IN PARTNERSHIP WITH:
CNRS

Université Rennes 1

Activity Report 2015

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 interpretation

Table of contents

1. Members **1**
2. Overall Objectives **2**
2.1. Introduction 2
2.2. Analysis and modeling for compact representation 2
2.3. Rendering, inpainting and super-resolution 2
2.4. Representation and compression of visual data 3
2.5. Distributed processing and robust communication 3
3. Research Program **3**
3.1. Introduction 3
3.2. Parameter estimation and inference 4
3.3. Data Dimensionality Reduction 4
3.4. Perceptual Modelling 5
3.5. Coding theory 5
4. Application Domains **6**
4.1. Introduction 6
4.2. Compression with advanced functionalities 6
4.3. Networked visual applications 7
4.4. Medical Imaging (CT, MRI, Virtual Microscopy) 7
4.5. Editing and post-production 7
5. Highlights of the Year **8**
6. New Software and Platforms **8**
6.1. Fixation Analysis 8
6.2. Salient object extraction 8
6.3. Saccadic model 8
6.4. Hierarchical super-resolution based inpainting 8
6.5. Video Inpainting for Loss Concealment 9
6.6. Video Inpainting for Editing 9
7. New Results **9**
7.1. Analysis and modeling for compact representation and navigation 9
7.1.1. Visual attention 9
7.1.2. Epitome-based video representation 11
7.1.3. Graph-based multi-view video representation 11
7.2. Rendering, inpainting and super-resolution 12
7.2.1. Color and light transfer 12
7.2.2. Image guided inpainting 13
7.2.3. Clustering on manifolds for image restoration 13
7.3. Representation and compression of large volumes of visual data 14
7.3.1. Manifold learning and low dimensional embedding for classification 14
7.3.2. Adaptive clustering with Kohonen self-organizing maps for second-order prediction 14
7.3.3. Rate-distortion optimized tone curves for HDR video compression 15
7.3.4. Local Inverse Tone Curve Learning for HDR Image Scalable Compression 15
7.3.5. HEVC-based UHD video coding optimization 15
7.4. Distributed processing and robust communication 16
7.4.1. Information theoretical bounds of Free-viewpoint TV 16
7.4.2. Compressed Sensing : a probabilistic analysis of the orthogonal matching pursuit algorithm 16
8. Bilateral Contracts and Grants with Industry **17**
8.1.1. CIFRE contract with Orange on Generalized lifting for video compression 17
8.1.2. CIFRE contract with Technicolor on High Dynamic Range (HDR) video compression 17

8.1.3. CIFRE contract with Technicolor on sparse modelling of spatio-temporal scenes 17
8.1.4. CIFRE contract with Thomson Video Networks (TVN) on Video analysis for HEVC based video coding 18
8.1.5. CIFRE contract with Envivio on LDR compatible HDR video coding 18
8.1.6. CIFRE contract with Technicolor on light fields editing 18
8.1.7. CIFRE contract with Technicolor on cloud-based video compression 19
9. Partnerships and Cooperations **19**
9.1. Regional Initiatives 19
9.2. International Initiatives 19
9.3. International Research Visitors 19
10. Dissemination **19**
10.1. Promoting Scientific Activities 19
10.1.1. Scientific events organisation 19
10.1.2. Scientific events selection 20
10.1.3. Journal 20
10.1.4. Invited talks 20
10.1.5. Leadership within the scientific community 20
10.1.6. Scientific expertise 20
10.1.7. Research administration 21
10.2. Teaching - Supervision - Juries 21
10.2.1. Teaching 21
10.2.2. Supervision (PhDs defended during the year) 21
10.2.3. Juries 21
11. Bibliography **22**

Project-Team SIROCCO

Creation of the Project-Team: 2012 January 01

Keywords:

Computer Science and Digital Science:

5. - Interaction, multimedia and robotics
5.3. - Image processing and analysis
5.4. - Computer vision
5.9. - Signal processing

Other Research Topics and Application Domains:

6. - IT and telecom

1. Members

Research Scientists

Christine Guillemot [Team leader, Inria, Senior Researcher, HdR]
Claude Labit [Inria, Senior Researcher, 20%, HdR]
Thomas Maugey [Inria, Researcher]
Aline Roumy [Inria, Researcher]

Faculty Member

Olivier Le Meur [Univ. Rennes I, Associate Professor, HdR]

PhD Students

Martin Alain [Technicolor, until Oct 2015, granted by CIFRE]
Jean Begaint [Technicolor, from Nov 2015, granted by CIFRE]
Nicolas Dhollande [Thomson Video Networks, until Nov. 2015, granted by CIFRE]
Thierry Dumas [Inria, from Oct 2015, granted by DGA]
Julio Cesar Ferreira [co-tutelle Univ. Uberlandia, Brazil, until Dec. 2015]
David Gommelet [Envivio, granted by CIFRE]
Matthieu Hog [Technicolor, from Sep 2015, granted by CIFRE]
Bihong Huang [Orange Labs, until Apr 2015, granted by CIFRE]
Mikael Le Pendu [Technicolor, granted by CIFRE]
Mira Rizkallah [Univ. Rennes 1, MESR grant, from Dec 2015]
Hristina Hristova [Univ. Rennes 1, MESR grant, also with FRVsense]

Post-Doctoral Fellows

Pierre Buyssens [Univ. Rennes I, from Oct 2015]
Xin Su [Inria, from May 2015]
Elif Vural [Inria, until May 2015]

Visiting Scientist

Reuben Farrugia [On leave from Univ. of Malta, from Sep 2015]

Administrative Assistant

Huguette Bechu [Inria]

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.

[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.

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.

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 (e.g. Packet Error Rate (PER)), the objective is then to adapt the data coding, protection and transmission scheme. However, the reliability of the estimated PER impacts the performance of FEC schemes. We investigate the problem of constructing codes which 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 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.

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.

3. Research Program

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 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 model 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 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 PCA and ICA, the non-negativity constraint makes the representation purely 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 unitary 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] D.D. Lee and H.S. Seung, “Algorithms for non-negative matrix factorization”, Nature 401, 6755, (Oct. 1999), pp. 788-791.

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 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 play an important role in our perception (see the example concerning the assumption that light comes from above-left). Our perception results from the combination of prior beliefs with data we gather from the environment. A Bayesian framework is an elegant solution to model these interactions [6]. We define a vector $\overrightarrow{v}_l$ of local measurements (contrast of color, orientation, etc.) and vector $\overrightarrow{v}_c$ of global and contextual features (global features, prior locations, type of the scene, etc.). The salient locations S for a spatial position $\overrightarrow{x}$ are then given by:

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

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

3.5. Coding theory

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.

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]

[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.

[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.

(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 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. This is particularly true for mobile applications, as the need for wireless transmission capacity will significantly increase during the years to come. Hence, efficient compression tools are required to satisfy the trend towards mobile access to larger image resolutions and higher quality. A new impulse to research in video compression is also brought by the emergence of new formats beyond High Definition TV (HDTV) towards high dynamic range (higher bit depth, extended colorimetric space), super-resolution, formats for immersive displays allowing panoramic viewing and 3DTV.

