Activity Report 2017

Team SIROCCO

Analysis Representation, Compression and Communication of Visual Data

Joint team with Inria Rennes – Bretagne Atlantique

D5 – Digital Signals and Images, Robotics

Table of contents

Project-Team SIROCCO

Creation of the Project-Team: 2012 January 01

Keywords:

Computer Science and Digital Science:

A5. - Interaction, multimedia and robotics
A5.3. - Image processing and analysis
A5.4. - Computer vision
A5.9. - Signal processing

Other Research Topics and Application Domains:

B6. - IT and telecom

1. Personnel

Research Scientists

Christine Guillemot [Team leader, Inria, Senior Researcher, HDR]
Thomas Maugey [Inria, Researcher]
Aline Roumy [Inria, Researcher]

Faculty Member

Olivier Le Meur [Univ de Rennes I, Associate Professor, until Sept. 2017, HDR]

Post-Doctoral Fellows

Oriel Frigo [Inria, until Aug 2017]
Xiaoran Jiang [Inria]
Lara Younes [Inria, from Mar 2017]

PhD Students

Jean Begaint [Technicolor]
Pierre David [Inria]
Elian Dib [Inria]
Thierry Dumas [Inria]
Simon Evain [Inria, from Dec 2017]
David Gommelet [ERICSSON, until Dec. 2017]
Fatma Hawary [Technicolor]
Matthieu Hog [Technicolor]
Hristina Hristova [Univ de Rennes 1, partly with FRVSense, until Sept. 2017]
Dmitry Kuzovkin [Technicolor, partly with FRVSense, until Sept. 2017]
Navid Mahmoudian Bidgoli [Inria]
Mira Rizkallah [Univ de Rennes I]
Jinglei Shi [Inria, from Oct 2017]

Technical staff

Laurent Guillo [CNRS, Research Engineer]
Pierre Allain [Inria, from May 2017]
Cédric Le Cam [Inria]
Mikael Le Pendu [Inria, until Jul 2017, granted by Google Inc]
Xin Su [Inria]

Administrative Assistant

Huguette Bechu [Inria]

Visiting Scientists

Ole Johannsen [University of Konstanz, from Nov. 2017 until Dec. 2017]
Chiara Galdi [Eurecom, May 2017]
Ehsan Miandji [University of Linkoping, from June 2017 until Jul. 2017]

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 visual data, i.e. natural 2D images, videos, multi-view sequences with a focus on recent imaging modalities such as high dynamic range imaging, light fields, and 360° videos. 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 processing
- Representation and compression of visual data
- Restoration, super-resolution, editing
- Distributed processing and robust communication

While aiming at generic approaches, some of the solutions developed are applied to practical problems in partnership with industry (Technicolor, Ericsson, DGA) or in the framework of national projects. The application domains addressed by the project are networked visual applications taking into account 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 and efficient processing

Analysis and modeling of the visual data are crucial steps for a number of video processing problems: 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 (natural 2D still and moving images, multi-view and multi-view plus depth video, high dynamic range images, light fields, 360° video content).

2.3. Restoration, super-resolution, editing

Depending on the application and the type of content, various issues are being addressed, such as restoring the data to cope with limitations of sensors or to recover the signal from compression artefacts. The design of efficient methods for enhancing the data resolution in spatial, temporal or angular (in the case of light fields) dimensions is also part of the project-team objectives to cope with limitations of the visual sensors or for anti-aliased image-based rendering, e.g. from light fields. View synthesis remains a difficult ill-posed problem related to angular super-resolution of multi-view data and light fields which keeps evolving with in particular the use of deep learning techniques. Estimating the scene geometry and the scene flow for dynamic scenes is a critical step of the above processing problems. Editing the content, or restoring disoccluded areas in multi-view processing calls for efficient içnpainting techniques.

2.4. Representation and compression of visual data

The objective is to develop algorithmic tools for constructing low-dimensional representations of various imaging modalities (2D images and videos, multi-view, light fields, ...). 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 using overcomplete dictionaries, transforms on graphs, or autoencoders based on deep learning architectures. Low rank and sparse models are the essence of transform coding and of many other processing methods (e.g., denoising, classification, registration, super-resolution, inpainting). Developing complete compression algorithms necessarily requires tackling topics beyond the issues of sparse data representation and dimensionality reduction. For example, the problem of spatial, inter-view or temporal prediction using deep learning techniques is also addressed. Finally, rate-distortion models for constructing rate-efficient representations with various features of scalability or low dynamic range compatibility in the case of high dynamic range content are also studied.

