view videos. The "omni-directional video" set-up refers to 360-degree view from one single viewpoint or spherical video. Stereoscopic video is composed of two-view videos, the right and left images of the scene which, when combined, can recreate the depth aspect of the scene. A multti-view video refers to multiple video sequences captured by multiple video cameras and possibly by depth cameras. Associated with a view synthesis method, a multi-view video allows the generation of virtual views of the scene from any viewpoint. This property can be used in a large diversity of applications, including Three-Dimensional TV (3DTV), and Free Viewpoint Video (FTV). The notion of "free viewpoint video" refers to the possibility for the user to choose an arbitrary viewpoint and/or view direction within a visual scene, creating an immersive environment. Multi-view video generates a huge amount of redundant data which need to be compressed for storage and transmission. In parallel, the advent of a variety of heterogeneous

---

[9] A. Wyner and J. Ziv. "The rate-distortion function for source coding ith side information at the decoder." IEEE Transactions on Information Theory, pp. 1-10, January 1976.

delivery infrastructures has given momentum to extensive work on optimizing the end-to-end delivery QoS (Quality of Service). This encompasses compression capability but also capability for adapting the compressed streams to varying network conditions. The scalability of the video content compressed representation, its robustness to transmission impairments, are thus important features for seamless adaptation to varying network conditions and to terminal capabilities.

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

## 4.3. Networked visual applications

*3D and Free Viewpoint TV:* The emergence of multi-view auto-stereoscopic displays has spurred a recent interest for broadcast or Internet delivery of 3D video to the home. Multiview video, with the help of depth information on the scene, allows scene rendering on immersive stereo or auto-stereoscopic displays for 3DTV applications. It also allows visualizing the scene from any viewpoint, for scene navigation and free-viewpoint TV (FTV) applications. However, the large volumes of data associated to multi-view video plus depth content raise new challenges in terms of compression and communication.

*Internet and mobile video:* Broadband fixed (ADSL, ADSL2+) and mobile access networks with different radio access technologies (RAT) (e.g. 3G/4G, GERAN, UTRAN, DVB-H), have enabled not only IPTV and Internet TV but also the emergence of mobile TV and mobile devices with internet capability. A major challenge for next internet TV or internet video remains to be able to deliver the increasing variety of media (including more and more bandwidth demanding media) with a sufficient end-to-end QoS (Quality of Service) and QoE (Quality of Experience).

*Mobile video retrieval:* The Internet has changed the ways of interacting with content. The user is shifting its media consumption from a passive to a more interactive mode, from linear broadcast (TV) to on demand content (YouTubes, iTunes, VoD), and to user-generated, searching for relevant, personalized content. New mobility and ubiquitous usage has also emerged. The increased power of mobile devices is making content search and retrieval applications using mobile phones possible. Quick access to content in mobile environments with restricted bandwidth resources will benefit from rate-efficient feature extraction and description.

*Wireless multi-camera vision systems:* Our activities on scene modelling, on rate-efficient feature description, distributed coding and compressed sensing should also lead to algorithmic building blocks relevant for wireless multi-camera vision systems, for applications such as visual surveillance and security.

## 4.4. Medical Imaging (CT, MRI, Virtual Microscopy)

The use of medical imaging has greatly increased in recent years, especially with *magnetic resonance images (MRI) and computed tomography (CT)*. In the medical sector, lossless compression schemes are in general used to avoid any signal degradation which could mask a pathology and hence disturb the medical diagnosis. Nethertheless, some discussions are on-going to use near-lossless coding of medical images, coupled with a detection and segmentation of region-of interest (ROIs) guided by a modeling stage of the image sensor, a precise knowledge of the medical imaging modalities and by the diagnosis and expertise of practitioners. New application domains using these new approaches of telemedecine will surely increase in the future. The second aspect deals with the legal need of *biomedical images storage* The legacy rules of such archives are changing and it could be interesting to propose adaptive compression strategies, i.e to explore reversible lossy-to-lossless coding algorithms and new storage modalities which use, in a first stage, the lossless representation and continuously introduce controlled lossy degradations for the next stages of archives. Finally, it seems promising to explore new representation and coding approaches for 3D biological tissue imaging captured by *3D virtual microscopy*. These fields of interest and scientific application domains commonly generate terabytes of data. Lossless schemes but also lossy approaches have to be explored and optimized, and interactive tools supporting scalable and interactive access to large-sized images such as these virtual microscopy slides need to be developed. 2D3Dinpainting

## 4.5. Editing and post-production

Video editing and post-production are critical aspects in the audio-visual production process. Increased ways of "consuming" video content also highlight the need for content repurposing as well as for higher interaction and editing capabilities. Content captured at very high resolutions may need to be repurposed in order to be adapted to the requirements of actual users, to the transmission channel or to the terminal. Content repurposing encompasses format conversion (retargeting), content summarization, and content editing. This processing requires powerful methods for extracting condensed video representations as well as powerful inpainting techniques. By providing advanced models, advanced video processing and image analysis tools, more visual effects, with more realism become possible. Other applications such as video annotation/retrieval, video restoration/stabilization, augmented reality, can also benefit from the proposed research.

# 5. Software

## 5.1. Oriented wavelet based image codec

**Participant:** Christine Guillemot [contact person].

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

## 5.2. M3DPlayer: 3D video player

**Participant:** Vincent Jantet [contact person].

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

A second version of the player has been registered at the APP (Agence de Protection des Programmes) under the number IDDN.FR.001.090023.000.S.P.2008.000.21000. In 2009-2010, we focused on improving the rendering engine, based on recent OpenGL extensions, to be able to render the viewed scenes on an auto-stereoscopic display with low-end graphic cards. In our case, auto-stereoscopic display requires the rendering of eight $1920x1200$ frames instead of just one for a standard display. This player is also used to render LDI (Layered Depth Images) and LDV (Layered Depth Videos) and to visualize 3D scenes on autostereoscopic displays taking multiple input views rendered from the LDI representation.

## 5.3. Depth maps extractor in mono-view (M3dAnalyzer2)

**Participant:** Josselin Gauthier [contact person].

This software estimates depth maps from a video captured by a unique camera moving in a static 3D environment with Lambertian surfaces. These sequences are of interest to specialized applications such as augmented reality, remote-controlled robots operating in hazardous environments or remote exploration by drones. This software has been filed at the APP (Agence de Protection des Programmes) under the number IDDN.FR.001.110031.000.S.P.2010.000.31235.

## 5.4. Depth maps extractor in multi-view (MV2MVD)

**Participant:** Josselin Gauthier [contact person].

This software estimates depth maps from multi-view videos, to provide Multi-View plus Depth (MVD) videos. MVD videos can be used to synthesize virtual views of the scene, or to render a different number of views than captured in the original video, for instance on an auto-stereoscopic display. This software produces depth maps of higher quality than those generated by the Depth Estimation Reference Software from the MPEG-3DV group, in terms of virtual views synthesis quality. This software has been filed at the APP (Agence de Protection des Programmes) under the number IDDN.FR.001.110034.000.S.P.2010.000.31235.



*Figure 1. Depth maps extracted for the kendo sequence (right and left views) with DERS (3rd column) and TEMICS software (middle column).*

## 5.5. LDI builder

**Participant:** Vincent Jantet [contact person].

This software constructs a Layered Depth Image (LDI) representation of un-rectified Multi-View + Depth (MVD) sequences. The Incremental construction scheme reduces inter-layer correlation. The generated I-LDI is compatible with the M3DPlayer, permitting 3D visualisation and free viewpoint rendering of the 3D scene. The software also implements a virtual-view rendering technique which significantly reduces ghosting artefacts by eliminating untrusted texture boundaries detected in depth maps, as well as cracking artefacts thanks to an epipolar geometry aided inpainting method.

## 5.6. ADT PICOVIN

**Participants:** Ronan Boitard, Laurent Guillo [contact person], Thomas Guionnet, Tangi Poirier.

The ADT Picovin is a technological development action, which works closely with the project-team TEMICS. This is a development structure which gives its support to the project-team to integrate new and relevant algorithms into the state-of-the-art codec and to take part in standardization.

The ITU-T Study Group 16 (VCEG) and ISO/IEC JTC 1/SC 29/WG 11 (MPEG) have created in 2010 the Joint Collaborative Team on Video Coding (JCT-VC) in order to develop a new generation video coding standard that will further reduce by 50

In 2011, the ADT mainly focused on the improvement and integration of algorithms dedicated to intra prediction. A part of our work was integrated in the HM 1.0 and then submitted and presented as a proposal in Daegu during the 4th JCT-VC meeting in January 2011. Bit rate gains that we obtained were significant but to the detriment of encoding and decoding times. That is why, all along this year, the ADT tried to reach the best tradeoff between performances and encoding/decoding times. Our solution based on linear combination of template matching predictors has been improved by taking into account several shapes of template. The very last integration in the HM (HM4.0) shows that it performs well especially for the class "screen content". A new very promising intra prediction method, Short Distance Intra Prediction (SDIP) is being tested as part of a Core Experiment in the HM 4.0. Once it is associated to our approach, bit save gains are additive. These results will be presented during the 7th JCT-VC meeting in Geneva in November 2011.

During 2011, the ADT also took part in cross checks which aims at evaluating and testing tools studied in core experiments. As part of cross checks the ADT has run 9 tests jointly with companies such as Technicolor, Mitsubushi, Huawei, Qualcom and Canon.

