IN PARTNERSHIP WITH:
CNRS

Université Rennes 1

Activity Report 2012

Project-Team SIROCCO

Analysis, representation, compression and communication of visual data

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

RESEARCH CENTER
Rennes - Bretagne-Atlantique

THEME
Vision, Perception and Multimedia Understanding

Table of contents

Project-Team SIROCCO

Keywords: Image Processing, Signal Processing, Video, Sparse Representations

Creation of the Project-Team: January 01, 2012 .

1. Members

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

Faculty Member
Olivier Le Meur [Assistant Professor, ESIR, University of Rennes 1]

Engineer
Laurent Guillo [IR, CNRS]

PhD Students
Martin Alain [CIFRE contract with Technicolor, since 1st Oct. 2012]
Marco Bevilacqua [Inria grant, Joint Inria/Alcatel Lucent lab.]
Jeremy Aghaei Mazaheri [Inria grant, contract with Astrium]
Safa Cherigui [CIFRE contract with Technicolor, until 30th Nov. 2012]
Nicolas Dhollande [CIFRE contract with Thomson Video Networks, since 1st Oct. 2012]
Bihong Huang [CIFRE contract with Orange, since April 2012]
Darya Khaustova [CIFRE contract with Orange]
Mounira Ebdelli [Inria grant, ANR-ARSSO project]
Josselin Gautier [Ministry of research grant, until 30th Sept. 2012]

Post-Doctoral Fellow
Raul Martinez Noriega [Inria]

Visiting Scientist
Zhi Liu [EU-IIF Grant, University of Rennes 1, from 1st Aug. 2012 till 31st July2014]

Administrative Assistant
Huguette Béchu [TR, Inria, shared with Fluminance and Serpico]

Others
Thomas Guionnet [Inria, until 30 Sept. 2012]
Ronan Le Boulch [Inria, Joint Inria/Alcatel Lucent Lab and ANR-ARSSO project]
Fabien Racapé [Inria, ANR-PERSEE project, since May 2012]
Alan Bourasseau [since 1st Sept. 2012]

2. Overall Objectives

2.1. Introduction

The goal of the SIROCCO project-team is the design and development of algorithms and practical solutions in the areas of analysis, modelling, coding, and communication of images and video signals. The objective is to cover several inter-dependent algorithmic problems of the end-to-end transmission chain from the capturing, compression, transmission to the rendering of the visual data. The project-team activities are 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]
- Rendering, inpainting and super-resolution of visual data
- 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, SIROCCO, in partnership with industrial companies, seeks to promote its results in standardization (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, Thomson Video Networks) 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 and network adaptation, of advanced functionalities such as navigation, and of high quality rendering.

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 3D scenes, 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 considers various types of image modalities (medical and satellite images, natural 2D still and moving images, multi-view and multi-view plus depth video content).

2.3. Rendering, inpainting and super-resolution

This research axis addresses the problem of high quality reconstruction of various types of visual data after decoding. Depending on the application and the corresponding type of content (2D, 3D), various issues are being addressed. For example, to be able to render 3D scenes, depth information is associated with each view as a depth map, and transmitted in order to perform virtual view generation. 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. Inpainting techniques are also required in transmission scenarios, where packet losses result in missing data in the video after decoding. The design of efficient mono-view and multi-view super-resolution methods is also part of the project-team objectives to improve the rendering quality, as well as to trade-off quality against transmission rate.

2.4. 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 which we address for further progress in the 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.

2.5. 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.

2.6. Highlights of the Year

- The paper “Hybrid template and block matching algorithm for image intra prediction” by Safa Cherigui, Christine Guillemot, Dominique Thoreau, Philippe Guillotel and Patrick Perez has received one of the three best student paper awards at IEEE-ICASSP, Kyoto, March 2012 [21].
- The paper “Map-aided locally linear embedding methods for image prediction” by Safa Cherigui, Christine Guillemot, Dominique Thoreau, Philippe Guillotel and Patrick Perez has been among the 11 finalists (out of 500 student papers) for the best student paper award at IEEE Intl. Conf. on Image Processing, ICIP, Oct. 2012.

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 Human Visual System (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

Bayesian estimation, Expectation-Maximization, stochastic modelling

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

manifolds, locally linear embedding, non-negative matrix factorization, principal component analysis

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, on locally linear embedding (LLE) and on the "non negative matrix factorization" (NMF) 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 exist many solutions x to $Ax = b$. 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 ∞ 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

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

additive. Classical data representation methods like PCA or Vector Quantization (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 of data-dependent dimensionality reduction will be at the core of our visual data analysis and compression activities.

3.4. Perceptual Modelling

Saliency, visual attention, cognition

The human visual system (HVS) 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 process. 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 a sensitivity function such as the Contrast Sensitivity Function (CSF).

Our capacity of analysis is also related to our visual attention. Visual attention which is closely linked to eye movement (note that this attention is called overt while 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:

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

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

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

[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”, 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.

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 [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.

[7] T. M. Cover and 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] 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.

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 multi-view video refers to multiple video sequences captured by multiple video cameras and possibly by depth cameras. Associated with a view synthesis method, a multi-view video allows the generation of virtual views of the scene from any viewpoint. This property can be used in a large diversity of applications, including Three-Dimensional TV (3DTV), and Free Viewpoint Video (FTV). The notion of "free viewpoint video" refers to the possibility for the user to choose an arbitrary viewpoint and/or view direction within a visual scene, creating an immersive environment. Multi-view video generates a huge amount of redundant data which need to be compressed for storage and transmission. In parallel, the advent of a variety of heterogeneous delivery infrastructures has given momentum to extensive work on optimizing the end-to-end delivery QoS (Quality of Service). This encompasses compression capability but also capability for adapting the compressed streams to varying network conditions. The scalability of the video content compressed representation, its robustness to transmission impairments, are thus important features for seamless adaptation to varying network conditions and to terminal capabilities.

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

4.3. 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.

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: Laurent Guillo [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: Laurent Guillo [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: Laurent Guillo [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.

5.5. JPF-Joint Projection Filling

Participant: Fabien Racapé [contact person].

In the context of multi-view videos, this software generates virtual views of the scene from any viewpoint using a proposed method named Joint Projection Filling (JPF). The latter belongs to Depth-Image-Based Rendering (DIBR) methods, relying 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. The JPF method performs forward projection on depth map, using connectivity information to fill in disocclusions in a single step. Depth- based inpainting can then be used to fill in color disocclusions.

5.6. LDI builder

Participant: Fabien Racapé [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.

5.7. Visual Fixation Analysis

Participant: Olivier Le Meur [contact person].

From a set of fixation data and a picture, the software called Visual Fixation Analysis extracts from the input data a number of features (fixation duration, saccade length, orientation of saccade...) and computes an human saliency map. The software can also be used to assess the degree of similarity between a ground truth (eye fixation data) and a predicted saliency map. This software is dedicated to people working in cognitive science and computer vision. This software has been registered at the APP (Agence de Protection des Programmes).

5.8. ADT PICOVIN-P

Participants: Laurent Guillo [contact person], Thomas Guionnet.

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

During this year, the ADT first pursued its developments on Intra prediction in the context of the standardization initiative referred to as High Efficiency Video Coding (HEVC) and led by the Joint ITU/MPEG Collaborative Team on Video Coding (JCT-VC). HEVC is implemented as a test model, the HEVC test Model (HM) in which the ADT tools have been integrated. We then followed the standardization activities within the Joint Collaborative Team on 3D Video Coding Extension (JCT-3V). JCT-VC and JCT-3V have been both created by the ITU-T Study Group 16 (VCEG) and the ISO/IEC JTC 1/SC 29/WG 11 (MPEG). While JCT-VC aims at developing a new generation 2D video coding standard, JCT-3V aims at developing 3D extensions for video codecs, which are AVC (ATM) or HEVC (HTM) based.