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 [1]. 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.

3.4. Perceptual Modelling

Saliency, visual attention, cognition

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

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 [2]. This detection is also influenced by prior knowledge about the environment of the scene [3]. 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 [4]. 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 [5] 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 [6] (SW) and Wyner-Ziv [7] (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

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

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

[4] 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.

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

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

[7] 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.

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. Overview

The application domains addressed by the project are:

- Compression with advanced functionalities of various imaging modalities
- Networked multimedia applications taking into account needs in terms of user and network adaptation (e.g., interactive streaming, resilience to channel noise)
- Content editing, post-production, and computational photography.

4.2. Compression of emerging imaging modalities

Compression of visual content 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), or of formats for immersive displays allowing panoramic viewing, Free Viewpoint Video (FVV) 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). 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

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. This application sector suffers from an accommodation-vergence conflict which arises with conventional 3D displays (with or without glasses). Since each eye receives a single view, the eyes 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).

On the other hand, Free-viewpoint television (FTV) is a system for watching videos in which the user can choose its viewpoint freely and change it at anytime. To allow this navigation, many views are proposed and the user can navigate from one to the other. The goal of FTV is to propose an immersive sensation without the disadvantage of Three-dimensional television (3DTV). With FTV, a look-around effect is produced without any visual fatigue since the displayed images remain 2D. However, technical characteristics of FTV are large databases, huge numbers of users, and requests of subsets of the data, while the subset can be randomly chosen by the viewer. This requires the design of coding algorithms allowing such a random access to the pre-encoded and stored data which would preserve the compression performance of predictive coding. This research also finds applications in the context of Internet of Things in which the problem arises of optimally selecting both the number and the position of reference sensors and of compressing the captured data to be shared among a high number of users.

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).

4.4. Editing, post-production and computational photography

Editing and post-production are critical aspects in the audio-visual production process. Increased ways of "consuming" visual content also highlight the need for content repurposing as well as for higher interaction and editing capabilities. 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. Our activies around light field imaging also find applications in computational photography which refers to the capability of creating photographic functionalities beyond what is possible with traditional cameras and processing tools.

5. Highlights of the Year

5.1. Highlights of the Year

5.1.1. Awards

BEST PAPER AWARD:

[27]
P. DAVID, M. LE PENDU, C. GUILLEMOT. *White Lenslet Image Guided Demosaicing for Plenoptic Cameras*, in "MMSP 2017 - IEEE 19th International Workshop on Multimedia Signal Processing", Luton, United Kingdom, Multimedia Signal Processing (MMSP), 2017 IEEE 19th International Workshop on, October 2017, https://hal.archives-ouvertes.fr/hal-01590345

6. New Software and Platforms

6.1. SaccadicModel

Saccadic model of visual attention
KEYWORDS: Visual saliency maps - Visual scanpath

FUNCTIONAL DESCRIPTION: Saliency models compute a saliency map from an input image. Saliency maps are a 2D map encoding the ability of every location to attract our gaze. There exist many models in the literature and tremendous progresses have been made. However, they remain quite limited when applied to natural scene exploration. Indeed, the vast majority of these models ignore fundamental properties of our visual system. The most important one is that they overlook the sequential and time-varying aspects of overt attention. Saccadic models aim to predict the visual scanpath itself, i.e. the series of fixations and saccades an observer would perform to sample the visual environment. We propose a new and efficient method to simulate the visual scanpath. It provides scanpaths in close agreement with human behavior and the model can be tailored to simulate scanpaths in specific conditions and for various observer profiles.

- Author: Olivier Le Meur
- Contact: Olivier Le Meur

6.2. QuantizationAE

KEYWORDS: Compression - Machine learning
FUNCTIONAL DESCRIPTION: This code learns an autoencoder to compress images. The learning is performed under a rate-distortion criterion, and jointly learns a transform (the autoencoder) and the quantization step for target rate points.The code is organized as follows. It first builds a set of luminance images (B1) for the auto-encoder training, a set of luminance images (B2) to analyze how the auto-encoder training advances and a set of luminance images (B3) to evaluate the auto-encoders in terms of rate-distortion. It then trains several auto-encoders using a rate-distortion criterion on the set B1. The quantization can be either fixed or learned during this training stage. The set B2 enables to periodically compute indicators to detect overfitting. It finally compares the auto-encoders in terms of rate-distortion on the set B3. The quantization can be either fixed or variable during this test.