This ADT started in October 2008. It will go on for one more year through the ADT PICOVIN-P. During this year, one permanent engineer from the SED Rennes (development and experimentation department of INRIA Rennes) and one senior engineer specialized in video compression are involved in the ADT. It is supported by the technological development department of INRIA.

# 6. New Results

## 6.1. Analysis and modeling for compact representation and navigation

### 6.1.1. *Joint projection/filling method for virtual view synthesis*
**Participants:** Christine Guillemot, Vincent Jantet.

This study is carried out in collaboration with INSA/IETR (Luce Morin). Associated with a view synthesis method, a multi-view plus depth video allows the generation of virtual views of the scene from any viewpoint. Many algorithms have thus been developed to synthesize virtual views from one or several input views video plus depth data. These rendering algorithms are either based on Image-Based Rendering (IBR) techniques or Geometry-Based Rendering (GBR) techniques, according to the amount of 3D information they use. IBR techniques require limited geometric information to synthesize intermediate views and allow the generation of photo-realistic virtual views at the expense of virtual camera freedom. GBR techniques require detailed 3D models of the scene to synthesize arbitrary viewpoints (points of view). GBR techniques are sensitive to the accuracy of the 3D model, which is difficult to estimate from real multi-view videos. Depth-Image-Based Rendering (DIBR) techniques include hybrid rendering methods between IBR and GBR techniques. DIBR methods are based on warping equations, which project a reference view onto a virtual viewpoint. Each input view is defined by a "color" (or "texture") map and a "depth" map, which associates a depth value to each image pixel.

In classical DIBR schemes, the rendering proceeds in several distinct steps, each one designed to solve a specific problem. First, the input depth map is warped onto the virtual viewpoint. The obtained warped depth map contains disocclusions, cracks and ghosting artifacts. Second, this virtual depth map is filtered with a median filter, in order to remove the cracks, and then to dilate disocclusion areas on the background side, in order to avoid ghosting artifacts during view synthesis. Third, the filtered depth map is used in a backward warping to compute the color of each pixel of the virtual view. Fourth, this resulting depth map is inpainted, to fill in disocclusion areas. Finally, this complete depth map is used by a depth-aided inpainting algorithm to fill in disocclusions in the color map. However, all these steps are inter-dependent, and errors introduced by each one are amplified by the following one. Connectivity information is lost during the first projection step, as shown in Fig. 2. Without this connectivity information, every inpainting method fails to fill in background disocclusions if the disoccluded area is surrounded by foreground objects. This case may happen each time a foreground object is not convex, and contains holes, as shown in Fig. 2-(a). As a result, depth-aided inpainting uses wrong foreground patches to fill in background disocclusions, producing annoying artifacts, as shown in Fig. 2-(b).

We have developed two DIBR techniques, both based on a novel forward projection technique, called the Joint Projection Filling (JPF) method [16]. The JPF method performs forward projection, using connectivity information to fill in disocclusions in a single step. The first proposed DIBR method is designed to extrapolate virtual views from a single input view plus depth video sequence. The synthesis of virtual depth maps by the JPF method avoids the use of dedicated filtering and inpainting processes and leads to synthsized depth maps of higher quality. The second proposed DIBR method is designed to interpolate intermediate views from multiple input view plus depth sequences. This interpolation method uses the Floating Texture approach to register multiple inputs view plus depth sequences before blending. The JPF method fills in disocclusion areas during the projection, to ensure that geometrical structures are well preserved. The method uses the occlusion-compatible ordering presented by McMillan, which uses epipolar geometry to select a pixel scanning order.The occlusion-compatible ordering is used to handle disocclusions gracefully. Cracks are filled in by interpolation of neighboring pixels, whereas disocclusions are only filled in by background pixels. This technique can also be used with non-rectified views, avoiding prior creation of parallax maps.
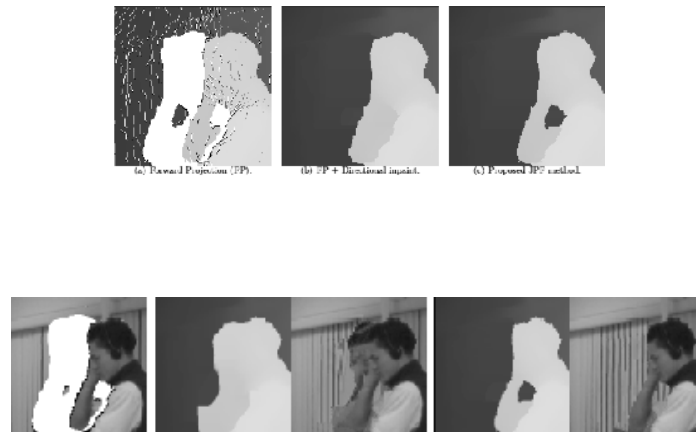


*Figure 2. Virtual depth map synthesized by three forward projection methods. The point-based projection method generates cracks and disocclusions (top-left). Median filtering and directional inpainting fills holes with foreground depth (top-middle). The proposed JPF method fills cracks and disocclusions with realistic background (top-right). Synthesized view with disocclusion (bottom-left), Synthesized depth maps, obtained with a Navierstokes inpainting algorithm (column 2), with the developed JPF method (column 4) and the corresponding synthesized views with the two depth maps.*

### 6.1.2. 2D/3D image inpainting for virtual view synthesis

**Participants:** Josselin Gauthier, Christine Guillemot, Mouid Keskes, Olivier Le Meur.

Inpainting methods play an important role in a wide range of applications. Removing text and advertisements (such as logos), removing undesired objects, noise reduction and image reconstruction from incomplete data are the key applications of inpainting methods. Algorithms can be classified into two categories: PDE (Partial Derivative Equation)-based schemes and examplar-based schemes. The former uses diffusion schemes in order to propagate structures in a given direction. Their drawback is the introduction of blur due to diffusion. The latter relies on the sampling and the copying of texture from the known parts of the picture.

We have proposed a novel inpainting algorithm combining the advantages of both aforementionned methods. As in Criminisi et al's approach [10], the proposed method involves two steps: first, a filling order is defined to favor the propagation of structure in the isophote direction. Second, a template matching is performed in order to find the best candidates to fill in the hole. Compared to previous approaches, the main contributions concern the use of structure tensors to define the filling order instead of field gradients. The structure tensor is defined as follow:

$$\mathbf{J} = \sum_{i=\{R,G,B\}} \nabla I_i \nabla I_i^T \qquad (2)$$

$\mathbf{J}$ is the sum of the scalar structure tensors $\nabla I_i \nabla I_i^T$ of each image channel $I_i$ ($i \in \{R, G, B\}$). Information about local geometry can be deduced by computing the eigenvalues and eigenvectors of $\mathbf{J}$. The local vector geometry is computed from the structure tensor $\mathbf{J}$. Its eigenvectors $\mathbf{v}_{1,2}$ ($\mathbf{v}_i \in R^n$) define an oriented orthogonal basis and its eigenvalues $\lambda_{1,2}$ define the amount of structure variation. $\mathbf{v}_1$ is the orientation with the highest fluctuations (orthogonal to the image contours), and $\mathbf{v}_2$ gives the preferred local orientation. This eigenvector (having the smallest eigenvalue) indicates the isophote orientation. The use of structure tensor allows to retrieve a more coherent local geometry. The computation of the filling order as proposed by Criminisi et al is then replaced by a term coming from PDE-based schemes, called Coherence Enhancing Diffusion. The use of structure tensor in an examplar-based approach leads to a more robust algorithm that visually improves the quality of the inpainted areas.

Additionnaly the simple template matching originally used in previous methods has been improved by using a K-nearest neighbor approach. The weigths of the linear combination of the first $K$ best candidate are adjusted by taking into account that all candidate patches are not equally reliable. Note that the number $K$ is also locally adjusted in function of the local spatial complexity.

The 2D inpainting algorithm described above has been extended to deal with 3D content. In this work the goal is to synthesize novel views directly from the original images. Image-based rendering (IBR) is commonly used to render a virtual view. It generates a nearby viewpoint image by projecting a point from the reference view to the virtual view using the depth data. However, when the viewpoint is shifted, occluded regions in the original view point are dissoccluded. Handling these disocclusions (holes) is a difficult problem. We propose to use an extension of the 2D inpainting method to fill in these holes. For this goal, we have modified the computation of the structure tensor by adding the depth information. Equation (2) is simply modified as follow:

$$\mathbf{J} = \sum_{i=\{R,G,B,Z\}} \nabla I_i \nabla I_i^T \qquad (3)$$

where $Z$ represents the depth map. As previously, the tensor is used to compute the filling order. A directional term is also included in order to favor a filling direction. Specifically, when the viewpoint is shifted from left to right in the horizontal direction, occluded regions in the left image appear in the right image around the right side of the object. Therefore, it is recommended to start the filling from the right to the left. This filling is performed by a modified template matching using texture information as well as depth data. Figure 3 illustrates the inpainting quality for different approaches.