As part of JCT-3V, we submitted several proposals related to the handling of the merge list of predictor candidates. They were about the re-ordering of the candidates in the list and about the addition of new candidates. Two of them have been accepted in the dedicated core experiment (CE) during the 1st JCT-3V meeting which was held in Stockholm in July 2012. An improved version related to the addition of candidates has been accepted in a CE during the 2nd JCT-3V meeting in Shanghai in October 2012. It will be integrated in the coming test model (HTM) and evaluated during the next meeting in Geneva in January 2013. During 2012, the ADT also took part in cross checks which aims at evaluating and testing tools studied in core experiments. As part of cross checks related to JCT-VC or JCT-3V the ADT ran 9 tests jointly with companies such as Canon, Huawei, HiSilicon and Nokia. The ADT Picovin-P started in October 2011 and lasted one year. 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

3D modelling, multi-view plus depth videos, Layered depth images (LDI), 2D and 3D meshes, epitomes, image-based rendering, inpainting, view synthesis

6.1.1. Computational modelling of visual attention

Participants: Josselin Gautier, Olivier Le Meur, Zhi Liu.

6.1.1.1. Time-dependent saliency map

The study related to the deployment of visual attention in 2D and 3D has been completed in 2012. The purpose of this study was to investigate whether or not there is a difference between eye movements recorded while observers viewed natural images in 2D and 3D conditions. 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. Based on these conclusions, a computational model of visual attention taking into account the temporal dimension has been designed. An Expectation-Maximisation (EM) algorithm has been used to infer the weight of different visual features (saliency, depth, center bias) over time. Results have been published in the journal Cognitive Computation.

A new study on a similar subject has started during the summer 2012. The purpose is again to investigate the influence of binocular disparity, scene complexity on visual scanpaths obtained in 2D and 3D viewing conditions. The main differences with the previous study are twofold. First, a new database of content has been designed. All parameters such as the amount of disparity are accurately mastered. Second is about the context of the study which deals with quality assessment of 3D video content.

6.1.1.2. Salient object detection

In 2012, Dr. Liu, who has joined the team in August for 2 years has started a study dealing with salient object detection. The goal is to extract automatically the most interesting object in an image or video sequence. The proposed approach is based on low-level visual features and extensively used a superpixel method. Starting from the superpixel representation of an image, the saliency measure of each superpixel is evaluated based on its global uniqueness and local contrasts with other superpixels. A saliency-directed region merging algorithm with a dynamic scale control scheme is then exploited to generate more meaningful regions. The region merging process is recorded using a Binary Partition Tree (BPT), in which each leaf node represents each superpixel and each non-leaf node represents each generated region during the region merging process. Finally, a node selection algorithm based on saliency density difference is used to select suitable nodes from BPT to form the salient object detection result. First experimental results on a public dataset (MSRA) are promising and demonstrate the effectiveness of the proposed approach.

6.1.2. Similarity metrics for image processing

Participants: Mounira Ebdelli, Christine Guillemot, Olivier Le Meur, Raul Martinez Noriega, Aline Roumy.

Several image processing problems addressed by the team (inpainting, loss concealment, super-resolution, denoising) require having patch objective similarity metrics as close as possible to ground truth visual similarity. The derivation of such metrics has been investigated along several directions. First, a performance analysis of the most used fidelity metrics (SSD, SSIM, two SSD-weighted Battacharya metrics) has been carried out to assess the perceptual similarities between patches. A statistical analysis of subjective tests has shown that some of these metrics (the SSD-weighted Battacharya) are more suitable than others to respect human decisions in terms of patch similarities. This conclusion has been confirmed with the results of Non Local means (NL-means) denoising algorithm which are highly sensitive to the used similarity metrics. The value of each pixel p in the blurred image is updated using a weighted average of the collocated pixels values in the most similar patches to the block centered on p. We show that SSD, which is the most used similarity metric, is not necessary the best correlated with the perceptual criteria.

Greedy algorithms for inpainting are based on the assumption of self-similarity within an image. A patch located on the boundary of the hole to be filled in, contains a known part and an unknown part. The known part is used to select other (completely known) patches and called exemplars. Then, these exemplars are used to reconstruct the unknown part of the patch being processed. Such an approach faces two main problems, decision of filling-in order and selection of good exemplars from which the missing region is synthesized. In [29], we proposed an algorithm that tackles these problems with improvements in the preservation of linear edges, and reduction of error propagation compared to well-known algorithms from the literature. Our improvement in the filling-in order is based on a combination of priority terms, previously defined, that better encourages the early synthesis of linear structures. The second contribution helps reducing the error propagation thanks to a better detection of outliers from the candidate patches carried. This is obtained with a new metric based on the Hellinger distance between the patches that incorporates the whole information of the candidate patches.

6.1.3. Epitome-based image representation

Participants: Safa Cherigui, Christine Guillemot.

This 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 in the literature, 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. The epitome is constructed from disjoint pieces of texture ("epitome charts") taken from the original image and a transform map which contains translational parameters. 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% on some images, including the rate cost of the epitome texture and of the transform map. The entire image can be reconstructed from the epitome texture with the help of the transform map.

6.2. Rendering, inpainting and super-resolution

image-based rendering, inpainting, view synthesis, super-resolution

6.2.1. Joint projection/filling method for virtual view synthesis

Participants: Christine Guillemot, Fabien Racapé.

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. State-of-the-art synthesizers use Depth-Image-Based Rendering (DIBR) techniques based on warping equations, which project a reference view onto a virtual viewpoint. In classical DIBR schemes, the rendering proceeds in several distinct steps, each one designed to solve a specific problem. First, the depth map is warped onto the virtual viewpoint and filtered with a median filter. The filtered depth map is then used in a backward warping of the virtual view (as illustrated in FIg.1). The 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, producing annoying artifacts, as shown in Fig. 2-(c).

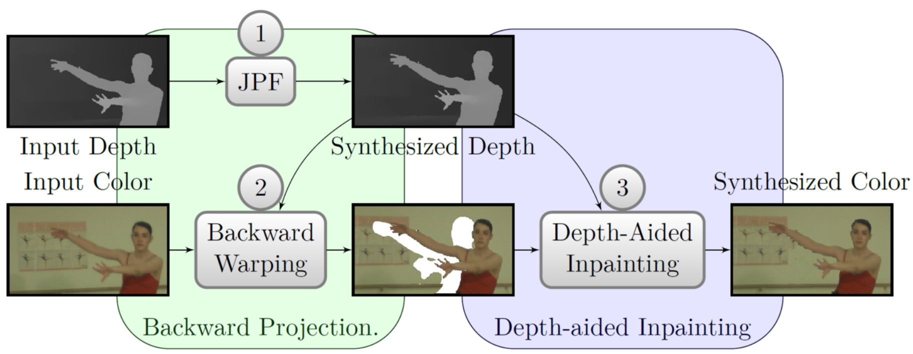

Figure 1. Virtual view generation chain, based on Joint Filling Projection. The depth map is jointly warped and inpainted. Depth-aided inpainting can be then used on disoccluded areas.

The proposed Joint Projection Filling (JPF) method performs forward projection, using connectivity information to fill in disocclusions in a single step. Applied on the depth map warping, JPF enables a depth-aided inpainting of color disocclusions after backward projection, as shown in Fig. 1. Fig. 2-(e) presents a resulting synthesis which contains less artifacts.

In the context of multi-view plus depth video coding (3D-HEVC standardization), inter view coding tools are added in the vein of temporal inter frame coding. We have tested our method as a projection tool for View Synthesis Prediction (VSP). However, the 3D-HEVC common test conditions, limited to rectified views as input, restrict the possible gains induced by efficient projection tools. Moreover, JPF outperforms other methods in synthesizing disoccluded areas with a good visual quality where VSP tools are not selected by MSE-based decision. JPF remains an efficient tool for extrapolating multi-view plus depth content with a minimum of artifacts on disoccluded areas.