Different video data formats and technologies are envisaged for interactive and immersive 3D video applications using omni-directional videos, stereoscopic or multi-view videos. The "omni-directional video" set-up refers to 360-degree view from one single viewpoint or spherical video. Stereoscopic video is composed of two-view videos, the right and left images of the scene which, when combined, can recreate the depth aspect of the scene. A 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 and its robustness to transmission impairments are thus important features for seamless adaptation to varying network conditions and to terminal capabilities.

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. Nevertheless, some discussions are on-going to use near-lossless coding of regions-of-interest (ROI) in medical images. The detection and segmentation of region-of interest (ROIs) can be guided by a precise knowledge of the medical imaging modalities and by the diagnosis and expertise of practitioners. It seems also to be 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. The increased number of 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. Highlights of the Year

5.1. Highlights of the Year

- C. Guillemot has received a Google faculty research award
- T. Maugey has received an AIS grant ("Aide à installation scientitifique") from the region of Brittany.
- The papers [31], [28] have been recognized as “Top 10%” at the IEEE international conference ICIP 2015.

6. New Software and Platforms

6.1. Fixation Analysis

FUNCTIONAL DESCRIPTION

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 a 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.

- Participants: Olivier Le Meur and Thierry Baccino
- Contact: Olivier Le Meur

6.2. Salient object extraction

FUNCTIONAL DESCRIPTION

This software detects salient object in an input picture in an automatic manner. The detection is based on super-pixel segmentation and contrast of histogram. This software is dedicated to people working in image processing and post production.

- Participants: Zhi Liu and Olivier Le Meur
- Contact: Olivier Le Meur

6.3. Saccadic model

The software called Scanpath Prediction aims at predicting the visual scanpath of an observer. The visual scanpath is a set of fixation points. The computational model is based on bottom-up saliency maps, viewing tendencies (that have been learned from eye tracking datasets) and inhibition-of-return. This study is based on the following paper [20]. This software is dedicated to people working in computer science, computer vision and cognitive science. This software is being registered at the APP (Agence de Protection des Programmes).

- Participants: Olivier Le Meur
- Contact: Olivier Le Meur

6.4. Hierarchical super-resolution based inpainting

From an input binary mask and a source picture, the software performs an examplar-based inpainting. The method is based on the combination of multiple inpainting applied on a low resolution of the input picture. Once the combination has been done, a single-image super-resolution method is applied to recover the details and the high frequency in the inpainted areas. The developments have been pursued in 2014, in particular by introducing a Poisson blending step in order to improve the visual quality of the inpainted video. This software is dedicated to people working in image processing and post production. This software is being registered at the APP (Agence de Protection des Programmes).

- Participants: Olivier Le Meur
- Contact: Olivier Le Meur

6.5. Video Inpainting for Loss Concealment

KEYWORDS: Video Inpainting - Motion informations - Loss concealment - BMFI (Bilinear Motion Field Interpolation)
FUNCTIONAL DESCRIPTION

From an input binary mask and a source video, the software performs an examplar-based inpainting. The motion information of the impaired areas is first recovered with a Bilinear Motion Field Interpolation (BMFI). The texture information is then recovered using a spatio-temporal examplar-based inpainting algorithm. The method to recover the texture proceeds in two steps: it first inpaints a low resolution version using an examplar-based method. Details of the inpainted corrupted areas of the input video are then retrieved using a nearest neighbor field (NNF) based super-resolution technique. A NNF is computed between an interpolated version of the concealed LR video and the known part of the received video at native resolution. In the same vein as in single-image super-resolution, the NNF is used to recover the high frequencies of the inpainted areas of the video.

- Participants: Ronan Le Boulch
- Contact: Olivier Le Meur

6.6. Video Inpainting for Editing

KEYWORDS: Video Inpainting - Editing
FUNCTIONAL DESCRIPTION

This software performs video inpainting for both static or free-moving camera videos. The method can be used for object removal, error concealment, and background reconstruction applications. To inpaint a frame, the method starts by aligning all the frames of a group of pictures (GOP). This is achieved by a region-based homography computation method which allows us to strengthen the spatial consistency of aligned frames. Then, from the stack of aligned frames, an energy function based on both spatial and temporal coherency terms is globally minimized. This energy function is efficient enough to provide high quality results even when the number of pictures in the GoP is rather small, e.g. 20 neighboring frames. This reduces the algorithm complexity and makes the approach well suited for near real-time video editing applications as well as for loss concealment applications.

- Participants: Mounira Ebdelli
- Contact: Olivier Le Meur

7. New Results

7.1. 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

7.1.1. Visual attention

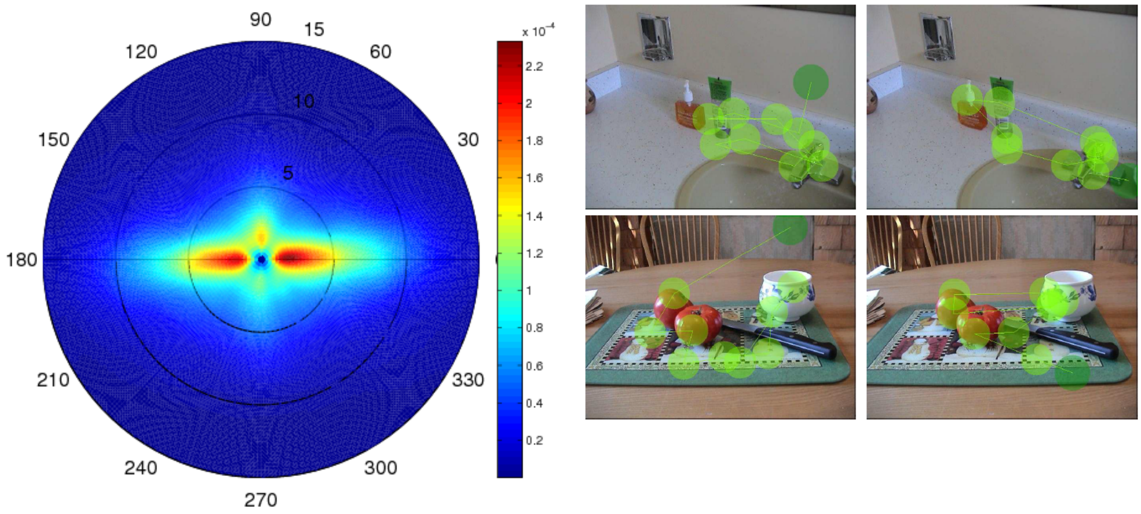
Participants: Pierre Buyssens, Olivier Le Meur.

Visual attention is the mechanism allowing to focus our visual processing resources on behaviorally relevant visual information. Two kinds of visual attention exist: one involves eye movements (overt orienting) whereas the other occurs without eye movements (covert orienting). Our research activities deals with the understanding and modeling of overt attention as well as saliency-based image editing. These research activities are described in the following sections.