---

[10]Criminisi et al., Region filling and object removal by examplar-based image inpainting, IEEE Trans. On Image Processing, vol. 13, pp. 1200-1212, 2004

*Figure 3. Virtual synthesized view. From left to right: original view projected into the new viewpoint; disocclusions filled by Criminisi's approach, Daribo's approach and the proposed method.*

### 6.1.3. Computational modelling of visual attention

**Participants:** Josselin Gauthier, Olivier Le Meur.

*6.1.3.1. Eye-movement study:*

In 2011, we have investigated whether two populations of visual fixation exist in 2D context. The question is simple: do all visual fixations have the same role in the free viewing of natural scenes? Recent studies suggest that there are at least two types of visual fixations: focal and ambient fixations. The former is believed to be used to inspect local areas accurately, whereas the latter is used to obtain the context of the scene.

From a collaboration with Technicolor (P. Guillotel and C. Chamaret) and LUTIN (T. Baccino), we found new evidence to support a focal-ambient dichotomy. Our results published in the journal i-Perception [14] indicate that the determining factor to classify the visual fixations is the saccade amplitude. We proposed an automatic system to cluster visual fixations in two groups using four types of natural scene images. From this automatic classification, the terms focal saliency map and ambient saliency map have been introduced. The dependence on the low-level visual features and the time course of these two kinds of visual fixations were examined. Our results demonstrate that there is an interplay between both fixation populations and that focal fixations are more dependent on low-level visual features than ambient fixations. These results might have a strong impact on both the computational modelling of visual attention and their performance assessment.

A second study related to eye-movement dealt with the role of the binocular disparity depth cue in the deployment of visual attention. To address this point, we compared eye tracking data recorded while observers viewed natural images in 2D and 3D conditions. The influence of disparity on the inter-observers congruency, saliency, center and depth bias was first examined. Results show that visual exploration in depth layer detection task is affected by the binocular disparity. In particular, participants tend to look first at closer areas just after the stimuli onset with the introduction of disparity, and then direct their gaze to more widespread locations. Our results has been submitted in the journal Cognitive Computation.

*6.1.3.2. Model of visual attention:*

Since 1998 with the publication of the influential work of Itti, Kock and Niebur [11], the computational modelling of the visual attention has known an increasing interest. The former models only used the low-level visual features for getting a saliency map. They perform well in a number of cases in predicting where an observer would look at. However, to improve the quality of the prediction, it seems unavoidable to use high-level information in order to account for visual deployment.

This work aims at designing a computationnal model mixing low-level and high-level features. Among the different factors influencing our gaze, we have focused our works on two cues: the dominant depth and the horizon line position. The dominant depth and the spatial position of the horizon line were inferred from the low-level visual features. A training database has been set up to perform a learning. Results indicate that the proposed model outperforms state-of-the-art models [37].

---

[11]Itti et al, A model for saliency-based visual attention for rapid scene analysis, IEEE Trans. on PAMI, vol. 20, pp. 1254-1259, 1998

From behavioual studies on eye-movement in a 3D context, we have proposed a model of visual attention able to predict saliency of 3D pictures. The method developed aims at using the depth cue, the central bias and the low-level visual features. The predicted saliency is obtained by linearly combining these cues. The weights of the linear combination are learnt from a training database and are time-dependent. This study is under revision in the journal Cognitive Computation.

*6.1.3.3. Predicting the inter-observer visual congruency:*

This work aims at predicting the inter-observer visual congruency (IOVC), indicating the congruence or the variability among different subjects looking at the same image [35]. Predicting this congruence is of interest for image processing applications where the visual perception of a picture matters such as website design, advertisement, etc. We proposed a computational model of the IOVC. This new model is a mixture of low-level visual features extracted from the input picture. Model's parameters are learned by using a large eye-tracking database. In this study, we also proposed a new scheme to compute the depth of field of a picture. Finally, once the training and the feature extraction have been carried out, a score ranging from 0 (minimal congruency) to 1 (maximal congruency) is computed. A value of 1 indicates that observers would focus on the same locations and suggests that the picture presents strong locations of interest. To illustrate the interest of the proposed model, we have used it to automatically rank personalized photograph. Figure 4 illustrates the proposed approach.
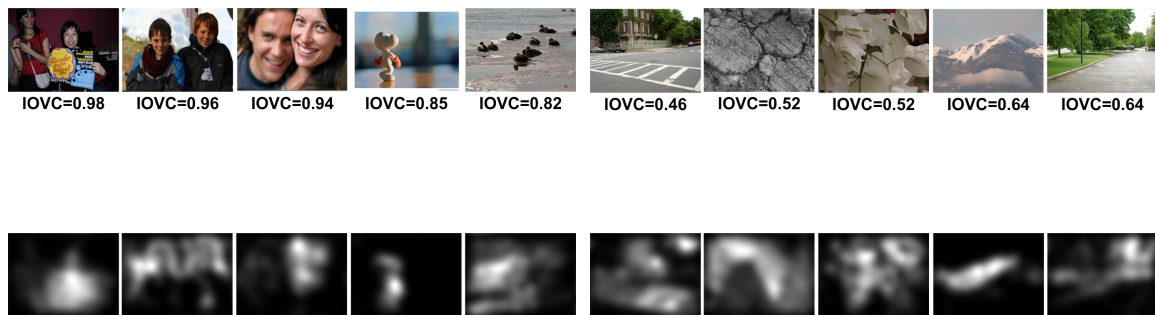


Figure 4. Top: pictures having high IOVC (first five) and pictures having low IOVC (last five). Bottom: saliency maps of pictures. Bright areas correspond to the most salient parts.

## 6.1.4. Visual cues analysis and modelling

**Participants:** Safa Cherigui, Christine Guillemot.

Thiw work is carried out in collaboration with Technicolor (D. Thoreau, Ph. Guillotel, P. Perez) and aims at designing a compresion algorithm based on the concept of epitomes. An epitome is a condensed representation of an image (or a video) signal containing the essence of the textural properties of this image. Different forms of epitomes have been proposed, such as a patch-based probability model learned either from still image patches or from space-time texture cubes taken from the input video. These probability models together with appropriate inference algorithms, are useful for content analysis inpainting or super-resolution. Another family of approaches makes use of computer vision techniques, like the KLT tracking algorithm, in order to recover self similarities within and across images. In parallel, another type of approach consists in extracting epitome-like signatures from images using sparse coding and dictionary learning.

The method developed aims at tracking self-similarities within an image using a block matching (BM) algorithm [25]. The epitome is constructed from disjoint pieces of texture ("epitome charts") taken from the original image and a transform map which contains translational parameters (see Fig.5-middle row). Those parameters keep track of the correspondences between each block of the input image and a block of the

epitome. An Intra image compression scheme based on the epitome has been developed showing a rate saving of up to 12
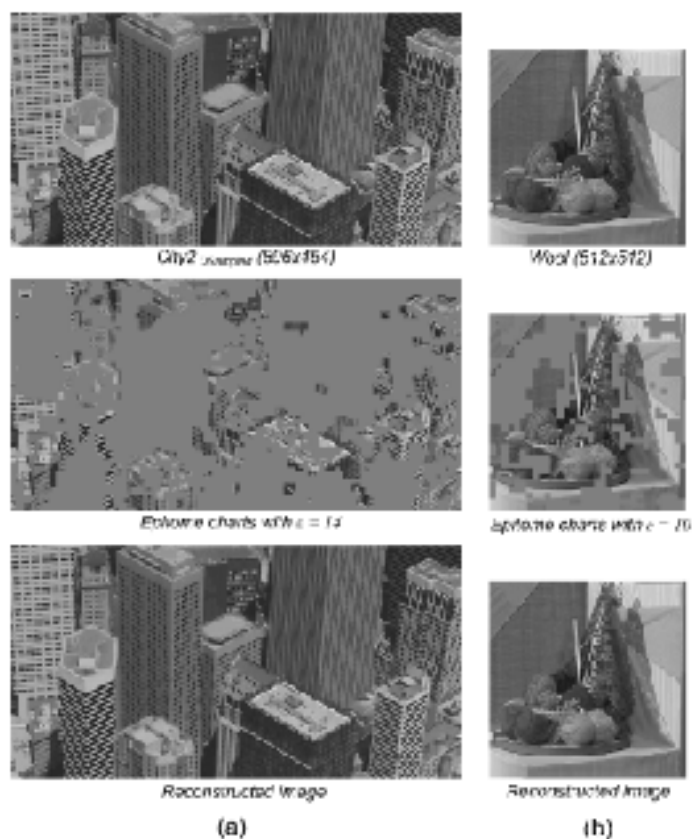


*Figure 5. Epitome and reconstructed images: (top row) Original images (columns a and b); (middle) Epitomes; (bottom) Reconstructed images from epitome texture and transform map.*

## 6.2. Representation and compression of large volumes of visual data

### 6.2.1. 3d representations for multi-view video sequences

**Participants:** Christine Guillemot, Vincent Jantet.

Multi-view plus depth video content represent very large volumes of input data wich need to be compressed for storage and tranmission to the rendering device. The huge amount of data contained in multi-view sequences indeed motivates the design of efficient representation and compression algorithms. In collaboration with INSA/IETR (Luce Morin), we have studied layered depth image (LDI) and layered depth video (LDV) representations as a possible compact representation format of multi-view video plus depth data. LDI give compact represenations of 3D objects, which can be efficiently used for photo-realistic image-based rendering (IBR) of different scene viewpoints, even with complex scene geometry. The LDI extends the 2D+Z representation, but instead of representing the scene with an array of depth pixels (pixel color with associated depth values), each position in the array may store several depth pixels, organised into layers.