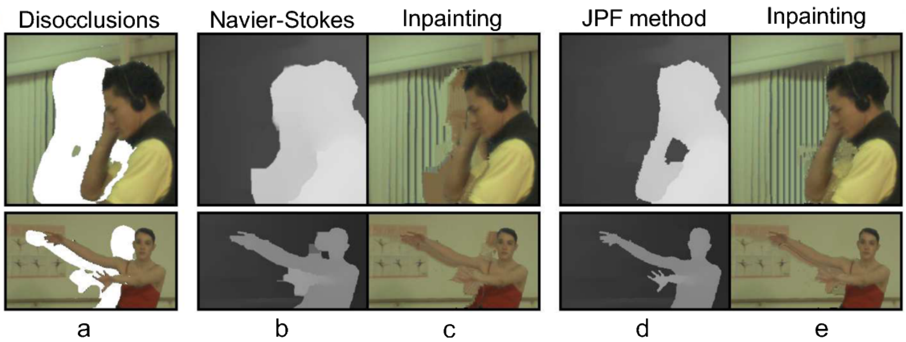

Figure 2. Disocclusion filling. (a) warped image before inpainting. Depth map inpainting: Navier-Stokes (b), JPF (d). Resulting depth-aided inpainting: Navier-Stokes (c), JPF (e).

6.2.2. Image inpainting using neighbor embedding and super-resolution

Participants: Mounira Ebdelli, Christine Guillemot, 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.

Image inpainting is a problem of texture synthesis. Given observations, or known samples in a spatial neighborhood, the goal is to estimate unknown samples of the patch to be filled in. Novel inpainting methods have been developed in the team along complementary directions: 1/- considering new priority functions exploiting the structure within the patch for defining the patch processing order; 2/- investigating various neighbor embedding techniques for estimating the unknown pixels; 3/- considering a coarse to fine multi-resolution approach where a low resolution version of the input image is first inpainted, this first step being followed by a a super-resolution based enhancement of the image.

Priority functions: Different priority functions using structure tensors and edge based information have been considered and their advantage over classical functions projecting isophote directions on the normal to the front line has been demonstrated.

Neighbor-embedding based inpainting: Examplar-based inpainting algorithms using various neighbor embedding techniques (LLE, LLE-LDNR, NMF with various solvers) have been developed. The methods have 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, or similarity weights (NLM) (see Fig. 3).

Figure 3. Inpainting results: Original image; Mask of the image to be inpainted; Inpainting results with examplar-based inpainting using similarity weights; Inpainting results with neighbor embedding (LLE-LDNR).

Super-resolution aided inpainting: A novel super-resolution aided inpainting approach has been introduced which consists in first inpainting a coarse version of the input image and then in a second step, using a hierarchical super-resolution algorithm, to recover the native resolution [28]. However, to be less sensitive to the setting of the inpainting methods, the low-resolution input picture is inpainted several times with different settings. Results are efficiently combined with a loopy belief propagation. A super-resolution algorithm is then applied to recover the details. Experimental results in a context of image editing, texture synthesis and 3D view synthesis demonstrate the effectiveness of the proposed method. Fig.4 show texture synthesis results obtained with this approach.

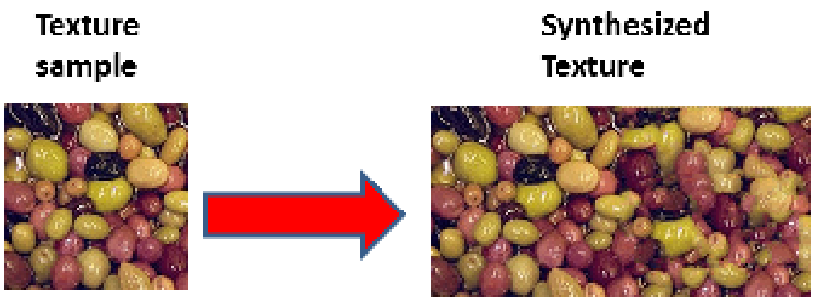

Figure 4. Texture synthesis results obtained with super-resolution aided inpainting.

6.3. Representation and compression of large volumes of visual data

sparse representations, data dimensionality reduction, compression, scalability, perceptual coding, rate-distortion theory

6.3.1. Multi-view plus depth video compression

Participants: Christine Guillemot, Thomas Guionnet, Laurent Guillo, Fabien Racapé.

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 represention 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. A novel object-based LDI representation which is more tolerant to compression artifacts, as well as being compatible with fast mesh-based rendering techniques has been developped.

The team has also studied motion vector prediction in the context of HEVC-compatible Multi-view plus depth (MVD) video compression. The HEVC compatible MVD compression solution implements a 6 candidate vector list for merge and skip modes. As part of the 3D video encoding, an inter–view motion vector predictor is added at the first position of this list. Our works show that this new list can be improved in optimizing the order of the candidates and in adding two more relevant candidates. When a merge or a skip mode is selected, a merge index is written in the bitstream. This index is first binarized using a unary code, then encoded with the CABAC. A CABAC context is dedicated to the first bin of the unary coded index while the remaining bins are considered as equiprobable. This strategy is efficient as long as the candidate list is ordered by decreasing index occurrence probability. However, this is not always the case when the inter-view motion vector predictor is added. To dynamically determine which candidate is the most probable, a merge index histogram is computed on the fly at the encoder and decoder side. Thus a conversion table can be calculated. It allows deriving the merge index to encode given the actual index in the list, and conversely, the actual index in the list given a decoded index. When using dynamic merge index, index re-allocation can happen at any time. Statistics of the first bin, which is encoded with CABAC, are modified. That is why a set of 6, one for each possible permutation of indexes, CABAC contexts dedicated to the first bin is defined. A bit rate gain of 0.1% for side views is obtained with no added complexity. These results are improved and reach 0.4% when additional CABAC contexts are used to take into account also the first three bins.

Candidates added by default in the merge list are not always the most relevant. As part of 3D video encoding using multiple rectified views, having a fine horizontal adjustment might be meaningful for efficient disparity compensated prediction. Therefore, we have proposed to replace some candidates in the merge list with candidates pointing to the base view and shifted by the horizontal offsets +4 and -4. To do so, the merge list is scanned to get among the first four candidates the first disparity compensated candidate. Once this vector found, the +4 and -4 offsets are added to its horizontal component and the two resulting vectors are inserted in the list two positions further if there is still room just after otherwise. With this improvement, a bit rate gain of 0.3% for side views is obtained with no added complexity.

6.3.2. Diffusion-based depth maps coding

Participants: Josselin Gautier, Olivier Le Meur.

A novel approach to compress depth map has been developed [26]. The proposed method exploits the intrinsic depth maps properties. Depth images indeed represent the scene surface and are characterized by areas of smoothly varying grey levels separated by sharp edges at the position of object boundaries. Preserving these characteristics is important to enable high quality view rendering at the receiver side. The proposed algorithm proceeds in three steps: the edges at object boundaries are first detected using a Sobel operator. The positions of the edges are encoded using the JBIG algorithm. The luminance values of the pixels along the edges are then encoded using an optimized path encoder. The decoder runs a fast diffusion-based inpainting algorithm which fills in the unknown pixels within the objects by starting from their boundaries.

6.3.3. Neighbor embedding for image prediction

Participants: Safa Cherigui, Christine Guillemot.

The problem of texture prediction 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. We have in 2010 and 2011 developed texture prediction methods as well as inpainting algorithms based on neighbor embedding techniques which come from the area of data dimensionality reduction [18], [31], [27]. The methods which we have more particularly considered are Locally Linear Embedding (LLE), LLE with Low-dimensional neighborhood representation (LDNR), and Non-negative Matrix Factorization (NMF) using various solvers.