Saccadic model: Most of the computation models of visual attention output a 2D static saliency map. This single topographic saliency map which encodes the ability of an area to attract our gaze is commonly computed from a set of bottom-up visual features. Although the saliency map representation is a convenient way to indicate where we look within a scene, these models do not completely account for the complexities of our visual system. One obvious limitation concerns the fact that these models do not make any assumption about eye movements and viewing biases. For instance, they implicitly make the hypothesis that eyes are equally likely to move in any direction.

There is evidence for the existence of systematic viewing tendencies. Such biases could be combined with computational models of visual attention in order to better predict where we look. Such a model, predicting the visual scanpath of observer, is termed as saccadic model. We recently propose a saccadic model ([20]) that combines bottom-up saliency maps, viewing tendencies and short-term memory. The viewing tendencies are related to the fact that most saccades are small (less than 3 degrees of visual angle) and oriented in the horizontal direction. Figure 1 (a) illustrates the joint probability distribution of saccade amplitudes and orientations. Examples of predicted scanpaths are shown in Figure 1 (b). We demonstrated that the proposed model outperforms the best state-of-the-art saliency models.

In the future, the goal is to go further by considering that the joint distribution of saccade amplitudes and orientations is spatially variant and depends on the scene category.



(a) (b)

Figure 1. (a) Joint probability distribution of saccade amplitudes and orientations shown on a polar plot. Radial position indicates saccadic amplitudes expressed in degree of visual angle. (b) Predicted scanpaths composed of ten fixations represented by green circles. The dark green circle corresponds to the first fixation which is randomly chosen.

Perceptual-based image editing: Since the beginning of October, we have started new studies related to perceptual-based image editing. The goal is to combine the modelling of visual attention with image/video editing methods. More specifically it aims at altering images/video sequences in order to attract viewers attention over specific areas of the visual scene. We intend to design new computational editing methods for emphasizing and optimizing the importance of pre-defined areas of the input image/video sequence. There exist very few studies in the literature dealing with this problem. Current methods simply alter the content by using blurring operation or by recoloring the image locally so that the focus of attention falls within the pre-defined areas of interest. One avenue for improving current methods is to minimize a distance computed between a user's defined visual scanpath and predicted visual scanpath. The content would be edited (i.e. recoloring, region rescaling, local contrast/resolution adjustment, removing disturbing object, etc) in an iterative manner in order to move the focus of attention towards the regions selected by the user.

7.1.2. Epitome-based video representation

Participants: Martin Alain, Christine Guillemot.

In 2014, we have developed fast methods for constructing epitomes from images. An epitome is a factorized texture representation of the input image, and its construction exploits self-similarities within the image. Known construction methods are memory and time consuming. The proposed methods, using dedicated list construction on one hand and clustering techniques on the other hand, aim at reducing the complexity of the search for self-similarities.

In 2015, we have developed methods for quantization noise removal (after decoding) exploiting the epitome representations together with local learning of either LLE (locally linear embedding) weights, which has proved to be a powerful tool for prediction [14], or using linear mapping functions between original and noisy patches. Compared to classical denoising methods which, most of the time, assume additive white Gaussian noise, the quantization turns out to be correlated to the signal which makes the problem more difficult. The methods have been experimented both in the contexts of single layer encoding and scalable encoding. The same methodology has been applied to super-resolution learning this time mapping functions between the low resolution and high resolution spaces in which lie the patches of the epitome [32].

7.1.3. Graph-based multi-view video representation

Participants: Christine Guillemot, Thomas Maugey, Mira Rizkallah, Xin Su.

One of the main open questions in multiview data processing is the design of representation methods for multiview data, where the challenge is to describe the scene content in a compact form that is robust to lossy data compression. Many approaches have been studied in the literature, such as the multiview and multiview plus depth formats, point clouds or mesh-based techniques. All these representations contain two types of data: i) the color or luminance information, which is classically described by 2D images; ii) the geometry information that describes the scene 3D characteristics, represented by 3D coordinates, depth maps or disparity vectors. Effective representation, coding and processing of multiview data partly rely on a proper representation of the geometry information. The multiview plus depth (MVD) format has become very popular in recent years for 3D data representation. However, this format induces very large volumes of data, hence the need for efficient compression schemes. On the other hand, lossy compression of depth information in general leads to annoying rendering artefacts especially along the contours of objects in the scene.

Instead of lossy compression of depth maps, we consider the lossless transmission of a geometry representation that captures only the information needed for the required view reconstructions. Our goal is to transmit "just enough" geometry information for accurate representation of a given set of views, and hence better control the effect of geometry lossy compression.

More particularly, in [23], we proposed a new Graph-Based Representation (GBR) for geometry information, where the geometry of the scene is represented as connections between corresponding pixels in different views. In this representation, two connected pixels are neighboring points in the 3D scene. The graph connections are derived from dense disparity maps and provide just enough geometry information to predict pixels in all the views that have to be synthesized.

GBR drastically simplifies the geometry information to the bare minimum required for view prediction. This "task-aware" geometry simplification allows us to control the view prediction accuracy before coding compared to baseline depth compression methods (Fig. 2). This work has first been carried out for multi-view configurations, in which cameras are parallel. We are currently investigating the extension of this promising GBR to complex camera transitions. An algorithm has already been implemented for two views and is being extended for multiple views. The next steps will be to develop color coding tools adapted to these graph structures.

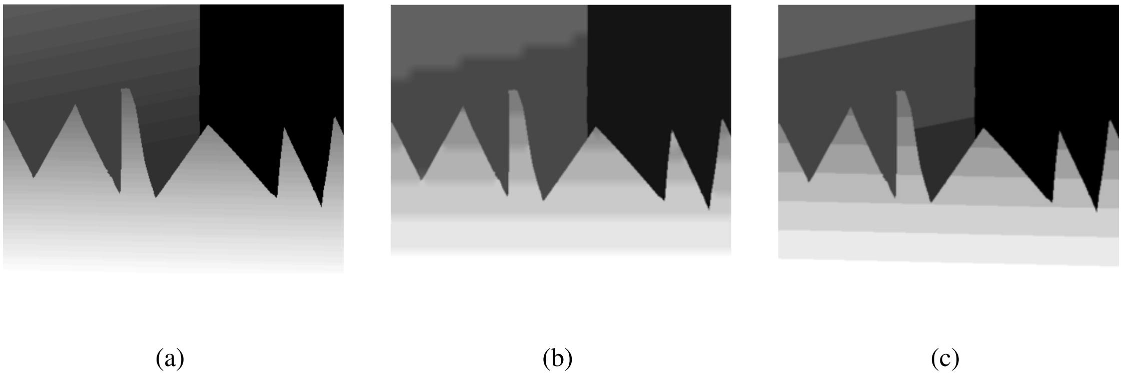


Figure 2. (a) original depth map, (b) depth map compressed with edge-adaptive method at 10kb with compression artifacts (c) depth image retrieved from the graph of our proposed GBR transmitted at 10kb keeping the original scene structure.

7.2. Rendering, inpainting and super-resolution

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

7.2.1. Color and light transfer

Participants: Hristina Hristova, Olivier Le Meur.