Various approaches exist to construct LDI, which all organize layers by visibility. The first layer contains all pixels visible from a chosen reference viewpoint. The other layers contain pixels hidden by objects in previous layers. With classical construction solutions, each layer may contain pixels from the background and pixels from objects in a same neighbourhood, creating texture and depth discontinuities within the same layer. These discontinuities are blurred during the compression process which in turn significantly reduces the rendering quality.

We have thus developed a novel object-based LDI representation which is more tolerant to compression artifacts, as well as being compatible with fast mesh-based rendering techniques [34]. This representation organises LDI pixels into two separate layers (foreground and background) to enhance depth continuity (see Fig.6). The construction of this object-based LDI makes use of a foreground-background region-growing segmentation algorithm followed by inpainting of both colour and texture images to have a complete background layer (without the holes corresponding to disocclusion areas). The costly inpainting algorithm is thus processed once, during the LDI classification, and not during each view synthesis, which helps to speed up the rendering step.
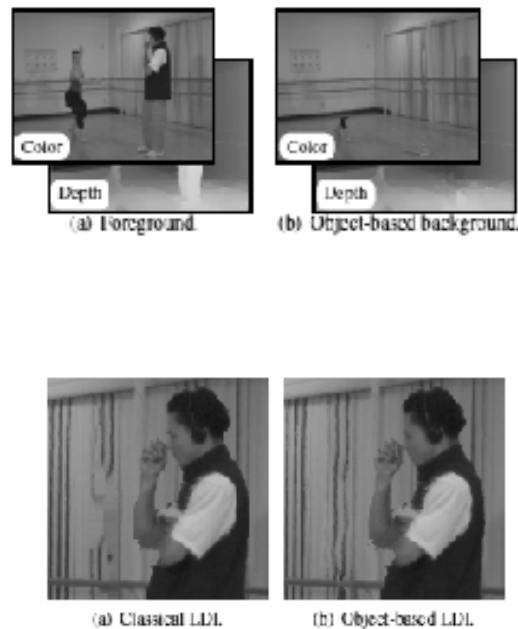


*Figure 6. Object-based LDI: (top) Foreground and background layers; (bottom) Rendering results classical and object LDI.*

### 6.2.2. *From sparse to spread representations*
**Participant:** Jean Jacques Fuchs.

Sparse representations, where one seeks to represent a vector on a redundant basis using the smallest number of basis vectors, appear to have numerous applications. The other extreme, where one seeks a representation that uses all the basis vectors, might be of interest if one manages to spread the information nearly equally over all of them. Minimizing the $\ell_\infty$-norm of the vector of weights is one way to find such a representation. Properties of the solution and a dedicated fast algorithm have been developed. While the application of such models in robust data coding and in improving achievable data rates over amplitude constrained channels

seems to be wishful thinking, its use in indexing techniques appears to be promising. In this context, one further replaces the optimal vector by its sign vector (potentially associated with a re-evaluated scalar weight) to get a binary vector that is not only cheap to store and (somehow) easy to search for but also allows for an explicit reconstruction unlike all other Hamming embedding functions used to map real vectors into binary vectors.

### 6.2.3. *On-line dictionary learning methods for prediction*
**Participants:** Christine Guillemot, Mehmet Turkan.

One crucial question to the problem of sparse approximation, and hence also of prediction based on sparse approximation, is the choice of the dictionary. Various advanced dictionary learning schemes have been proposed in the literature for the sparse signal approximation problem, so that the dictionary used is well suited to the data at hand. The popular dictionary learning algorithms include the K-SVD, the Method of Optimal Directions (MOD), Sparse Orthonormal Transforms (SOT), and (Generalized) Principle Component Analysis (PCA). However, the above learning methods are often used off-line since their computational complexity, which results from the number and the dimension of training samples, makes them inappropriate for online learning. In addition, these methods are adapted to the learning of basis to be used for approximating input data vectors, but not to the problem of predicting unknown samples from noisy observed samples in a causal neighborhood.

In 2011, we have developed a method for on-line training dictionaries adapted to the prediction problem [41]. Let A be the input dictionary, which is divided into two sub-dictionaries: $A_c$ and $A_t$. The goal is to have a simple on-line dictionary learning method which is adapted to the prediction problem, i.e., which will learn both sub-dictionaries so that sparse vectors found by approximating the known samples (the template) using the first sub-dictionary ($A_c$) will also lead to a good approximation of the block to be predicted when used together with the second sub-dictionary ($A_t$). When dealing with the prediction problem, the sparse signal approximation is indeed first run with a set of masked basis functions (dictionary $A_c$), the masked samples corresponding to the location of the pixels to be predicted. The principle of the approach is to first search for the linear combination of basis functions which best approximates known sample values in a causal neighborhood, and keep the same linear combination of basis functions, but this time with the unmasked samples (dictionary $A_t$) to approximate the block to be predicted. The decoder similarly runs the algorithm with the masked basis functions and taking the previously decoded neighborhood as the known support. The use of masked basis functions converts the complete approximation problem into an overcomplete approximation problem. Because of its simplicity, of the limited number of training samples it requires, can be used for online learning of dictionaries, i.e. while doing the block-wise encoding of the image. The training samples are all possible previously coded/decoded texture patches (blocks of pixels) within a search window located in a causal neighborhood of the block to be predicted.

### 6.2.4. *Neighbor embedding methods for image prediction and inpainting*
**Participants:** Christine Guillemot, Mehmet Turkan.

The problem of texture prediction as well as image inpainting can be regarded as a problem of texture synthesis. Given observations, or known samples in a spatial neighborhood, the goal is to estimate unknown samples of the block to be predicted or of the patch to be filled in inpainting. We have developed texture prediction methods as well as a new inpainting algorithm based on neighbor embedding techniques which come from the area of data dimensionality reduction. The methods which we have more particularly considered are Locally Linear Embedding (LLE) and Non-negative Matrix Factorization (NMF). The first step in the developed methods consists in searching, within the known part of the image, for the $K$ nearest (KNN) patches to the set of known samples in the neighborhood of the block to be predicted or of samples to be estimated in the context of inpainting. This first step can be seen as constructing a dictionary matrix by stacking in the matrix columns the vectorized K-NN texture patches. The non-negative dictionary matrix $\mathbf{A} \in \mathbb{R}^{N \times M}$, is formed by K nearest neighbors to the vector formed by the known samples in the neighborhood of the samples to be predicted. These K nearest neighbors are texture patches of the same shape taken from the known part of the image. This dictionary can then be used for approximating the known samples by masking the rows of

the matrix which correspond to the position of the unknown samples, solving a least squares problem under the constraint of sum-to-one of the wieghts in the case of LLE, or under the constraint of non-negativity of the weights for NMF. It is actually a variant of NMF since one of the components matrices is fixed (the one corresponding to the dictionary matrix) and only the matrix containing the weights of the linear approximation must then be found. The approaches are thus intended to explore the properties of the manifolds on which the input texture patches are assumed to reside. The underlying assumption is that the corresponding uncomplete and complete patches have similar neighborhoods on some nonlinear manifolds. The new prediction methods give RD performances which are signficantly better than the ones given by the H.264 Intra prediction modes, in particular for highly textured images [21], the highest gain being achieved with NMF.

A new examplar-based inpainting algorithm using neighbor embedding techniques has been developed. A new priority order has been proposed in order to inapint first patches containing structures or contour information. The methods have also been shown to enhance the quality of inpainted images when compared to classical examplar-based solutions using simple template matching techniques to estimate the missing pixels, (see Fig. 7).
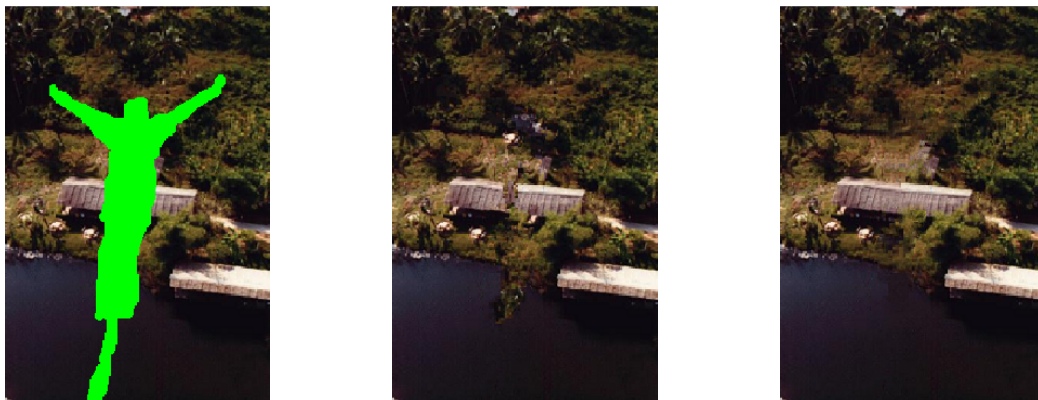


*Figure 7. Inpainting results: (left) mask of the image to be inpainted; (middle) Inpainting results with classical examplar-based inpainting; (right) Inpainting results with LLE (right).*

### 6.2.5. *Lossless coding for medical images*

**Participants:** Claude Labit, Jonathan Taquet.

Last year, we developed a hierarchical oriented prediction (HOP) algorithm, for resolution scalable lossless and near lossless compression of biomedical images. In 2011, the algorithm has been slightly improved with an iterative optimization of the predictors in order to get better results on less noisy/smooth images [39].