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). In a prediction (compression) context, in order for the decoder to proceed similarly, the K nearest neighbors are found by computing distances between the known pixels in a causal neighborhood (called template) of the input block and the co-located pixels in candidate patches taken from a causal window. Similarly, the weights used for the linear approximation are computed in order to best approximate the template pixels. Although efficient, these methods suffer from limitations when the template and the block to be predicted are not correlated, e.g. in non homogenous texture areas. To cope with these limitations, we have developed new image prediction methods based on neighbor embedding techniques in which the K-NN search is done in two steps and aided, at the decoder, by a block correspondence map, hence the name Map-Aided Neighbor Embedding (MANE) method. Another optimized variant of this approach, called oMANE method, has also been introduced. The resulting prediction methods are shown to bring significant Rate-Distortion (RD) performance improvements when compared to H.264 Intra prediction modes (up to 44.75%) [13]. Figure 5 illustrates the prediction quality obtained with different neighbor embedding methods, as well as the encoder selection rate of the oMANE-based prediction mode. This method has been presented at the IEEE International ICIP conference and the paper has been selected among the 11 finalists (out of 500 student papers) for a best student paper award.

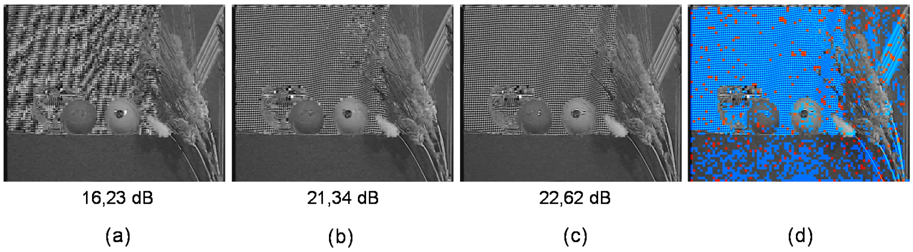

Figure 5. Spatial prediction for "Snook" with modes dynamically chosen according to a RD criterion with (a) H.264 Intra modes (High Profile), (b) LLE-based prediction, (c) Hybrid LLE-oMANE prediction and (d) selection rate of the two modes: LLE (red) and oMANE (blue).

6.3.4. Generalized lifting for video compression

Participants: Christine Guillemot, Bihong Huang.

This research activity is carried out in collaboration with Orange labs (Felix Henry) and UPC (Philippe Salembier) in Barcelona. The objective is to design new algorithmic tools for efficient loosless and lossy compression using generalized lifting concepts. The generalized lifting is a framework which permits the creation of nonlinear and signal probability density function (pdf) dependent and adaptive transforms. The use of such adaptive transforms for efficient coding of different HEVC syntax elements is under study.

6.3.5. Dictionary learning methods for sparse coding of satellite images

Participants: Jeremy Aghaei Mazaheri, Christine Guillemot, Claude Labit.

In the context of the national partnership Inria-Astrium, we explore novel methods to encode sequences of satellite images with a high degree of restitution quality and with respect to usual constraints in the satellite images on-board codecs. In this study, a geostationary satellite is used for surveillance and takes sequences of images. Then these pictures are stabilized and have to be compressed on-board before being sent to earth. Each picture has a high resolution and so the rate without compression is very high (about 70 Gbits/sec) and the goal is to achieve a rate after compression of 600 Mbits/sec, that is a compression ratio more than 100. On earth, the pictures are decompressed with a high necessity of reconstruction quality, especially for moving areas, and visualized by photo-interpreters. That is why the compression algorithm requires here a deeper study.The first stage of this study is to develop dictionary learning methods for sparse representations and coding of the images. These representations are commonly used for denoising and more rarely for image compression.

Sparse representation of a signal consists in representing a signal $y \in \Re^n$ as a linear combination of columns, known as atoms, from a dictionary matrix. The dictionary $D \in \Re^{n \times K}$ is generally overcomplete and contains K atoms. The approximation of the signal can thus be written $y \approx Dx$ and is sparse because a small number of atoms of D are used in the representation, meaning that the vector x has only a few non-zero coefficients. The choice of the dictionary is important for the representation. A predetermined transform matrix, as overcomplete wavelets or DCT, can be chosen. Another option is to learn the dictionary from training signals to get a well adapted dictionary to the given set of training data. Previous studies demonstrated that dictionaries have the potential to outperform the predetermined ones. Various advanced dictionary learning schemes have been proposed in the literature, 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).

Recently, the idea of giving relations between atoms of a dictionary appeared with tree-structured dictionaries. Hierarchical sparse coding uses this idea by organizing the atoms of the dictionary as a tree where each node corresponds to an atom. The atoms used for a signal representation are selected among a branch of the tree. The learning algorithm is an iteration of two steps: hierachical sparse coding using proximal methods and update of the entire dictionary. Even if it gives good results for denoising, the fact to consider the tree as a single dictionary makes it, in its current state, not well adapted to efficiently code the indices of the atoms to select when the dictionary becomes large. We introduce in this study a new method to learn a tree-structured dictionary offering good properties to code the indices of the selected atoms and to efficiently realize sparse coding. Besides, it is scalable in the sense that it can be used, once learned, for several sparsity constraints. We show experimentally that, for a high sparsity, this novel approach offers better rate-distortion performances than state-of-the-art "flat" dictionaries learned by K-SVD or Sparse K-SVD, or than the predetermined overcomplete DCT dictionary. We recently developped a new sparse coding method adapted to this tree-structure to improve the results. Our dictionary learning method associated with this sparse coding method is also compared to other methods previously introduced in the recent litterature such as TSITD (Tree-Structured Iteration-Tuned Dictionary) or BITD (Basic Iteration-Tuned Dictionary) algorithms.

6.4. Distributed processing and robust communication

information theory, stochastic modelling, robust detection, maximum likelihood estimation, generalized likelihood ratio test, error and erasure resilient coding and decoding, multiple description coding, Slepian-Wolf coding, Wyner-Ziv coding, information theory, MAC channels

6.4.1. Loss concealment based on video inpainting

Participants: Mounira Ebdelli, Christine Guillemot, Ronan Le Boulch, 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 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. We have considered two neighbor embedding approaches namely locally linear embedding (LLE) and non-negative matrix factorization (NMF), in a way that each patch of the target region is inpainted with the best estimation provided using template matching, LLE and NMF. The motion confidence introduced in the neighbor embedding improves the robustness of the algorithm with limiting the error propagation effects which may result from uncertainties on the motion information of the unknown pixels to be estimated. Evaluations of the algorithm in a context of video editing (object removal) show natural looking videos with less annoying artifacts [24].

This approach has then been adapted to the context of loss concealment that is to estimate unknown pixels after decoding when the corresponding transport packets have been lost on the transmission network. For this purpose, a preprocessing step is proposed to estimate the motion information of each corrupted block using Bilinear Motion Field Interpolation (BMFI) before inpainting the texture. The BMFI method computes the missing motion vector of each pixel in the lost block as a weighted combination of motion vectors of neighboring blocks. The estimated motion information is also used to limit the search space for the best matching patches in a motion-compensated window. Experiments of the proposed approach on several videos show a PSNR average gain about 2dB compared to state-of-art methods [25]. The next step will be to assess the performance of the approach in a context of free moving camera videos. To deal with this problem, we propose to consider a panoramic image mosaics in order to estimate the background of the video before inpainting the missing part of the foreground objects.

6.4.2. Unequal Erasure Protection and Object Bundle Protection

Participant: Aline Roumy.