Color transfer aims at modifying the look of an original image considering the illumination and the color palette of a reference image. It can be employed for image and video enhancement by simulating the appearance of a given image or a video sequence. It can also be applied to hallucinations of particular parts of the day. Current state-of-the-art methods focus mainly on the global transfer of the light and color distributions. Unfortunately, the use of a global distribution is questionable since the light and color of image can significantly vary within the same scene. In [27], we proposed a new method to deal with the limitations of existing methods. The proposed method aims at performing the partitions of the input and reference images into Gaussian distributed clusters by considering the main style of input and reference images. From this clustering, several novel policies are defined for mapping the clusters of the input and reference images. To complete the style transfer, for each pair of corresponding clusters, we apply a parametric color transfer method (i.e. Monge-Kantorovitch transformation) and a local chromatic adaptation transform. Results, subjective user evaluation as well as objective evaluation show that the proposed method obtains visually pleasing and artifact-free images, respecting the reference style. Some results are illustrated in Figure 3.

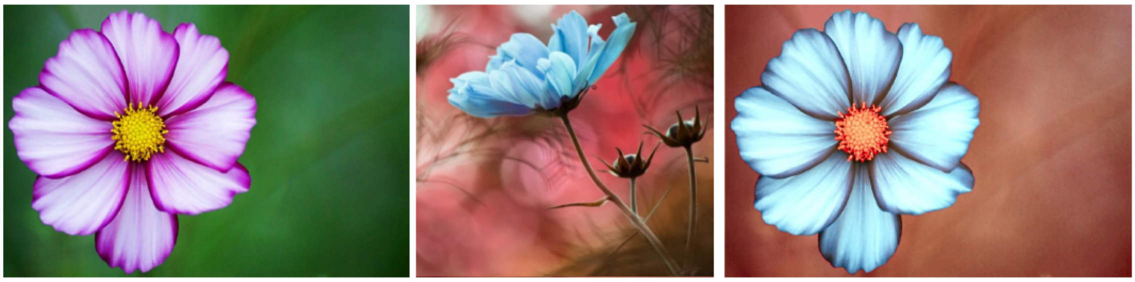


Figure 3. From left to right: input image, reference image and the result of the proposed method.

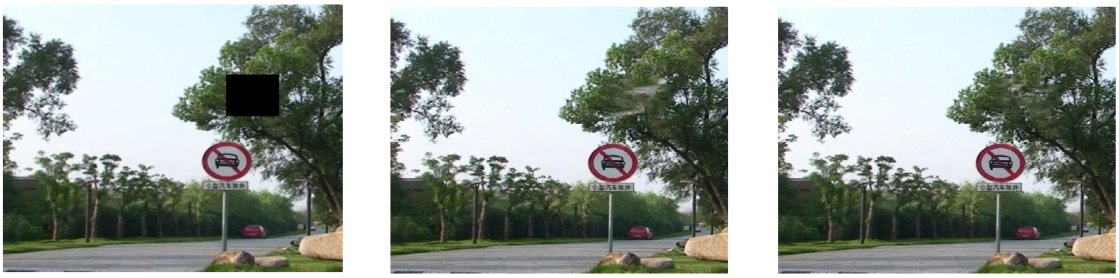
In [34], we extended the method presented in [27] to deal with a color transfer between two HDR images. One limitation of the two proposed methods is that we are still considering that the distributions of color and light follow a Gaussian law. We are currently investigating a more general approach by considering multivariate generalized Gaussian distribution.

7.2.2. Image guided inpainting

Participants: Christine Guillemot, Thomas Maugey.

Inpainting of images has been intensively studied in the past few years, especially for applications such as image restoration and editing [16]. Another application where inpainting techniques are useful is view synthesis, where holes are to be filled corresponding to areas that are no longer occluded. In the particular cases where one has access to ground truth images (like for example in multiview video coding where view synthesis is used for predicting the captured views from a reference one), auxiliary information can be generated to help inpainting, which leads to the concept of *guided inpainting*.

In [29], we have proposed a new auxiliary information that is used to refine the set of candidate patches for the hole filling step of the inpainting. Assuming that the patches of an image lie in a union of subspaces, *i.e.*, the images have different regions with different color textures, these patches are first clustered using a new recursive spectral clustering algorithm that extends existing sparse subspace clustering and replaces the sparse approximation by locally linear embedding, better suited for the inpainting context. Dictionaries are then built from these clusters and used for the hole filling process. However, the inpainting is not always able to "guess" in which cluster the patches of the hole belong to (especially around discontinuities). The auxiliary information that is built from the ground truth image may help to find the right cluster. We thus propose a new guided inpainting algorithm that forces the patch reconstruction to be done in one cluster only, if no auxiliary information is available, or in the cluster pointed by the auxiliary information, if it is available. Experiments (Fig. 4) show that auxiliary information helps to significantly improve the inpainting quality for a reasonable coding cost.



(a) (b) (c)

Figure 4. (a) input image to inpaint, (b) filled image using baseline not guided inpainting (c) filled image using proposed guided inpainting with an auxiliary information cost of 0.018 *bpp bitrate.*

We are currently working on the extension of this technique in order to place the guided inpainting problem in an information theoretic framework, and better answer the following questions: when additional information is actually needed? What type of auxiliary information is needed? how to optimize in a rate-distortion sense the guided inpainting problem?.

7.2.3. Clustering on manifolds for image restoration

Participants: Julio Cesar Ferreira, Christine Guillemot, Elif Vural.

Local learning of sparse image models has proven to be very effective to solve a variety of inverse problems in many computer vision applications. To learn such models, the data samples are often clustered using the K-means algorithm with the Euclidean distance as a dissimilarity metric. However, the Euclidean distance may not always be a good dissimilarity measure for comparing data samples lying on a manifold. We have developed two algorithms for determining a local subset of training samples from which a good local model can be computed for reconstructing a given input test sample, where we take into account the underlying geometry of the data. The first algorithm, called Adaptive Geometry-driven Nearest Neighbor search (AGNN), is an adaptive scheme which can be seen as an out-of-sample extension of the replicator graph clustering method for local model learning. The second method, called Geometry-driven Overlapping Clusters (GOC), is a less complex nonadaptive alternative for training subset selection. The AGNN and GOC methods have been evaluated in image super-resolution, deblurring and denoising applications and shown to outperform spectral clustering, soft clustering, and geodesic distance based subset selection in most settings.

7.3. Representation and compression of large volumes of visual data

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

7.3.1. Manifold learning and low dimensional embedding for classification

Participants: Christine Guillemot, Elif Vural.

Typical supervised classifiers such as SVM are designed for generic data types and do not make any particular assumption about the geometric structure of data, while data samples have an intrinsically low-dimensional structure in many data analysis applications. Recently, many supervised manifold learning methods have been proposed in order to take the low-dimensional structure of data into account when learning a classifier. Unlike unsupervised manifold learning methods which only take the geometric structure of data samples into account when learning a low-dimensional representation, supervised manifold learning methods learn an embedding that not only preserves the manifold structure in each class, but also enhances the separation between different classes.