Recently, there have been a growing interest for the compression of an emerging imaging modality : the virtual microscopy (VM). It is used in anatomopathology and may produce huge images of more than 1 Gigabytes. We have studied the efficiency for lossless and lossy compression of our previously developed algorithms HOP and OWD (optimized wavelet decomposition) and of two extensions of OWD : near-lossless and/or region of interest (ROI) coding. The lossless results, which are slightly better than JPEG-LS and JPEG-2000 standards with about 3:1 compression ratio, show that lossless compression is not suited to VM. By compressing only the information area (ROI) which represents about 20 percents of the size of test images, 9:1 ratio could be obtained, and combined with near-lossless approach, depending on the required quality, ratio can reach 17:1 with no visual losses to more than 30:1 with some visual losses (or approximately about 6:1 for ROI only data). We have concluded that it would probably be better to use lossy or efficient quality scalable compression. Because those images have specific contents (cellular tissus for example) we have also

introduced and investigated new learning based methods. We have developed an optimization process for designing multiple KLT (Karhunen-Loeve Transform) in order to get orthonormal bases that are optimal for decorrelation and quality scalability. This learning approach has been applied as an a-posteriori transform of a wavelet decomposition in order to propose transforms with no blocking artefacts. A fully quality-scalable coding algorithm allows to obtain interesting PSNR improvements compared to the optimized coding process of JPEG-2000. Gain is around 0.5 dB for 16:1 compression of ROI only data, and more than 1 dB for 8:1 compression ratio.

## 6.3. Distributed processing and robust communication

### 6.3.1. *Loss concealment based on video inpainting*
**Participants:** Mounira Ebdelli, Christine Guillemot, Olivier Le Meur.

In 2011, we have started developing a loss concealment scheme based on a new video examplar-based inpainting algorithm. The developed video inpainting approach relies on new patch priority functions as welll as on a motion confidence-aided neighbor embedding techniques. Neighbor embedding approaches aim at approximating input vectors (or data points) as a linear combination of their neighbors. The search of the weights of the linear combination (i.e. of the embedding) are formulated as constrained least squares problems. When using the locally linear embedding, the constraint is that the sum of the weights is equal to 1. We have also considered non-negative matrix factorization to solve the problem, in which case the constraint is that the weights and the other vector are non-negative. The motion confidence introduced in the neighbor embedding improves the robustness of the algorithm in the sense that it limits the error propgation effects which otherwise result from uncertainties on the motion information of the unknown pixels to be estimated. A new patch similarity measure which accounts for the correlation between motion information has been defined for the $K$-NN search inherent to neighbor embedding techniques. Evaluations of the algorithm in a context of video editing (object removal) are on-going. The next step will be to assess the performance of the approach in a context of loss concealment, that is to estimate unknown pixels after decoding when the corersponding transport packets have been lost on the transmission network.

### 6.3.2. *Unequal Erasure Protection and Object Bundle Protection*
**Participant:** Aline Roumy.

In 2011, we started a new collaboration in the framework of the Joint INRIA/Alcatel Lucent lab. In this work, carried out with V. Roca (Planete, INRIA), B. Sayadi and R. Imad (Alcatel Lucent), we proposed and analyzed a novel technique capable of providing both an unequal erasure protection service and an object bundle protection service.

*Unequal Erasure Protection:* When a data flow contains information of different priority levels, it is natural to try to offer an unequal protection where the high priority data benefits from a higher protection than the rest of data. In this work we focused on the "erasure channel", for instance the Internet where the UDP/IP datagram integrity is guaranteed by the physical layer FCS (or CRC) and the UDP checksum. In this context UEP refers to an Unequal Erasure Protection (rather than Error) and the FEC code being used is one of the various Application-Layer Forward Erasure Correction (AL-FEC) codes that have been designed and standardized in the past years, like Reed-Solomon, one of the LDPC variants, or Raptor(Q) codes. Offering an unequal protection in this context can be achieved by one of the following three general approaches: by using dedicated UEP-aware FEC codes, by using a dedicated UEP-aware packetization scheme, or by using an UEP-aware signaling scheme. In this work we ignored the first approach as we wanted to reuse existing AL-FEC codes. Instead we focused on and compared the last two approaches and more precisely the well known Priority Encoding Transmission (PET) scheme that belongs to the UEP-aware packetization category and a Generalized Object Encoding (GOE) scheme, we proposed [53], that belongs to the UEP-aware signaling category. Through a careful modeling of both proposals [55], whose accuracy has been confirmed by simulations, we have demonstrated that the protection performance (i.e. erasure resiliency and average decoding delay) of both approaches are equivalent, not only asymptotically but also in finite length conditions.

In fact the key differences between these approaches become apparent when applying them in practical systems. Such metrics as the simplicity of the solution, the number of packets processed, the maximum memory requirements, the number of FEC encoding and decodings, as well as the system of linear equations complexity (number of variables) are in favor of the GOE approach.

*Object Bundle Protection:* we considered the use of PET, more precisely an extension called Universal Object Delivery (UOD), and GOE in situations where one needs to send a bundle of small object (e.g. files). If both solutions can address this need, we showed that once again the GOE scheme is highly recommendable for practical realizations. This is mostly due to the lack of flexibility of the PET/UOD approach. For instance the limited size of a packet creates an upper bound to the number of objects that can be considered together (e.g. UOD limits this number to 255), the symbol size has a coarse granularity (e.g. UOD requires symbols to be multiple of 4 bytes when used with RaptorQ codes) which can create rounding problems with certain sets of objects (i.e. the actual packet size may be significantly shorter than the target, and/or the actual code rate significantly different than its target). GOE has no such constraints. In particular GOE offers the possibility to adjust the packet interleaving to the use-case and channel erasure features. One can easily trade robustness in front of long erasure bursts for very short decoding delays of high priority objects and low memory requirements, which can be a key asset in case of small, lightweight terminals or timely delivery services. This feature may be sufficiently important to justify by itself the use of a GOE FEC Scheme [55].

### 6.3.3. *Distributed compressed sensing*

**Participants:** Aline Roumy, Velotiaray Toto-Zarasoa.

This work has been performed in collaboration with E. Magli and G. Coluccia (Politecnico di Torino) in the framework of the FP7 IST NOE NEWCOM++ (Jan. 2008 - Apr. 2011). A new lossy compression scheme for distributed and sparse sources under a low complexity encoding constraint has been proposed in [26]. This problem naturally arises in wireless sensor networks. Indeed, nodes of a sensor network may acquire temperature readings over time. The temperature may vary slowly over time, and hence consecutive readings have similar values. However, they also have inter-sensor correlation, as the sensors may be in the same room, in which the temperature is rather uniform. The question hence arises of how to exploit intra- and inter-sensor correlations without communication between the sensors and with a low complexity acquisition process in order to save energy consumption at the sensor. Therefore, we consider *continuous*, *correlated*, *distributed* and *sparse* (in some domain) sources and perform *lossy universal* compression under a *low encoding complexity* constraint.

In order to meet the low complexity encoding constraint, the encoding stage is performed by a lossy distributed compressed sensing (CS). More precisely, the proposed architecture is based on the joint use of CS to capture memory of a signal, and DSC to take advantage of inter-sensor correlations. First, we showed that the resilience of CS to quantization error also holds in the distributed setup. Moreover, the optimal number of measurements can be chosen as the one guaranteeing (close-to-)perfect reconstruction. In addition, using joint decoding, dequantization and reconstruction techniques allows to boost performance even further. The joint use of CS and DSC allows to save 1.18 bit per source sample for the same PSNR quality w.r.t. the non-distributed CS scheme. Compared to the DSC scheme (without CS), we observe a gain increasing with the rate for the same PSNR quality. All these results makes the proposed scheme an attractive choice for environments such as sensor networks, in which the devices performing acquisition and processing are severely constrained in terms of energy and computations.

### 6.3.4. *Super-resolution as a communication tool*

**Participants:** Marco Bevilacqua, Christine Guillemot, Raul Martinez-Noriega, Aline Roumy.

In 2011, we started a new collaboration in the framework of the Joint INRIA/Alcatel Lucent lab. In this work, carried out with M-L. Alberi (Alcatel Lucent), we proposed a novel technique capable of producing a high-resolution (HR) image from a single low-resolution (LR) image. This method that belongs to the class of single-image super-resolution (SR), offers the promise of overcoming the inherent limitations of the video acquisition and transmission systems. More precisely, one can think of sending a low resolution video to adapt

to the complexity constraint of the encoder and/or the bandwidth limitation of the network, while the decoder reconstructs a high-resolution video.

As a first step toward the more ambitious goal of compressing video through SR, we proposed a novel method for single-image super-resolution based on a neighbor embedding technique. Each low-resolution input patch is approximated by a linear combination of nearest neighbors taken from a dictionary. This dictionary stores low-resolution and corresponding high-resolution patches taken from natural images and is thus used to infer the HR details of the super-resolved image. The entire neighbor embedding procedure is carried out in a feature space. Features which are either the gradient values of the pixels or the mean-subtracted luminance values are extracted from the LR input patches, and from the LR and HR patches stored in the dictionary. The algorithm thus searches for the $K$ nearest neighbors of the feature vector of the LR input patch and then computes the weights for approximating the input feature vector. The so-obtained weights are finally used to compute a linear combination of the corresponding HR patches, which yields the super-resolved image. The use of a positive constraint for computing the weights of the linear approximation is shown to have a more stable behavior than the use of sum-to-one constraint and lead to significantly higher PSNR values for the super-resolved images.