In 2011, we started a new collaboration on Unequal Erasure Protection (UEP) and Object Bundle Protection in the framework of the joint research lab Inria–Alcatel Lucent and the ANR ARSSO project. Protection is usually obtained by adding Forward error correction (FEC) to the object (or data) to be transmitted. However, when the object contains information with different importance levels (as in a video bitstream), providing a protection adapted to the importance of each subpart of the object, helps reducing the encoded bitrate. To implement UEP, traditional transport protocols based on FEC Schemes need to split the original object into say two sub-objects, one per important class, and to submit each sub-object separately to the FEC Scheme. This requires extra logic for splitting/gathering the data. A companion problem, is the case where the object size is smaller than the packetsize. In this case, FEC traditional approaches applied to each small object is wasting the bandwidth. An optimized solution consists in grouping the small objects with equal importance into a single file. This is the goal of object bundle protection. We proposed a novel method, called Generalized Object Encoding that can deal with both aspects [37], [38], [39]. In 2011, we analyzed our GOE approaches with average metrics such as average waiting time, average number of packets to be encoded. In 2012, we continued the analysis and considered memory requirements at the decoder [30].

6.4.3. Universal distributed coding

Participant: Aline Roumy.

In 2012, we started a new collaboration with Michel Kieffer and Elsa Dupraz (Supelec, L2S) on universal distributed source coding. Distributed source coding refers to the problem where several correlated sources need to be compressed without any cooperation at the encoders. Decoding is however performed jointly. This problem arises in sensor networks but also in video compression techniques, where the correlation between the successive frames is not directly used at the encoder as in [17], and are therefore seen as distributed. Traditional approaches (from an information theoretical but also practical point of view) assume that the correlation channel between the sources is perfectly known. Since this assumption is not satisfied in practice, a way to get around this is to use a feedback channel (from the decoder to the encoder), that can trigger the encoder. Instead, we consider universal distributed source coding, where the correlation channel is unknown and belongs to a class parametrized by some unknown parameter vector. In [23], we proposed four uncertainty models that depend on the partial knowledge we have on the correlation channel and derived the information theoretical bounds.

6.4.4. Super-resolution as a communication tool

Participants: Marco Bevilacqua, Christine Guillemot, Aline Roumy.

In 2012, we carried on the collaboration with Alcatel Lucent Bell Labs, represented by M-L. Alberi Morel, in the framework of a Joint Inria/Alcatel Lucent lab. In this work, we continued investigating super resolution (SR) as a potential tool to use in the context of video transmission. As SR refers to the task of producing a high-resolution (HR) image from one or several low-resolution (LR) input images, one can think of sending a LR video to adapt to the complexity constraint of the encoder and/or the bandwidth limitation of the network, and still being able to reconstruct a HR video at the encoder side, by applying a SR algorithm.

As a first step toward the more ambitious goal of compressing video through SR, we developed a novel method for single-image SR based on a neighbor embedding technique. In the neighbor embedding based SR procedure, the LR input image is first divided into small patches, namely sub-windows of image. Each input patch is approximated by a linear combination of its nearest neighbors (LR candidate patches) taken from a dictionary. Then, the corresponding HR output patch is created by combining similarly the corresponding HR candidates of the dictionary. The SR image is finally obtained by aggregating all the single HR patches reconstructed. A key point of this approach is represented by the above mentioned dictionary, which is a stored set of LR and HR patch correspondences extracted from training natural images.

The studies undertaken led us to have two publications in international conferences [20], [19]: ICASSP (International Conference on Acoustics, Speech, and Signal Processing) and BMVC (British Machine Vision Conference). In [20] we presented a neighbor embedding based SR method, by following the general scheme, but also introducing a new method to compute the weights of the linear combinations of patches. The weights of a certain input patch are computed as the result of a least squares problem with a nonnegative constraint. The so resulting nonnegative weights, that intuitevely represent a reasonable solution as they allow only additive combinations of patches, are shown to perform better than other weight computation methods described in the literature. The least squares problem is solved in a original fashion by means of SNMF, a tool for matrix factorization with one nonnegative factor. In [19] we refined the proposed algorithm, by focusing more on a low complexity target and by giving some theoretical insights about the choice of the nonnegative embedding. An analysis about the representation of the patches (either by the straight luminance values of its pixels or by some "features" conveniently computed) is also performed. The algorithm is shown to have better results, both in terms of quality performance and running time, than other similar SR algorithms that also adopt a one-pass procedure; and comparable visual results with respect to more sophisticated multi-pass algorithms, but still presenting a much reduced computational time. During the year, some other studies have been conducted, e.g. on the creation of the dictionary and on alternative ways to select the candidate patches from the dictionary. These extra studies, together with the already consolidated work of the published papers, represent the point of departure to the next step of designing a framework for video super resolution.

7. Bilateral 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 : Compression of satellite images.
- Partners : Astrium, Inria-Rennes.
- Funding : Astrium.
- Period : Oct.11-Sept.14.

This contract with Astrium addresses the problem of sparse representation and dictionary learning for efficient sparse coding of video signals captured from a geostationary satellite. The goal is to develop a compact spatio-temporal representation taking advantage of the high redundancy present in the video of very high resolution and characterized by low motion. The first year has been dedicated to the study of different methods for learning tree-structured dictionaries, which can be well-tailored to the characteristics of the signals to be processed at each iteration of the greedy matching pursuit algorithms, while allowing efficient encoding of the produced sparse vectors. The performance of the dictionaries both in terms of PSNR versus sparsity and versus actual bit rate has been assessed.

7.1.2. Collaboration with Alcatel on robust video compression

Participants: Marco Bevilacqua, Christine Guillemot, Ronan Le Boulch, Aline Roumy.

- Title: Self adaptive video codec
- Funding: Joint research laboratory between Inria and Alcatel
- Period: Oct. 2010 - Dec. 2013.

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). The objective is, jointly with the Alcatel Lucent team, to develop video representations and compression tools allowing smooth network adaptation on one hand and loss resilience on the other hand. In that context, the PhD thesis of M. Bevilacqua focuses on the development and study of image and video super-resolution as a tool for constructing scalable representations, hence enabling network adaptation of transmitted video streams. Despite the use of scalable representations and of error correcting codes, the reconstructed video after decoding may still suffer from residual losses. The effect of these residual losses can be mitigated using loss concealment techniques, as developed within the ANR-ARSSO project. However, loss concealment does not allow perfect recovery of the lost data. The reconstructed video after loss concealment is hence still corrupted by noise. Forward error correction using wyner-ziv encoded video is therefore also envisaged as an additional protection means, to cope with the residual noise after loss concealment.

7.1.3. Contract with EutelSat on video traffic analysis

Participants: Laurent Guillo, Aline Roumy.

- Title : Bit rate statistical analysis of HEVC encoded video in a broadcast transmission.
- Partners : EutelSat, Inria-Rennes.
- Funding : EutelSat.
- Period : Aug.12-Feb.13.

This contract with EutelSat (starting in August 2012) is a consulting contract and aims at analyzing the variation of the video traffic, when the video is encoded by HEVC. Indeed, the main characteristic of satellite broadcasting, as proposed by Eutelsat, is to provide a nearly constant video quality, which is obtained by variable video traffic (bit rate). Then, to address this variability issue, statistical multiplexing is used to share the resource among the users. However, statistical multiplexing needs a precise analysis of this variability. In this contract, we therefore analyze this variability, when the video is compressed with the upcoming video compression standard HEVC.

7.1.4. Contract with SHOM (Service Hydrographique et Océanographique de la Marine)

Participants: Alan Bourasseau, Olivier Le Meur.

- Title: Oceanograhic data compression
- Partners: SHOM, Alyotech, Univ. Rennes 1
- Funding: SHOM
- Period: 09/2012-02/2013.