An important factor that influences the performance of classification is the separability of different classes in the computed embedding. We have done a theoretical analysis of separability of data representations given by supervised manifold learning. In particular, we have focused on the nonlinear supervised extensions of the Laplacian eigenmaps algorithm and have examined the linear separation between different classes in the learned embedding. We have shown that, if the graph is such that the inter-group graph weights are sufficiently small, the learned embedding becomes linearly separable at a dimension that is proportional to the number of groups. These theoretical findings have been confirmed by experimentation on synthetic data sets and image data.

We have then considered the problem of out-of-sample generalizations for manifold learning. Most manifold learning methods compute an embedding in a pointwise manner, i.e., data coordinates in the learned domain are computed only for the initially available training data. The generalization of the embedding to novel data samples is an important problem, especially in classification problems. Previous works for out-of-sample generalizations have been designed for unsupervised methods. We have studied this problem for the particular application of data classification and proposed an algorithm to compute a continuous function from the original data space to the low-dimensional space of embedding. In particular, we have constructed an interpolation function in the form of a radial basis function that maps input points as close as possible to their projections onto the manifolds of their own class. Experimental results have shown that the proposed method gives promising results in the classification of low-dimensional image data such as face images.

7.3.2. Adaptive clustering with Kohonen self-organizing maps for second-order prediction

Participants: Christine Guillemot, Bihong Huang.

The High Efficiency Video Coding standard (HEVC) supports a total of 35 intra prediction modes which aim at reducing spatial redundancy by exploiting pixel correlation within a local neighborhood. However the correlation remains in the residual signals of intra prediction, leading to some high energy prediction residuals. In 2014, we have studied several methods to exploit remaining correlation in residual domain after intra prediction. These methods are based on vector quantization with codebooks learned and dedicated to the different prediction modes in order to model the directional characteristics of the residual signals. The best matching code vector is found in a rate-distortion optimization sense. Finally, the index of the best matching code vector is sent to the decoder and the vector quantization error, the difference between the intra residual vector and the best matching code vector, is processed by the conventional operations of transform, scalar quantization and entropy coding.

In a first approach called MDVQ (Mode Dependent Vector Quantization), the codebooks were learned using the k-means algorithm [26]. More recently, we have developed a variant of the approach, called AMDVQ (Adaptive MDVQ) by adding a codebook update step based on Kohonen Self-Organized Maps which aims at capturing the variations of the residual signal statistical charateristics. The Kohonen algorithm uses previously reconstructed residual vectors to continuously update the code vectors during the encoding and decoding of the video sequence [12].

7.3.3. *Rate-distortion optimized tone curves for HDR video compression*

Participants: David Gommelet, Christine Guillemot, Aline Roumy.

High Dynamic Range (HDR) images contain more intensity levels than traditional image formats. Instead of 8 or 10 bit integers, floating point values requiring much higher precision are used to represent the pixel data. These data thus need specific compression algorithms. In collaboration with Envivio, we have developed a novel compression algorithm that allows compatibility with the existing Low Dynamic Range (LDR) broadcast architecture in terms of display, compression algorithm and datarate, while delivering full HDR data to the users equipped with HDR display. The developed algorithm is thus a scalable video compression offering a base layer that corresponds to the LDR data and an enhancement layer, which together with the base layer corresponds to the HDR data. The novelty of the approach relies on the optimization of a mapping called Tone Mapping Operator (TMO) that maps efficiently the HDR data to the LDR data. The optimization has been carried out in a rate-distortion sense: the distortion of the HDR data is minimized under the constraint of minimum sum datarate (for the base and enhancement layer), while offering LDR data that are close to some "aesthetic" a priori. Taking into account the aesthetic of the scene in video compression is novel, since video compression is traditionally optimized to deliver the smallest distortion with the input data at the minimum datarate.

7.3.4. *Local Inverse Tone Curve Learning for HDR Image Scalable Compression*

Participants: Christine Guillemot, Mikael Le Pendu.

In collaboration with Technicolor, we have developed local inverse tone mapping operators for scalable high dynamic range (HDR) image coding. The base layer is a low dynamic range (LDR) version of the image that may have been generated by an arbitrary Tone Mapping Operator (TMO). No restriction is imposed on the TMO, which can be either global or local, so as to fully respect the artistic intent of the producer. The method which has been developed successfully handles the case of complex local TMOs thanks to a block-wise and non-linear approach [28]. A novel template based Inter Layer Prediction (ILP) is designed in order to perform the inverse tone mapping of a block without the need to transmit any additional parameter to the decoder. This method enables the use of a more accurate inverse tone mapping model than the simple linear regression commonly used for blockwise ILP [21]. In addition, this paper shows that a linear adjustment of the initially predicted block can further improve the overall coding performance by using an efficient encoding scheme of the scaling parameters. Our experiments have shown an average bitrate saving of 47% on the HDR enhancement layer, compared to previous local ILP methods.

7.3.5. *HEVC-based UHD video coding optimization*

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

The HEVC (High Efficiency Video Coding) standard brings the necessary quality versus rate performance for efficient transmission of Ultra High Definition formats (UHD). However, one of the remaining barriers to its adoption for UHD content is the high encoding complexity. We address the reduction of HEVC encoding complexity by investigating different strategies: First we have proposed to infer UHD coding modes and quad-tree from a first encoding pass which consists in encoding a lower resolution version of the input video. In the context of our study, the first encoding pass encodes a HD video sequence. A speed-up by a factor of 3 is achieved compared to directly encoding the UHD format without compromising the final video quality. The second strategy focuses on the block partitioning of intra frame coding. The Coding Tree Unit (CTU) is the root of the coding tree and can be recursively split into four square Coding Unit (CU), given that the smallest block size is 8×8. Once the partitioning procedure is fully completed, the final quad-tree can be obtained by choosing the configuration leading to the best rate-distortion trade-off. Rather than performing an exhaustive partitioning, we aim to predict the quad-tree partition into coding units (CU). This prediction is based on low-level visual features extracted from the video sequences. The low-level features are related to gradient-based statistics, structure tensors statistics or entropy etc. From these features, we trained a probabilistic model on a set of UHD training sequences in order to determine whether the coding unit should be further split or not. The proposed methods yield a significant encoder speed-up ratio (up to 5.3 times faster) with a moderate loss in terms of compression efficiency [33].

7.4. 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

7.4.1. Information theoretical bounds of Free-viewpoint TV

Participants: Thomas Maugey, Aline Roumy.

Free-viewpoint television FTV is a new system for viewing video where the user can choose its viewpoint freely and change it at anytime. The goal is to propose an immersive sensation without the disadvantage of Three-dimensional (3D) television (TV). Indeed, the conventional 3D displays (with or without glasses) occur, by construction, an accommodation-vergence conflict: since the eye tend to focus on the display screen (accommodation), whereas the brain perceives the depth of 3D images due to the different views seen by each eye (vergence). Instead, with FTV, a look-around effect is produced without any visual fatigue since the displayed images remain 2D. Therefore, FTV presents nice properties that makes it a serious competitor for 3DTV. Existing compression algorithms for FTV consider to send all the views, which would require about 100 Mbits/s (for 100 views, as needed to propose a true navigation within the scene). Since this amount does not fit the current datarate for transmission in a streaming scenario, we investigate a solution where the server only send the request. In [31], [30], we have shown a very surprising and positive result: if all the views are compressed once and if the server extracts from the compressed bitstream the request (i.e. one view at a time), the datarate is exactly the same as if the whole database was entirely decoded, and the requested views reencoded. This very positive result shows that it is possible to send FTV with the same datarate as single view television with very limited computational cost at the server (only extraction from the bistream). This result is an information theoretical result and the goal is now to build a practical system that can achieve this performance.