# 7. Contracts and Grants with Industry

## 7.1. Contracts with Industry

### 7.1.1. *Contract with Astrium on compression of satellite images*
**Participants:** Jeremy Aghaei-Mazaheri, Christine Guillemot, Claude Labit.

- Title : Comression of satellite images.
- Partners : Astrium, Inria-Rennes.
- Funding : Astrium.
- Period : Oct.11-Sept.14.

This contract with Astrium (starting in Oct. 2011) addresses the problem of compression of video signals captured from a geostationary satellite. The focus will be on the spatio-temporal modelling of scenes captured by the satellite in order to develop a compact representation taking advantage of the high redundancy present in the video of very high resolution and caracterized by low motion.

### 7.1.2. *Collaboration with Alcatel on robust video compression*
**Participants:** Marco Bevilacqua, Christine Guillemot, Laurent Guillo, Aline Roumy, Velotiaray Toto-Zaratosoa.

- Title: Self adaptive video codec
- Funding: Joint research laboratory between INRIA and Alcatel
- Period: Oct. 2010 - May. 2011.

In the framework of the joint research lab between Alcatel-Lucent and INRIA, we participate in the ADR (action de recherche) Selfnets (or Self optimizing wireless networks). More precisely, we collaborate with the Alcatel-Lucent team on an adaptive video codec. The goal is to design a video codec, which is able to adapt to the existing underlying transport network and/or to the complexity constraint of the encoder. Therefore the video codec has to include

- Means at the encoder to adapt dynamically the output bitrate to the estimated channel throughput and to the effective transport QoS while maintaining the video quality requirements.
- Means at the decoder to be resilient to any remaining packet losses.

## 7.2. Grants with Industry

### 7.2.1. CIFRE contract with Orange on 3D scene analysis

**Participants:** Christine Guillemot, Mouid Keskes, Olivier Le Meur.

- Title : 3D scenes analysis.
- Research axis : § 6.1.2.
- Partners : Orange Labs, Inria-Rennes.
- Funding : Orange Labs.
- Period : Dec.11-Nov.14.

This contract with Orange labs. (starting in Dec. 2011) aims at extracting, identifying and removing objects in a 3D unconstrainted environment. To reach this objective, several views of the same scene will be used. Spatio-temporal segmentation of 3D video content will be developped.

### 7.2.2. CIFRE contract with Orange on 3D quality assessment

**Participants:** Darya Khaustova, Olivier Le Meur.

- Title : Objective Evaluation of 3D Video Quality.
- Research axis : § 6.1.3.
- Partners : Orange Labs, Inria-Rennes.
- Funding : Orange Labs.
- Period : Dec.11-Nov.14.

This contract with Orange labs. (starting in Dec. 2011) aims at developping a video quality metric for 3D content. The usage of 3D video is expected to increase in the next years. In order to ensure a good QoE (Quality of Experience), the 3D video quality must be monitored and accuraly measured. The goal of this thesis is to study objective measures suitable for estimating 3D video quality. A comparison with ground truth as well as with the state-of-the-art 2D metrics should be carried out. To be as effective as possible, the feature of the human visual system should be taken into account.

### 7.2.3. CIFRE contract with Thomson/Technicolor on sparse modelling of spatio-temporal scenes

**Participants:** Safa Cherigui, Christine Guillemot.

- Title : Sparse modeling of spatio-temporal scenes
- Research axis : § 6.1.4.
- Partners : Thomson, Irisa/Inria-Rennes.
- Funding : Thomson, ANRT.
- Period : Nov.09- Oct.12.

This CIFRE contract concerns the Ph.D of Safa Cherigui. The objective is to investigate texture and scene characterization methods and models based on concepts of spatio-temporal epitomes and/or signatures for different image processing problems focusing in particular on video compression and editing. A novel method has been developed for constructing the epitome representation of an image. The epitome has been used for image compression showing significant performance gain with respect to H.264 Intra prediction modes.

# 8. Partnerships and Cooperations

## 8.1. National Initiatives

### 8.1.1. ANR-PERSEE

**Participants:** Josselin Gauthier, Christine Guillemot, Laurent Guillo, Olivier Le Meur.

- Title : Perceptual coding for 2D and 3D images.
- Research axis : § 6.1.2, 6.1.3.
- Partners : IRCCYN-Polytech Nantes, INSA-Rennes, Telecom Paris Tech.
- Funding : ANR.
- Period : 10/2009-09/2012

The objective of the project is to develop perceptually driven coding solutions for mono-view and multi-view video. TEMICS contributes on different problems relevant for mono-view and multi-view video coding: visual attention modeling (see Section 6.1.3), on texture synthesis and inpainting for both 2D and 3D content. Several methods for 2D image inpainting and 2D/3D inpainting to handle disocclusions in virtual view synthesis have been developed (see Sections 6.1.2. A computational model for 3D content has also been studied (see Section 6.1.3)

### 8.1.2. ANR-TCHATER

**Participant:** Jean-Jacques Fuchs.

The RNRT project TCHATER (Terminal Coherent Hétérodyne Adaptatif TEmps Reel) whose coordinator is Alcatel has started in January 2008. Its aim is to fully implement coherent detection in an optical fibers transmission systems, with among others the real time implementation on dedicated FPGAs. TEMICS has collaborated with ASPI on the design of solutions to adapt the extremely high channel rate, 4 ADC (analog-to-digital converters) and to accommodate the FPGA to the output rate, as well as temporal multiplexing of order 40. The project ended successfully in June 2011 with an operational prototype consisting in 4 analog-to-digital converters followed by 4 (Stratix IV) FPGAs implementing in real time the Chromatic Dispersion annihilating filters, the Constant Modulus Algorithm thau equalizes the channel, the Carrier Frequency Estimators and the Carrier Phase Estimators.

### 8.1.3. ANR-ARSSO

**Participants:** Mounira Ebdelli, Christine Guillemot, Laurent Guillo, Aline Roumy.

- Title : Adaptable, Robust, Streaming SOlutions.
- Partners : INRIA/Planète, TESA-ISAE, CEA-LETI/LNCA, ALCATEL LUCENT BELL LABS, THALES Communications, EUTELSAT SA.
- Funding : ANR.
- Period : 06/2010-11/2013

The ARSSO project focuses on multimedia content communication systems, characterized by more or less strict real-time communication constraints, within highly heterogeneous networks, and toward terminals potentially heterogeneous too. It follows that the transmission quality can largely differ in time and space. The solutions considered by the ARSSO project must therefore integrate robustness and dynamic adaptation mechanisms to cope with these features. The overall goal is to provide new algorithms, develop new streaming solutions and study their performances. TEMICS contributes on the development and improvement of scalable video coding techniques and components to make the video codec robust to losses. More specifically loss concealments methods will be developed.

## 8.2. European Initiatives

### *8.2.1. IST-NEWCOM++*

**Participants:** Christine Guillemot, Aline Roumy.

Program: VII Framework Program, Objective ICT-2007.1.1

Project acronym: IST-NEWCOM++

Project title:NEWCOM: Network of Excellence in Wireless Communication.

Duration: Jan. 2008 - Apr. 2011.

Coordinator: Politecnico Di Torino

Other partners: 17 partners across Europe

Abstract:

The NEWCOM++ project proposal (Network of Excellence in Wireless COMmunication) intends to create a trans-European virtual research centre on the topic "The Network of the Future". It was submitted to Call 1 of the VII Framework Program under the Objective ICT-2007.1.1: The Network of the Future, mainly in its target direction "Ubiquitous network infrastructure and architectures". We participate in the workpackage WPR7 - Joint source and channel co-decoding which we coordinate together with the task TR7.3 Tools for multi-terminal JSCC/D. WPR7 addresses issues related to the robust transmission of multimedia, and essentially video, over wireless channels (possibly terminating a wired IP network). In this framework, we proposed a novel distributed compressed sensing algorithm (see Section ADD REF for details) in collaboration with E. Magli and G. Coluccia (Politecnico di Torino). A paper has been published at the EUSIPCO 2011 conference and awarded the 2011 Best Paper Award "Francesco Carassa".

# 9. Dissemination

## 9.1. Animation of the scientific community

- C. Guillemot is associate editor of the Eurasip International Journal on Image Communication.

- C. Guillemot is member of the award committee of the Eurasip Image communication journal.

- C. Guillemot is member of the Selection and Evaluation Committee of the << Pôle de Compétitivité >> Images and Networks of the Region of Ouest of France.

- C. Guillemot has been a member of the technical program committees of the international conferences IEEE-MMSP 2011, EUSIPCO 2011, and of the International Workshop on Acoustics and Video Coding and Communication, held jointly with IEEE-ICME 2011, Barcelona.

- C. Guillemot is member of the << commission personnels >> in charge of the postdoc recrutments.

- C. Labit is member of the coordination group of the Allistene alliance devoted to ICT.