The project consists in developing lossless and lossy compression algorithms for oceanographic data in partnership with ALYOTECH. The SIROCCO team contributes on the design and development of compression algorithms for this specific type of data, based on diffusion methods. The main constraint is the limited bandwidth used by the navy to transmit the data, i.e. an emitted message must be smaller than 4 kilo bytes. The obtained quality versus rate performances will be assessed against those given by state of the art solutions (HEVC-Intra and JPEG-2000).

7.2. Grants with Industry

7.2.1. CIFRE contract with Orange on Generalized lifting for video compression

Participants: Christine Guillemot, Bihong Huang.

- Title : Generalized lifting for video compression.
- Research axis : § 6.3.4.
- Partners : Orange Labs, Inria-Rennes.
- Funding : Orange Labs.
- Period : Apr.12-Mar.15.

This contract with Orange labs. (starting in April. 2012) concerns the PhD of Bihong Huang and aims at investigating various forms of generalized lifting as efficient lossless and lossy coding operators in a video encoder. The generalized lifting is an extension of the lifting scheme of classical wavelet transforms which permits the creation of nonlinear and signal probability density function (pdf) dependent and adaptive transforms. This study is also carried out in collaboration with UPC (Ph. Salembier) in Barcelona.

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 : § 6.1.1.
- Partners : Orange Labs, Inria-Rennes.
- Funding : Orange Labs.
- Period : Dec.11-Nov.14.

This contract with Orange labs. (starting in Dec. 2011) concerns the PhD of Darya Khaustova and 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 accuratly 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 Technicolor on sparse modelling of spatio-temporal scenes

Participants: Martin Alain, Safa Cherigui, Christine Guillemot.

- Title : Spatio-temporal analysis and characterization of video scenes
- Research axis : § 6.1.3.
- Partners : Technicolor, Irisa/Inria-Rennes.
- Funding : Technicolor, ANRT.
- Period : Ph.D Of S. Cherigui (Nov.09- Oct.12); Ph.D. of M. Alain (Oct.12-Sept.15).

A first CIFRE contract has concerned the Ph.D of Safa Cherigui from Nov.2009 to Oct.2012, in collaboration with Dominique Thoreau (Technicolor). The objective was to investigate texture and video scene characterization methods and models based on sparse and data dimensionality reduction techniques, as well as on concepts of epitomes. The objective was then to use these models and methods in different image processing problems focusing in particular on video compression. Among the main achievements, one can first cite novel Intra prediction techniques using data dimensionality reduction methods with correspondance map-aided K-NN search. A novel method has also been introduced for constructing epitome representations of an image, with rate-distortion optimization criteria. The epitome coupled with inpainting is used for image compression, showing significant performance gains with respect to H.264 Intra prediction modes. The above results have led to a best student paper award at IEEE-ICASSP 2012 [21], one nominated paper among the 11 finalists for a best student paper award at IEEE-ICIP 2012 [22], one journal publication in the IEEE Trans. on Image Processing [13] and to 6 patents.

The first PhD thesis has focused on spatial analysis, processing, and prediction of image texture. A second CIFRE contract (PhD thesis of Martin Alain) has hence just started to push further the study by addressing issues of spatio-temporal analysis and epitome construction, with applications to temporal prediction, as well as to other video processing problems such as video inpainting.

7.2.4. CIFRE contract with Thomson Video Networks (TVN) on Video analysis for HEVC based video coding

Participants: Nicolas Dhollande, Christine Guillemot, Olivier Le Meur.

- Title : Coding optimization of HEVC by using preanalysis approaches.
- Partners : Thomson Video Networks, Univ. Rennes 1.
- Funding : Thomson Video Networks (TVN).
- Period : Nov.12-Sept.15.

This contract with TVN (starting in Oct. 2012) concerns the PhD of Nicolas Dhollande and aims at performing a coding mode analysis and developing a pre-analysis software. HEVC standard is a new standard of compression including new tools such as advanced prediction modes. Compared to the previous standard H.264, HEVC's complexity is three to four times higher. The goal of this thesis is to to infer the best coding decisions (prediction modes...) in order to reduce the computational complexity of HEVC. A pre-analysis or a first coding pass. This latter would provide useful estimates such as rate distortion characteristics of video, the set of quantization parameters (often called just QPs). These estimated characteristics would help to select quantization parameters for the second pass in order to reduce the computational complexity by keeping a good quality.

8. Partnerships and Cooperations

8.1. National Initiatives

8.1.1. ANR-PERSEE

Participants: Josselin Gautier, Christine Guillemot, Laurent Guillo, Olivier Le Meur, Fabien Racapé.

- Title : Perceptual coding for 2D and 3D images.
- Research axis : § 6.2.2, 6.1.1.
- Partners : IRCCYN-Polytech Nantes, INSA-Rennes, Telecom Paris Tech.
- Funding : ANR.
- Period : 10/2009-08/2013

The objective of the project is to develop perceptually driven coding solutions for mono-view and multi-view video. The SIROCCO project-team contributes on different problems relevant for mono-view and multi-view video coding: visual attention modeling (see Section 6.1.1), 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 6.2.2. A computational model for 3D content has also been studied (see Section 6.1.1)

8.1.2. ANR-ARSSO

Participants: Mounira Ebdelli, Christine Guillemot, Ronan Le Boulch, Olivier Le Meur, 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. The SIROCCO project-team contributes on the development of loss concealment methods based on video inpainting. A first approach using examplar-based inpainting with neighbor embedding techniques has been developed. This method is currently being improved along three directions: 1/- he use of new distance metrics for finding the best matching patches; 2/- using a multi-resolution approach to both reduce the computational time and improve the robustness of the method; 3/- using mosaicking techniques for enhancing steps of stationary background and spatial inpainting. These solutions are studied in the context of a video compression and transmission chain using the emerging HEVC coding standard.

8.2. European Initiatives

8.2.1. FP7-PEOPLE-SHIVPRO

Participants: Olivier Le Meur, Zhi Liu.

- Title : Saliency-aware High-resolution Video Processing.
- Research axis : 6.1.1.
- Partners : Visting professor from Beijing University.
- Funding : EC-FP7 MC-IIF International Incoming Fellowships (IIF).
- Period : 08/2012-07/2014

The proposal SHIVPRO (Saliency-aware High-resolution Video Processing) submitted to the call FP7-PEOPLE-2011-IIF (funding scheme: MC-IIF International Incoming Fellowships (IIF)) has been accepted. Dr. Z. Liu, from Beijing University, has joined the team since August 2012 for two years. The objective of this project is to propose an efficient spatio-temporal saliency model to predict salient regions in High-Resolution (HR) videos, and fully exploit it to ease the design and improve the performance of HR video compression and retargeting applications. With the aim to overcome the drawbacks of existing saliency models, based on a multiscale region representation, the proposed model systematically realizes statistical model saliency measuring, intra-scale saliency modification, inter-scale saliency propagation and flexible incorporation of top-down information, to generate a novel saliency representation form with scalability, saliency tree, from which a multiscale saliency fusion scheme is used to derive high-quality saliency maps at various scales.

8.3. International Research Visitors

8.3.1. Visits of International Scientists

Mattei Mancas, researcher from the Univ. of Mons, Belgium has visited the team for two months (June-July 2012).
Dr. Zhi Liu, from Beijing University, is visiting the team since August 2012 for two years. His stay is funded by the FP7-PEOPLE-2011-IIF program. The funding scheme is the MC-IIF International Incoming Fellowships (IIF).