7.4.2. Compressed Sensing : a probabilistic analysis of the orthogonal matching pursuit algorithm

Participant: Aline Roumy.

Compressed sensing (CS) is an efficient acquisition scheme, where the data are projected onto a randomly chosen subspace to achieve data dimensionality reduction. The projected data are called measurements. The reconstruction is performed from these measurements, by solving underdetermined linear systems under a sparsity a priori constraint. It is generally believed that the greedy algorithm Orthogonal Matching pursuit performs well and can determine which variables are active (i.e. non zero). In contrast, we showed that this is not the case even in the noiseless context. We derived an exact probabilistic analysis of the iterative algorithm in the large system regime, when all dimensions tend to infinity. We showed that as the number of iterations grows, the algorithm will make errors with probability one.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Grants with Industry

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

Participants: Christine Guillemot, Bihong Huang.

- Title : Generalized lifting for video compression.
- Research axis : § 7.3.2.
- Partners : Orange Labs, Inria-Rennes, UPC-Barcelona.
- Funding : Orange Labs.
- Period : Apr.2012-Mar.2015.

This contract with Orange labs. (started in April. 2012) concerns the PhD of Bihong Huang and aims at modelling the redundancy which remains in spatial and temporal prediction residues. The analysis carried out in the first year of the PhD has shown that this redundancy (hence the potential rate saving) is high. In 2013, different methods have been investigated to remove this redundancy, such as generalized lifting and different types of predictors. 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 (Prof. Philippe Salembier) in Barcelona.

8.1.2. CIFRE contract with Technicolor on High Dynamic Range (HDR) video compression

Participants: Mikael Le Pendu, Christine Guillemot.

- Title : Floating point high dynamic range (HDR) video compression
- Research axis : § 7.3.4.
- Partners : Technicolor, Inria-Rennes.
- Funding : Technicolor, ANRT.
- Period : Dec.2012-Nov.2015.

High Dynamic Range (HDR) images contain more intensity levels than traditional image formats, leading to higher volumes of data. HDR images can represent more accurately the range of intensity levels found in real scenes, from direct sunlight to faint starlight. The goal of the thesis is to design a visually lossless compression algorithm for HDR floating-point imaging data. The first year of the thesis has been dedicated to the design of a quantization method converting the floating point data into a reduced bit depth representation, with minimal loss. The method leads to a bit rate saving of 50% compared to the existing Adaptive LogLuv transform.

8.1.3. CIFRE contract with Technicolor on sparse modelling of spatio-temporal scenes

Participants: Martin Alain, Christine Guillemot.

- Title : Spatio-temporal analysis and characterization of video scenes
- Research axis : § 7.1.2.
- Partners : Technicolor, Inria-Rennes.
- Funding : Technicolor, ANRT.
- Period : Oct.2012-Sept.2015.

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 using models based on sparse and data dimensionality reduction techniques, as well as based on epitomes. The objective was then to use these models and methods in different image processing problems focusing in particular on video compression. While, 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 started in Oct. 2012 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 denoising and super-resolution.

8.1.4. 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 pre-analysis approaches.
- Research axis : § 7.3.5.
- Partners : Thomson Video Networks, Univ. Rennes 1.
- Funding : Thomson Video Networks (TVN).
- Period : Nov.2012-Sept.2015.

This contract with TVN (started 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 infer the best coding decisions (prediction modes...) in order to reduce the computational complexity of HEVC thanks to a pre-analysis step. The pre-analysis is expected to provide useful estimates of local video characteristics which will then help selecting the prediction and transform partitions as well as a number of other parameters such as the quantization parameters or the prediction modes.

8.1.5. CIFRE contract with Envivio on LDR compatible HDR video coding

Participants: Christine Guillemot, David Gommelet, Aline Roumy.

- Title : LDR-compatible coding of HDR video signals.
- Research axis : § 7.3.3.
- Partners : Envivio.
- Funding : Cifre Envivio.
- Period : Oct.2014-Sept.2017.

The goal of this Cifre contract is to design solutions for LDR-compatible coding of HDR videos. This involves the study of rate-distortion optimized tone mapping operators taking into account constraints of temporal coherency to avoid the temporal flickering which results from a direct frame-by-frame application of classical tone mapping operators. The goal is also to design a coding architecture which will build upon these operators, integrating coding tools tailored to the statistics of the HDR refinement signals.

8.1.6. CIFRE contract with Technicolor on light fields editing

Participants: Christine Guillemot, Matthieu Hog.

- Title : Light fields editing
- Research axis : *just started*
- Partners : Technicolor, Inria-Rennes.
- Funding : Technicolor, ANRT.
- Period : Oct.2015-Sept.2018.

Editing is quite common with classical imaging. Now, if we want light-fields cameras to be in the future as common as traditional cameras, this functionality should also be enabled with light-fields. The goal of the PhD will therefore be to develop methods for light-field editing focusing first on object removal thanks to light-fields inpainting and for constructing panoramic images based on light-fields stitching. This objective also includes the development of algorithms for dynamic light fields spatio-temporal segmentation with spatio-temporal coherence constraints across sub-aperture images.

8.1.7. CIFRE contract with Technicolor on cloud-based video compression

Participants: Jean Begaint, Christine Guillemot.

- Title : Cloud-based video compression
- Research axis : *just started*
- Partners : Technicolor, Inria-Rennes.
- Funding : Technicolor, ANRT.
- Period : Nov.2015-Oct.2018.

The goal of this Cifre contract is to develop a novel image compression scheme exploiting similarity between images in a cloud. The objective will therefore be to develop rate-distortion optimized affine or homographic estimation and compensation methods which will allow us to construct prediction schemes and learn adapted bases from most similar images retrieved by image descriptors. One issue to be addressed is the rate-distortion trade-off induced by the need for transmitting image descriptors.

9. Partnerships and Cooperations

9.1. Regional Initiatives

- T. Maugey has received a grant for scientific intallation from Rennes Metropole.
- The postdoc of Xin Su on multi-view data representation and compression is partly funded (at the level of 75%) by the Brittany region.

9.2. International Initiatives

9.2.1. Inria International Partners

9.2.1.1. Informal International Partners

- The study on guided image inpainting is carried out in collaboration with Prof. Pascal Frossard from EPFL (Ecole Polytechique Federal de Lausanne).
- The study on adaptive clustering with Kohonen self-organizing maps for second-order prediction has been carried out in collaboration with Prof. Philippe Salembier from UPC (Universitat Politecnica De Catalunya).