- C. Labit is president of the Rennes-Atalante Science Park and of the start-up incubator Emergys.

- C. Labit is member of the GRETSI association board.

- C. Labit is the Scientific Board chairman of Rennes 1 University (since June 1st, 2008).

- C. Labit is president of Rennes-Atalante Science Park.

- O. Le Meur has been member of the technical program committee of the international workshop on << Multimedia Quality of Experience: Modeling, Evaluation, and Directions >>, held in conjunction with the IEEE Int. Symp. on Multimedia (ISM2011, http://ism.eecs.uci.edu/).

- A. Roumy has been a member of the technical program committee of the international conference EUSIPCO 2011.

## 9.2. Standardization and Patents

- Submission of an Internet-draft to the IETF/Reliable Multicast Transport (RMT) Working Group: In March 2011, Qualcomm submitted a new FEC scheme to the RMT working group, << Universal Object Delivery (UOD) >> using RaptorQ, that aims at adding Unequal Erasure Protection (UEP) capabilities to their RaptorQ AL-FEC solution. This approach has been analyzed and compared to a new approach, we proposed (Generalized Object Encoding (GOE)), via simulations and modeling. A new Internet-Draft has been presented in July 2011 during the IETF meeting.

- Presentation of several contributions to the ISO/ITU HEVC (High efficiency Video Coding) initiative [44],[45],[49],[46],[48],[47],[43].

- S. Chérigui, C. Guillemot, D. Thoreau, P. Guillotel, " Epitome-based image compression ", joint patent Technicolor/INRIA, filed in July 2011.

## 9.3. Teaching

- J.J. Fuchs gives the following courses:
    - Master Research 2 SISEA: Optimization and sparse representations
    - Master of Science in Electronics and Telecommunications in the Joint International Program of the University of Rennes 1 and the SouthEast University of China (Nanjing): Advanced Signal Processing in the International

- C. Guillemot gives the following courses:
    - Engineering degree, Enic, Villeneuve-d'Ascq: Video communication (13.5 hours)
    - Master Recherche 2 Informatique (Computer Science), Univ. of Rennes 1: Image and video compression, (12 hours)
    - Master Recherche 2 SISEA, University Rennes 1, Image and video compression, (22 hours)

- L. Guillo gives the following courses:
    - Engineering degree, DIIC, University Rennes 1, Video streaming;
    - Master, Network Engineering, university of Rennes 1;

- O. Le Meur gives the following courses:
    - Engineering degree Diic3-INC, Visual communication, 65 hours, Univ. of Rennes 1, France
    - Engineer degree ESIR2, Image Processing, video analysis and compression, 54 hours, Univ. of Rennes 1, France
    - Master 2 Psychologie de la cognition, Psychologie des processus cognitifs, Mesure et modelisation: Selective visual attention: from experiments to computational models, 9 hours, University of Paris 8
    - Master 2 MITIC: Acquisition/Image Processing/Compression, 15 hours, Univ. of Rennes 1, France

- A. Roumy gives the following courses:
    - Magistère program: Information Theory, Computer science and telecommunications, 9 hours, Ecole Normale Supérieure de Cachan, Ker Lann campus, France;
    - Engineering degree, DIIC, Video streaming, 3 hours, Univ. of Rennes 1, France.
    - Master 2 MITIC, Acquisition/Image processing/Compression, 11 hours, Univ. rennes 1.

PhD & HdR :

PhD : Ana Charpentier, Identification de copies de documents multimédia grâce aux codes de Tardos, University of Rennes 1, 21 Oct. 2011, T. Furon and C. Fontaine

PhD : Jonathan Taquet, Techniques avancées pour la compression d'images médicales, University of Rennes 1, 21 Oct. 2011, C. Labit (HDR)

PhD : Mehmet Turkan, New texture synthesis methods: application to image prediction and inpainting, University of Rennes 1, 19 Dec. 2011, C. Guillemot (HDR)

PhD and HDR committees outside the team:

- J.J. Fuchs has been member of the jury of the PhD thesis of:
  - P. Sudhakara Murthy , University of Rennes 1, 21/02/2011
  - S. Yameogo, University of Rennes 1, 30/09/2011
  - Q. Duong, University of Rennes 1, 15/11/2011

- C. Guillemot has been member of the jury of the PhD thesis of:
  - L. Blondé, INSA-University of Rennes 1, 19/01/2011
  - Ho Lee, Liris, University Claude Bernard Lyon 1, 22/06/2011
  - V. Thirumalai, EPFL, Lausanne, 07/12/2011

- C. Labit has been member of the jury of the PhD thesis of:
  - M. Moinard, TelecomParistech, 01/07/2011

- O. Le Meur has been member of the jury of the PhD thesis of:
  - Matia Pizzoli, University of Rome, Nov. 2011.

- A. Roumy has been examiner of the PhD thesis of:
  - I. Hadj Kacem, Ecole Nationale Ingénieur de SFAX, Sfax, Tunisia, July 6th, 2011.

# 10. Bibliography

## Major publications by the team in recent years

[1] V. CHAPPELIER, C. GUILLEMOT. *Oriented wavelet transform for image compression and denoising*, in "IEEE Trans. on Image Processing", Oct. 2006, vol. 15, n⁰ 10, p. 2892-2903.

[2] T. COLLEU, S. PATEUX, L. MORIN, C. LABIT. *A polygon soup representation for multiview coding*, in "Journal of Visual Communication and Image Representation", Feb 2010, p. 1–32, http://hal.inria.fr/hal-00457634.

[3] J. J. FUCHS. *A robust matched detector*, in "IEEE Trans. on Signal Processing", Nov. 2007, vol. 55, n⁰ 11, p. 5133-5142.

[4] C. GUILLEMOT, A. ROUMY. *in book "Distributed source coding"*, in "Chapter entitled "Towards constructive Slepian-Wolf coding schemes""", Elsevier Inc., 2008.

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

[6] H. JEGOU, C. GUILLEMOT. *Robust multiplexed codes for compression of heterogeneous data*, in "IEEE Trans. on Information Theory", Apr. 2005, vol. 51, n⁰ 4, p. 1393-1403.

[7] O. LE MEUR, P. LE CALLET, D. BARBA. *Predicting visual fixations on video based on low-level visual features*, in "Vision Research", Sep. 2007, vol. 47, n<sup>o</sup> 19, p. 2493-2498.

[8] O. LE MEUR, P. LE CALLET, D. BARBA, D. THOREAU. *A coherent computational approach to model the bottom-up visual attention*, in "IEEE Trans. On PAMI", May 2006, vol. 28, n<sup>o</sup> 5, p. 802-817.

[9] L. OISEL, E. MEMIN, L. MORIN, F. GALPIN. *One-dimensional dense disparity estimation for three-dimensional reconstruction*, in "IEEE Trans. on Image Processing", Sept. 2003, vol. 12, n<sup>o</sup> 9, p. 1107–1119.

## Publications of the year

### Doctoral Dissertations and Habilitation Theses

[10] A. CHARPENTIER. *Identification de copies de documents multimédia grâce aux codes de Tardos*, University Rennes 1, Oct. 2011.

[11] V. JANTET. *Layered Depth Images for Multi-View Coding*, University Rennes 1, Jan. 2012.

[12] J. TAQUET. *Techniques avancées pour la compression d'images médicales*, Université de Rennes 1, Dec. 2011, http://tel.archives-ouvertes.fr/tel-00629429/fr/.

[13] M. TURKAN. *Nouvelles méthodes de synthèse de texture ; application à la prédiction et à l'inpainting d'images*, Université de Rennes 1, Dec. 2011, http://tel.archives-ouvertes.fr/tel-00635301/fr/.

### Articles in International Peer-Reviewed Journal

[14] B. FOLLET, O. LE MEUR, T. BACCINO. *New insights into ambient and focal visual fixations using an automatic classification algorithm*, in "i-Perception", 2011, http://hal.inria.fr/inria-00628069/en/.

[15] B. FUCHS, J. J. FUCHS. *Optimal Polarization Synthesis of Arbitrary Arrays with Focused Power Pattern.*, in "IEEE Trans. on Antenna and Propagation", 2011.

[16] V. JANTET, C. GUILLEMOT, L. MORIN. *Joint Projection Filling method for occlusion handling in Depth-Image-Based Rendering*, in "International journal on 3D research, special issue on "3DTV"", 2011, http://hal.archives-ouvertes.fr/hal-00628019/fr/.

[17] A. MARTIN, J. J. FUCHS, C. GUILLEMOT, D. THOREAU. *Sparse representations for spatial prediction and texture refinement*, in "special issue on "emerging techniques for high performance video coding" of the international journal on Visual Communication and Image representation (JVCI)", Jan. 2011.

[18] G. SOURIMANT, T. COLLEU, V. JANTET, L. MORIN, K. BOUATOUCH. *Toward automatic GIS-video initial registration*, in "Annals of Telecommunications", Mar. 2011, vol. 1, p. 1-13, http://hal.archives-ouvertes.fr/hal-00628167/en/.

[19] V. TOTO-ZARASOA, A. ROUMY, C. GUILLEMOT. *Maximum Likelihood BSC parameter estimation for the Slepian-Wolf problem*, in "IEEE Communications Letters", Feb. 2011, vol. 15, n<sup>o</sup> 2, p. 232–234 [*DOI :* 10.1109/LCOMM.2011.122810.102182], http://hal.inria.fr/inria-00628996/en/.