9. Dissemination

9.1. Scientific Animation

- 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 “Technical co-chair” of the Packet Video International Conference, Munich, May 2012.
- C. Guillemot is member as scientific expert of the CCRRDT (Regional Committee of Research and Technological Development) of the Brittany region.
- C. Guillemot is member of the committee in charge of the IEEE Brillouin-Glavieux award.
- C. Guillemot has been a member of the technical program committees of the international conferences EUSIPCO 2012, PV-2012 and PCS-212.
- C. Guillemot is member of the “bureau du Comité des Projets” as well as of the “commission personnels” in charge of the postdoc and delegation recrutments.
- C. Guillemot was a member of the selection committee for recruting a Professor at Univ. Paris-Sud 11, section 63 (2012).
- C. Guillemot gave a keynote talk on data dimensionality reduction methods and application to different image processing problems at the International Symposium on Image, Video and Communications (ISIVC), Valenciennes, July 2012.
- C. Labit is the Scientific Board chairman of Rennes 1 University (since June 1st, 2008).
- C. Labit is president of the Rennes-Atalante Science Park and of the start-up incubator Emergys (since April, 2007).
- C. Labit is member of the GRETSI association board.
- C. Labit is member of the coordination group of the Allistene alliance devoted to ICT.

- O. Le Meur co-organized a special session about visual attention at the conference SPIE Photonics Europe 2012 (presentations are available on http://www.irisa.fr/temics/staff/lemeur/visualAttention/specialsessions/).
- A. Roumy is a member of the GRETSI association board.
- A. Roumy has been a member of the technical program committee of the international conference EUSIPCO 2012.
- A. Roumy was nominated a titular member of the National Council of Universities (CNU section 61, 2012-2015).
- A. Roumy is a local liaison officer for Eurasip.
- A. Roumy co-organized a two-day workshop on Sparse Models and Machine Learning in Rennes (15-16 Oct. 2012) (sponsored by the Signal, Image & Robotics Department of IRISA. In partnership with EURASIP, Inria & ORANGE LABS).

9.2. Patents and Standardization

The team contributes to the HEVC and 3DV standardization initiatives of the ITU/MPEG joint collaborative team on video coding (JCT-VC) [32], [34], [33], [36], [35].
Four patents have been filed jointly by Technicolor and Inria in the area of 2D video coding. Two more patents have been filed by Inria on motion and disparity prediction tools for 3D video coding.

9.3. Teaching - Supervision - Juries

9.3.1. Teaching

Master: C. Guillemot, Image and video compression, 12 hours, M2 computer science, Univ. of Rennes 1, France.

Master: C. Guillemot, Image and video compression, 12 hours, M2 SISEA, Univ. of Rennes 1, France.

Master: L. Guillo, Multimedia Communication, 15 hours, M2 Network Engineering, Univ. of Rennes 1, France.

Master: O. Le Meur, Selective visual attention, 13 hours, M2, Univ. of Paris 8, France.

Master: O. Le Meur, Acquisition/Image Processing/Compression, 22 hours, M2 MITIC, Univ. of Rennes 1, France.

Master; A. Roumy, Acquisition/Image processing/Compression, 16 hours, M2 MITIC, Univ. of Rennes 1, France.

Master: A. Roumy, Magistère program, Information Theory, Computer science and telecommunications, 18 hours, Ecole Normale Supérieure de Cachan, Ker Lann campus, France.

Engineering degree: C. Guillemot, Video communication, 16 hours, Télécom Lille 1, Villeneuve-d'Ascq, France.

Engineering degree: L. Guillo, Video streaming, 3 hours, ESIR, University Rennes 1, France.

Engineering degree: C. Labit, Entrepreneurship and innovation, 3 hours, ESIR, Rennes, France.

Engineer degree: O. Le Meur, Image Processing, video analysis and compression, 54 hours, ESIR2, Univ. of Rennes 1, France.

Engineer degree: O. Le Meur, Visual communication, 65 hours, ESIR3, Univ. of Rennes 1, France.

Engineering degree: A. Roumy, Image processing and numerical analysis, 51 hours, ECAM Rennes, France.

9.3.2. Supervision

PhD & HdR

PhD : V. Jantet, Layered depth images for multi-view video coding, Univ. Rennes 1, 23rd Nov. 2012, C. Guillemot and L. Morin (INSA)

PhD : J. Gautier, A Dynamic Visual Attention Model for 2D and 3D conditions; Depth Coding and Inpainting-based Synthesis for Multiview Videos, Univ. Rennes 1, 5th Dec. 2012, O. Le Meur and C. Guillemot

PhD in progress : J. Aghaei Mazaheri, Sparse representations and dictionary learning for satellite image compression, Oct. 2011, C. Guillemot and C. Labit (contract with Astrium)

PhD in progress : M. Bevilacqua, Image and video super-resolution using neighbor embedding algorithms, Feb. 2011, A. Roumy and C. Guillemot (Contract with Alcatel/Lucent)

PhD in progress : S. Cherigui, Spatio-temporal scene modeling for video compression, Nov. 2009, C. Guillemot (Cifre contract with Technicolor)

PhD in progress : B. Huang, Video compression with generalized lifting, Apr. 2012, C. Guillemot (Cifre contract with Orange)

PhD in progress : D. Khaustova, Objective evaluation of 3D video quality, Dec. 2011, O. Le Meur (Cifre contract with Orange)

PhD in progress : M. Ebdelli, Video inpainting for editing and loss concealment, Dec. 2010, C. Guillemot and O. Le Meur

PhD in progress : M. Alain, Spatio-temporal linear embedding for epitome-based video compression, Oct. 2012, C. Guillemot (Cifre contract with Technicolor)

9.3.3. Juries

- C. Guillemot has been member (president) of the jury of the HDR committee of:
  - G. Coatrieux, Telecom Bretagne, 9/07/2012
  - E. Vincent, Univ. Rennes 1, 23/11/2012
- C. Guillemot has been member (president) of the jury of the PhD committee of:
  - C. Greco, Telecom ParisTech, 6/07/2012
- C. Labit has been member (president) of the jury of the HDR committee of:
  - Marie Babel, INSA Rennes, 29/06/2012
- C. Labit has been examiner and member of the jury of the PhD committee of:
  - Lu Zhang-Ge, Univ d'Angers, 28/11/2012

9.4. Popularization

- L. Guillo is advisor of secondary school pupils following the course "Informatique et Science du Numérique (ISN)" of the Lycée Jean Macé in Rennes, France.
- A. Roumy led a reading group on machine learning that was held weekly in the academic year 2011/12.
- Aline Roumy gave a lecture on video compression at the lycée Chateaubriand for highschool students (Terminale and classes préparatoires).
- Participation of Sirocco team members to the Rennes Science Festival (cultural dissemination of sciences).

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 Transactions on Image Processing", 2006, vol. 15, n^{o} 10, p. 2892-2903, http://hal.inria.fr/inria-00504227.

[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, n^{o} 11, p. 5133-5142.

[4] C. GUILLEMOT, A. ROUMY. *Chapter entitled "Towards constructive Slepian-Wolf coding schemes, in book "Distributed source coding"*, 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. JÉGOU, C. GUILLEMOT. *Robust multiplexed codes for compression of heterogeneous data*, in "IEEE Transactions on Information Theory", April 2005, vol. 51, n^{o} 4, p. 1393 - 1407, http://hal.inria.fr/inria-00604036.

[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^{o} 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^{o} 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^{o} 9, p. 1107–1119.

[10] A. ROUMY, S. GUEMGHAR, G. CAIRE, S. VERDU. *Design Methods for Irregular Repeat-Accumulate Codes*, in "IEEE Trans. on Information Theory", August 2004, vol. 50, n^{o} 8.

Publications of the year

Doctoral Dissertations and Habilitation Theses

[11] J. GAUTIER. *A dynamic atention model for 2D and 3D conditions; Depth coding and inpainting-based synthesis for multi-view videos*, University Rennes 1, Dec. 2012, http://tel.archives-ouvertes.fr/tel-00758112.