9.3. International Research Visitors

9.3.1. Visits of International Scientists

- Pr. Reuben Farrugia from Malta University, is spending one sabbatical year in the team from Sept. 2015 until August 2016.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific events organisation

10.1.1.1. Member of the organizing committees

- C. Guillemot has been area chair of Eusipco 2015 and IEEE-ICIP 2015, and award chair of IEEE-ICME 2015.
- A. Roumy is local liaison officer for Eurasip.

10.1.2. Scientific events selection

10.1.2.1. Member of the conference program committees

- C. Guillemot has been a member of technical program committees of international conferences: EUSIPCO 2015, IEEE-ICIP 2015.
- C. Labit is member of the GRETSI technical program committee.
- O. Le Meur has been a member of technical program committees of international conferences: IEEE-ICME 2015, QoMex 2015.
- O. Le Meur co-organized a special session on Visual attention modelling at EUSIPCO'2015.
- A. Roumy is member of the GRETSI technical program committee.

10.1.3. Journal

10.1.3.1. Member of the editorial boards

- C. Guillemot is associate editor of the Eurasip International Journal on Image Communication.
- C. Guillemot is associate editor of the IEEE Trans. on Image Processing.
- C. Guillemot is associate editor of the International Journal on Mathematical Imaging and Vision.
- C. Guillemot is senior member of the editorial board of the IEEE Journal on selected topics in signal processing.
- O. Le Meur has been guest editor of the Special Issue on Recent Advances in Saliency Models, Applications and Evaluations, in Signal Processing: Image communication journal.
- A. Roumy is associate editor of the Hindawi Journal on Mathematical Problems in Engineering.

10.1.4. Invited talks

- C. Guillemot has been invited as member of a panel on computational imaging at IEEE-ICME 2015.
- O. Le Meur gave a seminar dealing with examplar-based inpainting at INSA Rennes, June 2015.
- T. Maugey has been invited for a talk at INSA Rennes on “Interactive multi-view video coding", June 2015.
- T. Maugey has been invited for a talk at the GDR-Isis meeting on “Multi-view imaging from acquisition to rendering", June 2015.

10.1.5. Leadership within the scientific community

- C. Guillemot is member of the IEEE IVMSP technical committee.
- A. Roumy has co-animated with F. Bimbot and N. Bertin (PANAMA team, IRISA)the workshop “Les défis de dimensionnalité soulevés par les nouvelles générations de capteurs audiovisuels", at the Data Science Symposium, IRISA, Nov. 2015.

10.1.6. Scientific expertise

- 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 (until June 2015).
- 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. Labit is member of the ICT strategic steering committee (CPS-7) of the National Research Agency (ANR).

- A. Roumy is a member of the Selection and Evaluation Committee (CDT Commission Developpement Technologique) for the Inria technological development grants.

10.1.7. Research administration

- C. Guillemot is, since Sept. 2015, vice-chair of Inria's evaluation committee.
- C. Guillemot is member of the "bureau du Comité des Projets".
- C. Labit is the Vice-president of the Scientific Board, in charge of Research and Innovation, for the University of Rennes1 (since June 1st, 2008).
- C. Labit is president of the Rennes-Atalante Science Park and of the start-up incubator Emergys (since April, 2007).
- A. Roumy is titular member of the National Council of Universities (CNU section 61, 2012-2015).

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Master: C. Guillemot, Image and video compression, 8 hours, M2 computer science, Univ. of Rennes 1, France.

Master: C. Guillemot, Image and video compression, 8 hours, M2 SISEA, 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: T. Maugey, "Multi-view / 3D video coding", 2H, EPFL, Lausanne, Switzerland, March 2015.

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

Engineer degree: C. Guillemot, Video communication, 10 hours, Télécom Lille 1, Villeneuve-d'Ascq, France.

Engineer 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.

Engineer degree: A. Roumy, Image processing, 14 hours, ECAM Rennes, France.

Professional training: O. Le Meur, Image Processing and OpenCV, 42 hours, Technicolor Rennes.

10.2.2. Supervision (PhDs defended during the year)

PhD : B. Huang, Second-order prediction and residue vector quantization for video compression, University of Rennes 1, defense on the 8th of July 2015, C. Guillemot (contract with Orange labs.)

PhD : J. Aghaei Mazaheri, Representations parcimonieuses et apprentissage de dictionnaires pour la compression et la classification d'images satellites, Univ. of Rennes 1, defense on the 20th of July 2015, C. Labit and C. Guillemot (contract with Astrium)

10.2.3. Juries

- C. Guillemot has been member (rapporteur) of the jury of the PhD committee of:
  - Y. Xing, Télécom ParisTech, Jan. 2015
  - M. Suryanarayana, MidSweden Univ., June 2015
  - M. Farajallah, Univ. Nantes, June 2015
  - M. S. Farid, Universita Degli Studi Di Torino, Sept. 2015

  - S. Karygianni, EPFL, Oct. 2015
- C. Guillemot has been member (president) of the jury of the PhD committee of:
  - M. Daisy, Univ. of Caen, Dec. 2015
  - A. Arrufat, INSA of Rennes, Dec. 2015
  - A. Basset, Univ. Rennes 1, Dec. 2015
- O. Le Meur has been member (rapporteur) of the jury of the PhD committee of:
  - S. Lemonnier, Univ. Paris VIII, November 2015
- O. Le Meur has been member (examiner) of the jury of the PhD committee of:
  - L. Yi, INSA of Rennes, March 2015
  - J. Liang, Univ. of Hong Kong, April 2015
  - N. Riche, Univ. of Mons, October 2015

11. 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, pp. 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, pp. 1–32

[3] C. Guillemot, O. Le Meur. *Image inpainting: Overview and recent advances*, in "IEEE Signal Processing Magazine", Jan. 2014, vol. 31, n^{o} 1, pp. 127-144

[4] C. Guillemot, A. Roumy. *Towards constructive Slepian-Wolf coding schemes*, in "Distributed source coding", Elsevier Inc., 2008

[5] 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, pp. 1393 - 1407, http://hal.inria.fr/inria-00604036

[6] 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, pp. 2493-2498

[7] 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, pp. 802-817

[8] T. Maugey, A. Ortega, P. Frossard. *Graph-based representation for multiview image geometry*, in "IEEE Transactions on Image Processing", May 2015, vol. 24, n^{o} 5, pp. 1573-1586 [*DOI :* 10.1109/TIP.2015.2400817], https://hal.inria.fr/hal-01116211

[9] 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

[10] J. ZEPEDA, C. GUILLEMOT, E. KIJAK. *Image compression using sparse representations and the iteration-tuned and aligned dictionary*, in "IEEE Journal on Selected Topics in Signal Processing", Sep. 2011, vol. 5, pp. 1061-1073

Publications of the year

Doctoral Dissertations and Habilitation Theses

[11] J. AGHAEI MAZAHERI. *Sparse representations and dictionary learning for the compression and the classification of satellite images*, Université Rennes 1, July 2015, https://hal.archives-ouvertes.fr/tel-01205490

[12] B. HUANG. *Second-order prediction and residue vector quantization for video compression*, Université Rennes 1, July 2015, https://tel.archives-ouvertes.fr/tel-01206572

[13] D. KHAUSTOVA. *Objective assessment of stereoscopic video quality of 3DTV*, Université Rennes 1, January 2015, https://tel.archives-ouvertes.fr/tel-01193103