[20] V. TOTO-ZARASOA, A. ROUMY, C. GUILLEMOT. *Source modeling for Distributed Video Coding*, in "IEEE Trans. on Circuits and Systems for Video Technology", May 2011, http://hal.inria.fr/inria-00632708/en/.

[21] M. TURKAN, C. GUILLEMOT. *Image prediction based on neighbor embedding methods*, in "IEEE Trans. On Image Processing", 2011, http://hal.inria.fr/inria-00628054/en/.

[22] F. URBAN, B. FOLLET, C. CHAMARET, O. LE MEUR, T. BACCINO. *Medium Spatial Frequencies, a Strong Predictor of Salience*, in "Cogn Comput", 2011, vol. 3, p. 37-47, http://hal.inria.fr/inria-00628096/en/.

[23] J. ZEPEDA, C. GUILLEMOT, E. KIJAK. *Image Compression using the Iteration-tuned and Aligned Dictionary*, in "IEEE Journal of Selected Topics in Signal Processing, special issue on "Adaptive sparse representation of data and applications in signal and image processing"", Sept. 2011, vol. 5, p. 1061-1073.

### International Conferences with Proceedings

[24] E. BOSC, V. JANTET, M. PRESSIGOUT, L. MORIN, C. GUILLEMOT. *Bit-rate allocation for Multi-view video plus depth*, in "3DTV Conference", Antalya, Turquie, May 2011, http://hal.archives-ouvertes.fr/hal-00595580/en/.

[25] S. CHERIGUI, C. GUILLEMOT, D. THOREAU, P. PEREZ. *Epitome-based image compression*, in "IEEE Int. Conf. on MultiMedia Signal Processing (IEEE-MMSP)", May 2011, http://hal.inria.fr/inria-00628232/en/.

[26] G. COLUCCIA, E. MAGLI, A. ROUMY, V. TOTO-ZARASOA. *Lossy Compression of Distributed Sparse Sources: a Practical Scheme*, in "European Signal Processing Conference (EUSIPCO)", September 2011, http://hal.inria.fr/inria-00629001/en/.

[27] B. FOLLET, O. LE MEUR, T. BACCINO. *Features of ambient and focal fixations on natural visual scenes*, in "ECEM", Marseille, France, 2011.

[28] J. J. FUCHS. *Array and beamformer design for optimal directivity.*, in "45st Asilomar Conference on Signals, Systems and Computers", Pacific Grove, California, 2011.

[29] J. J. FUCHS. *Detection properties of some sparse representation approaches.*, in "45st Asilomar Conference on Signals, Systems and Computers", Pacific Grove, California, 2011.

[30] J. J. FUCHS. *Extension of the Global Matched Filter to structured groups of atoms: application to harmonic signals*, in "Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)", Prague, Tchèque, République, May 2011, p. 3996-3999, http://hal.archives-ouvertes.fr/hal-00628716/en/.

[31] J. J. FUCHS. *Spread representations*, in "45st Asilomar Conference on Signals, Systems and Computers", Pacific Grove, California, 2011.

[32] J. J. FUCHS. *Synthèse de diagramme de rayonnement : disposition optimale des éléments rayonnants*, in "GRETSI", Sep 2011, http://hal.archives-ouvertes.fr/hal-00628722/en/.

[33] J. GAUTIER, O. LE MEUR, C. GUILLEMOT. *DEPTH-BASED IMAGE COMPLETION FOR VIEW SYNTHESIS*, in "3DTVConf", ANTALYA, Turquie, 2011, http://hal.inria.fr/inria-00628084/en/.

[34] V. JANTET, C. GUILLEMOT, L. MORIN. *Object-based Layered Depth Images for improved virtual view synthesis in rate-constrained context*, in "IEEE International Conference on Image Processing (ICIP)", Sept. 2011, http://hal.archives-ouvertes.fr/hal-00628013/en/.

[35] O. LE MEUR, T. BACCINO, A. ROUMY. *Prediction of the Inter-Observer Visual Congruency (IOVC) and Application to Image Ranking*, in "ACM Multimedia", Phoenix, USA, 2011, http://hal.inria.fr/inria-00628077/en/.

[36] O. LE MEUR, J. GAUTIER, C. GUILLEMOT. *EXAMPLAR-BASED INPAINTING BASED ON LOCAL GEOMETRY*, in "ICIP", Brussel, Belgique, 2011, http://hal.inria.fr/inria-00628074/en/.

[37] O. LE MEUR. *Predicting Saliency Using Two Contextual Priors: the Dominant Depth and the Horizon Line*, in "ICME", Barcelona, Espagne, 2011, http://hal.inria.fr/inria-00628076/en/.

[38] O. LE MEUR. *Robustness and Repeatability of Saliency Models Subjected to Visual Degradations*, in "ICIP", Brussel, Belgique, 2011, http://hal.inria.fr/inria-00628072/en/.

[39] J. TAQUET, C. LABIT. *Compression sans perte et avec pertes contrôlées d'images de très haute définition : les lames virtuelles*, in "XXIIIeme Colloque du Groupe de Recherche et d'Etudes du Traitement du Signal et des Images (GRETSI 2011)", Sept. 2011, http://hal.inria.fr/inria-00632450/en/.

[40] M. TURKAN, C. GUILLEMOT. *Image prediction based on non-negative matrix factorization*, in "IEEE Int. Conf. on Acous. Speech and Signal Process. (ICASSP)", Prague, May 2011, http://hal.inria.fr/inria-00628057/en/.

[41] M. TURKAN, C. GUILLEMOT. *Online dictionaries for image prediction*, in "IEEE International Conference on Image Processing (ICIP)", Sept. 2011, http://hal.inria.fr/inria-00628058/en/.

[42] J. ZEPEDA, C. GUILLEMOT, E. KIJAK. *Image compression using the iteration-tuned and aligned dictionary*, in "IEEE Int. Conf. on Acous. Speech and Signal Process. (ICASSP)", Prague, May 2011.

### Research Reports

[43] L. GUILLO, R. BOITARD. *AHG18: Cross-check report of Weighted Prediction, proposal JCTVC-F265*, Joint Collaborative Team on Video Coding, Jul 2011, JCTVC-F436.

[44] L. GUILLO, R. BOITARD. *CE10: Cross-check report from INRIA on number of intra prediction directions*, Joint Collaborative Team on Video Coding, Jan 2011, JCTVC-D063.

[45] L. GUILLO, R. BOITARD. *CE2-subset 4: Cross-check report of Technicolors proposal on Geometry adaptive block partitioning Simplification from INRIA*, Joint Collaborative Team on Video Coding, Jan 2011, JCTVC-D173.

[46] L. GUILLO, R. BOITARD. *CE2: Cross-check report of Non Rectangular Motion Partitioning (NMRP) + Overlapped Block Motion Compensation (OBMC), proposal JCTVC-E373*, Joint Collaborative Team on Video Coding, Mar 2011, JCTVC-E252.

[47] L. GUILLO, R. BOITARD. *CE6.b: Cross-check report of Test 3 : Interaction between SDIP and MDCS, proposal JCTVC-F197*, Joint Collaborative Team on Video Coding, Jun 2011, JCTVC-F231.

[48] L. GUILLO, R. BOITARD. *Cross Verification Report of Canon proposal JCTVC-E219 on Robust solution for the AMVP parsing issue*, Joint Collaborative Team on Video Coding, Mar 2011, JCTVC-E252.

[49] L. GUILLO, R. BOITARD, T. POIRIER. *Integration into the TMuC of an Intra Prediction based on a linear combination of Template Matching predictors*, Joint Collaborative Team on Video Coding, Jan 2011, JCTVC-D193.

[50] T. GUIONNET, L. GUILLO. *CE6d: Cross-check report on Intra prediction with secondary boundary (Test 7) (JCTVC-F456)*, Joint Collaborative Team on Video Coding, Nov 2011, JCTVC-G561.

[51] T. GUIONNET, L. GUILLO. *Non-CE6: Intra prediction based on weighted template matching predictors (WTM)*, Joint Collaborative Team on Video Coding, Nov 2011, JCTVC-G598.

[52] T. GUIONNET, L. GUILLO. *Non-CE6a : Cross check on use of chroma phase in LM mode (JCTVC-G245)*, Joint Collaborative Team on Video Coding, Nov 2011, JCTVC-G488.

[53] V. ROCA, A. ROUMY, B. SAYADI. *The Generalized Object Encoding (GOE) Approach for the Forward Erasure Correction (FEC) Protection of Objects and its Application to Reed-Solomon Codes over GF($2^8$)*, July 2011, http://tools.ietf.org/html/draft-roca-rmt-goe-fec-00.

[54] V. ROCA, A. ROUMY, B. SAYADI. *The Generalized Object Encoding (GOE) LDPC-Staircase FEC Scheme*, October 2011, http://tools.ietf.org/id/draft-roca-rmt-goe-ldpc-00.txt.

[55] A. ROUMY, V. ROCA, B. SAYADI, R. IMAD. *Unequal Erasure Protection and Object Bundle Protection with the Generalized Object Encoding Approach*, INRIA, Jul. 2011, n^o RR-7699, submitted to Infocom 2012, http://hal.inria.fr/inria-00612583/en/.