[12] V. JANTET. *Compression multi-vues par representation LDI (Layered Depth Images)*, Université Rennes 1, November 2012, http://tel.archives-ouvertes.fr/tel-00758301.

Articles in International Peer-Reviewed Journals

[13] S. CHERIGUI, C. GUILLEMOT, D. THOREAU, P. GUILLOTEL, P. PEREZ. *Correspondence map-aided neighbor embedding for image intra prediction*, in "IEEE Transaction on Image Processing", November 2012, http://hal.inria.fr/hal-00755762.

[14] J. GAUTIER, O. LE MEUR. *A time-dependent saliency model combining center and depth biases for 2D and 3D viewing conditions*, in "Cognitive Computation", Jun 2012, vol. 4, n^{o} 2, p. 141-156, http://hal.inria.fr/hal-00757536.

[15] O. LE MEUR, T. BACCINO. *Methods for comparing scanpaths and saliency maps: strengths and weaknesses*, in "Behavior Research Methods", Jul 2012, p. 1-16, http://hal.inria.fr/hal-00757615.

[16] J. TAQUET, C. LABIT. *Hierarchical Oriented Predictions for Resolution Scalable Lossless and Near-Lossless Compression of CT and MRI Biomedical Images*, in "IEEE Trans. on Image Processing", May 2012, vol. 21, n^o 5, p. 2641-2652, http://hal.inria.fr/hal-00755741.

[17] V. TOTO-ZARASOA, A. ROUMY, C. GUILLEMOT. *Source modeling for Distributed Video Coding*, in "IEEE Trans. on Circuits and Systems for Video Technology", Feb. 2012, vol. 22, n^o 2, http://hal.inria.fr/inria-00632708/en/.

[18] M. TURKAN, C. GUILLEMOT. *Image prediction based on neighbor embedding methods*, in "IEEE Trans. On Image Processing", April 2012, vol. 4, n^o 21, http://hal.archives-ouvertes.fr/hal-00747007.

International Conferences with Proceedings

[19] M. BEVILACQUA, M. BEVILACQUA, A. ROUMY, C. GUILLEMOT, M.-L. ALBERI MOREL. *Low-Complexity Single-Image Super-Resolution based on Nonnegative Neighbor Embedding*, in "BMVC (British Machine Vision Conference)", 2012, http://hal.inria.fr/hal-00747054.

[20] M. BEVILACQUA, A. ROUMY, C. GUILLEMOT, M.-L. ALBERI MOREL. *Neighbor embedding based single-image super-resolution using Semi-Nonnegative Matrix Factorization*, in "ICASSP (IEEE International Conference on Acoustics, Speech, and Signal Processing)", 2012, http://hal.inria.fr/hal-00747042.

[21] S. CHERIGUI, C. GUILLEMOT, D. THOREAU, P. GUILLOTEL, P. PEREZ. *Hybrid Template and Block Matching for Image Intra Prediction*, in "IEEE International Conference on Acoustics and Signal Processing (IEEE-ICASSP)", March 2012, http://hal.archives-ouvertes.fr/hal-00695277.

[22] S. CHERIGUI, C. GUILLEMOT, D. THOREAU, P. GUILLOTEL, P. PEREZ. *Map-Aided Locally Linear Embedding Methods for Image Prediction*, in "ICASSP'12 - International Conference on Acoustics, Speech, and Signal Processing, 2012", March 2012, http://hal.inria.fr/hal-00749966.

[23] E. DUPRAZ, A. ROUMY, M. KIEFFER. *Source Coding with Side Information at the Decoder: Models with Uncertainty, Performance Bounds, and Practical Coding Schemes*, in "International Symposium on Information Theory and its Applications 2012", October 2012, http://hal-supelec.archives-ouvertes.fr/hal-00727780.

[24] M. EBDELLI, C. GUILLEMOT, O. LE MEUR. *Examplar-Based Video Inpainting With Motion-Compensated Neighbor Embedding*, in "IEEE International Conference on Image Processing", Oct. 2012, http://hal.inria.fr/hal-00747029.

[25] M. EBDELLI, O. LE MEUR, C. GUILLEMOT. *Loss Concealment Based on Video Inpainting For Robust Video Communication*, in "20th European Signal Processing Conference", August 2012, http://hal.inria.fr/hal-00747034.

[26] J. GAUTIER, O. LE MEUR, C. GUILLEMOT. *Efficient Depth Map Compression based on Lossless Edge Coding and Diffusion*, in "Picture Coding Symposium (PCS)", May 2012, http://hal.inria.fr/hal-00757469.

[27] C. GUILLEMOT, M. TURKAN. *Neighbor Embedding with Non-negative Matrix Factorization for image prediction*, in "IEEE International Conference on Acoustics and Signal Processing (IEEE-ICASSP)", March 2012, http://hal.inria.fr/hal-00755720.

[28] O. LE MEUR, C. GUILLEMOT. *Super-resolution-based inpainting*, in "ECCV", October 2012, http://hal.inria.fr/hal-00742827.

[29] R. MARTINEZ NORIEGA, A. ROUMY, G. BLANCHARD. *Exemplar-based image inpainting: fast priority and coherent Nearest Neighbor search*, in "IEEE Workshop on Machine Learning for Signal Processing (MLSP)", September 2012, http://hal.inria.fr/hal-00752506.

[30] A. ROUMY, V. ROCA, B. SAYADI. *Memory Consumption Analysis for the GOE and PET Unequal Erasure Protection Schemes*, in "IEEE International Conference on Communications", February 2012, http://hal.inria.fr/hal-00668826.

[31] M. TURKAN, C. GUILLEMOT. *Locally linear embedding based texture synthesis for image prediction and error concealment*, in "IEEE International Conference on on Image Processing (IEEE-ICIP)", Oct. 2012, http://hal.archives-ouvertes.fr/hal-00747011.

Research Reports

[32] T. GUIONNET, L. GUILLO, C. GUILLEMOT. *CE5.h related: Information on the impact of the merge candidate list pruning process on side views encoding*, Inria - IRISA, July 2012, http://hal.inria.fr/hal-00755731.

[33] T. GUIONNET, L. GUILLO, C. GUILLEMOT. *CE5.h related: Merge candidate list extension for disparity compensated prediction*, Inria - IRISA, July 2012, http://hal.inria.fr/hal-00755740.

[34] T. GUIONNET, L. GUILLO, C. GUILLEMOT. *CE5.h related: Reducing the coding cost of merge index by dynamic merge candidate list re-ordering*, Inria - IRISA, July 2012, http://hal.inria.fr/hal-00755737.

[35] T. GUIONNET, L. GUILLO, C. GUILLEMOT. *3D-CE5.h: Merge candidate list for disparity compensated prediction*, Inria - IRISA, October 2012, http://hal.inria.fr/hal-00755747.

[36] T. GUIONNET, L. GUILLO, C. GUILLEMOT. *3D-CE5.h: Reducing the coding cost of merge index by dynamic merge index re-allocation*, Inria - IRISA, October 2012, http://hal.inria.fr/hal-00755743.

[37] 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)*, Inria - IRISA - Alcatel Lucent Bell Labs, July 2012, http://tools.ietf.org/html/draft-roca-rmt-goe-fec-02.

[38] V. ROCA, A. ROUMY, B. SAYADI. *The Generalized Object Encoding (GOE) LDPC-Staircase FEC Scheme*, Inria - IRISA - Alcatel Lucent Bell Labs, July 2012, http://tools.ietf.org/html/draft-roca-rmt-goe-ldpc-01.

[39] V. ROCA, A. ROUMY, B. SAYADI. *The Need for Extended Forward Erasure Correction (FEC) Schemes: Problem Position*, Inria - IRISA - Alcatel Lucent Bell Labs, July 2012, http://tools.ietf.org/html/draft-roca-rmt-extended-fec-problem-00.