Articles in International Peer-Reviewed Journals

[14] M. ALAIN, C. GUILLEMOT, D. THOREAU, P. GUILLOTEL. *Inter-prediction methods based on linear embedding for video compression*, in "Signal Processing: Image Communication", September 2015, vol. 37, pp. 47-57 [*DOI :* 10.1016/J.IMAGE.2015.07.011], https://hal.inria.fr/hal-01204761

[15] A. DE ABREU, L. TONI, N. THOMOS, T. MAUGEY, F. PEREIRA, P. FROSSARD. *Optimal Layered Representation for Adaptive Interactive Multiview Video Streaming*, in "Journal of Visual Communication and Image Representation", 2015, forthcoming, https://hal.inria.fr/hal-01203250

[16] M. EBDELLI, O. LE MEUR, C. GUILLEMOT. *Video inpainting with short-term windows: application to object removal and error concealment*, in "IEEE Transaction on Image Processing", October 2015, vol. 24, n^{o} 10, pp. 3034-47 [*DOI :* 10.1109/TIP.2015.2437193], https://hal.inria.fr/hal-01204677

[17] Y. GAO, G. CHEUNG, T. MAUGEY, P. FROSSARD, J. LIANG. *Encoder-Driven Inpainting Strategy in Multiview Video Compression*, in "IEEE Transactions on Image Processing", 2015, forthcoming, https://hal.inria.fr/hal-01217115

[18] S. KHATTAK, T. MAUGEY, R. HAMZAOUI, S. AHMAD, P. FROSSARD. *Temporal and Inter-view Consistent Error Concealment Technique for Multiview plus Depth Video Broadcasting*, in "IEEE Transactions on Circuits and Systems for Video Technology", 2015, forthcoming [*DOI :* 10.1109/TCSVT.2015.2418631], https://hal.inria.fr/hal-01137927

[19] D. KHAUSTOVA, J. FOURNIER, O. LE MEUR. *An Objective Metric for Stereoscopic 3D Video Quality Prediction Using Perceptual Thresholds*, in "Smpte Motion Imaging Journal", March 2015, vol. 124, n^{o} 2, pp. 47-55, https://hal.inria.fr/hal-01204808

[20] O. LE MEUR, Z. LIU. *Saccadic model of eye movements for free-viewing condition*, in "Vision Research", February 2015, https://hal.inria.fr/hal-01204682

[21] M. LE PENDU, C. GUILLEMOT, D. THOREAU. *Local Inverse Tone Curve Learning for High Dynamic Range Image Scalable Compression*, in "IEEE Transactions on Image Processing", September 2015, vol. 24, n^{o} 12, pp. 5753-5763 [*DOI :* 10.1109/TIP.2015.2483899], https://hal.inria.fr/hal-01204722

[22] J. Li, Z. Liu, X. Zhang, O. Le Meur, L. Shen. *Spatiotemporal Saliency Detection Based on Superpixel-level Trajectory*, in "Signal Processing: Image Communication", May 2015, vol. 38, pp. 100–114 [*DOI :* 10.1016/j.image.2015.04.014], https://hal.inria.fr/hal-01204803

[23] T. Maugey, A. Ortega, P. Frossard. *Graph-based representation for multiview image geometry*, in "IEEE Transactions on Image Processing", May 2015, vol. 24, n^o 5, pp. 1573-1586 [*DOI :* 10.1109/TIP.2015.2400817], https://hal.inria.fr/hal-01116211

[24] L. Toni, T. Maugey, P. FROSSARD. *Optimized Packet Scheduling in Multiview Video Navigation Systems*, in "IEEE Transactions on Multimedia", September 2015, vol. 17, n^o 9, pp. 1604-1616 [*DOI :* 10.1109/TMM.2015.2450020], https://hal.inria.fr/hal-01169278

[25] W. Zou, Z. Liu, K. Kpalma, J. Ronsin, Y. Zhao, N. Komodakis. *Unsupervised Joint Salient Region Detection and Object Segmentation*, in "IEEE Transactions on Image Processing", November 2015, vol. 24, n^o 11 [*DOI :* 10.1109/TIP.2015.2456497], https://hal.archives-ouvertes.fr/hal-01243552

International Conferences with Proceedings

[26] B. HUANG, F. HENRY, C. Guillemot, P. Salembier. *Mode dependent vector quantization with a rate-distortion optimized codebook for residue coding in video compression*, in "IEEE Intl. Conf. on Acoustics and Signal Processing (IEEE-ICASSP)", Brisbane, Australia, April 2015, https://hal.archives-ouvertes.fr/hal-01204768

[27] H. Hristina, O. Le Meur, R. Cozot, K. Bouatouch. *Style-aware robust color transfer*, in "Computational Aesthetics in Graphics, Visualization, and Imaging", Istambul, Turkey, June 2015, https://hal.inria.fr/hal-01219789

[28] M. Le Pendu, C. Guillemot, D. Thoreau. *Template based Inter-Layer Prediction for High Dynamic Range Scalable compression*, in "IEEE International Conference on Image Processing (ICIP)", Quebec, Canada, September 2015, https://hal.inria.fr/hal-01204715

[29] T. Maugey, P. FROSSARD, C. Guillemot. *Guided inpainting with cluster-based auxiliary information*, in "IEEE ICIP", Québec, Canada, September 2015, https://hal.inria.fr/hal-01204698

[30] A. Roumy, T. Maugey. *Compression et interactivite : etude de la navigation au recepteur*, in "colloque international francophone de Traitement du Signal et de l'Image, GRETSI", Lyon, France, September 2015, https://hal.inria.fr/hal-01208129

[31] A. Roumy, T. Maugey. *Universal lossless coding with random user access: the cost of interactivity*, in "IEEE International Conference on Image Processing (ICIP)", Quebec, Canada, September 2015, https://hal.inria.fr/hal-01208128

[32] M. Turkan, M. Alain, D. Thoreau, P. Guillotel, C. Guillemot. *Epitomic image factorization via neighbor-embedding*, in "2015 IEEE International Conference on Image Processing (IEEE-ICIP)", Quebec City, Canada, September 2015, https://hal.inria.fr/hal-01204755

Conferences without Proceedings

[33] N. Dhollande, X. Ducloux, O. Le Meur, C. Guillemot. *Fast block partitioning method in HEVC Intra coding for UHD video*, in "IEEE International Conference on Consumer Electronics", Berlin, Germany, September 2015, https://hal.inria.fr/hal-01204835

[34] H. Hristina, R. Cozot, O. Le Meur, K. Bouatouch. *Color transfer between high-dynamic-range images*, in "SPIE 9599, Applications of Digital Image Processing XXXVIII", san diego, United States, August 2015, https://hal.inria.fr/hal-01219780

[35] D. Khaustova, J. Fournier, E. Wyckens, O. Le Meur. *An objective method for 3D quality prediction using perceptual thresholds and acceptability*, in "SPIE 9391, Stereoscopic Displays and Applications XXVI", San Francisco, United States, February 2015, https://hal.inria.fr/hal-01204819