- Participants: Aline Roumy, Christine Guillemot and Thierry Dumas
- Contact: Aline Roumy

6.3. LF-Inpainting

Light field inpainting based on a low rank model
KEYWORDS: Light fields - Low rank models - Inpainting
FUNCTIONAL DESCRIPTION: This code implements a method for propagating the inpainting of the central view of a light field to all the other views. To this end, it also implements a new matrix completion algorithm, better suited to the inpainting application than existing methods. A first option does not require any depth prior, unlike most existing light field inpainting algorithms. The code also implements an extended version to better handle the case where the area to inpaint contains depth discontinuities.

- Participants: Mikael Le Pendu and Christine Guillemot
- Contact: Christine Guillemot

6.4. LF-HLRA

Light fields homography-based low rank approximation
KEYWORDS: Compression - Light fields - Low rank models - Dimensionality reduction
FUNCTIONAL DESCRIPTION: This code jointly searches for homographies to align the views of an input light field together with the components of its low rank approximation model. The code either uses a global homography per view or multiple homographies, one per region, the region being extracted using depth information.

- Participants: Xiaoran Jiang, Mikael Le Pendu and Christine Guillemot
- Contact: Christine Guillemot

6.5. GBR-MVimages

Graph-based Representation for multi-view and light field images
KEYWORDS: Light fields - Multi-View reconstruction - Graph
FUNCTIONAL DESCRIPTION: Graph-Based Representation (GBR) describes color and geometry of multiview or light field image content using a graph. The graph vertices represent the color information, while the edges represent the geometry information, i.e. the disparity, by connecting corresponding pixels in neighboring images.

- Participants: Xin Su and Thomas Maugey
- Contact: Thomas Maugey

6.6. Platforms

6.6.1. Light field editor

Participants: Pierre Allain, Laurent Guillo, Christine Guillemot.

As part of the ERC Clim project, the EPI Sirocco is developing a light field editor, a tool analogous to traditional image editors such as the GNU image manipulation program Gimp or the raster graphic editor Photoshop but dedicated to light fields. As input data, this tool accepts for instance sparse light fields acquired with High Density Camera Arrays (HDCA) or denser light fields captured with microlens array (MLA). Two kinds of features are provided. Traditional features such as changing the angle of view, refocusing or depth map extraction are or will be soon supported. More advanced features are being integrated in our tool as libraries we have developed, such as segmenation or inpainting. For instance, a segmentation on a specific subaperture/view of light fields can be propagated to all subapertures/views. Thus, the so-segmented objects or zones can be colourized or even removed, the emptied zone being then inpainted. The tool and libraries are developed in C++ and the graphical user interface relies on Qt.

6.6.2. Acquisition of multi-view sequences for Free viewpoint Television

Participants: Cédric Le Cam, Laurent Guillo, Thomas Maugey.

The scientific and industrial community is nowadays exploring new multimedia applications using 3D data (beyond stereoscopy). In particular, Free Viewpoint Television (FTV) has attracted much attention in the recent years. In those systems, user can choose in real time its view angle from which he wants to observe the scene. Despite the great interest for FTV, the lack of realistic and ambitious datasets penalizes the research effort. The acquisition of such sequences is very costly in terms of hardware and working effort, which explains why no multi-view videos suitable for FTV has been proposed yet.

In 2017, in the context of the project ADT ATeP (funded by Inriahub), such datasets have been acquired and some calibration tools have been developed. First 40 omnidirectional cameras and their associated equipments have been acquired by the team (thanks to Rennes Metropole funding). We have first focused on the calibration of this camera, *i.e.,* the development of the relationship between a 3D point and its projection in the omnidirectional image. In particular, we have shown that the unified spherical model fits the acquired omnidirectional cameras. Second, we have developed tools to calibrate the cameras in relation to each other. Finally, we have made a capture of 3 multiview sequences that are currently in preparation for a sharing with the community (Fig. 1). This work has been published in [41].

6.6.3. Light fields datasets

Participants: Pierre Allain, Christine Guillemot, Laurent Guillo.

The EPI Sirocco makes extensive use of light field datasets with sparse or dense contents provided by the scientific community to run tests. However,it has also generated its own natural and synthetic contents.

Figure 1. The 40 omnidirectional cameras positioned for an indoor scene capture.

Natural content has been created with Lytro cameras (the original first generation Lytro and the Lytro Illum) and is already available to the community (https://www.irisa.fr/temics/demos/lightField/CLIM/DataSoftware.html). The team also owns a R8 Raytrix plenoptic cameras with which still and video contents have been captured. Applications taking advantage of the Raytrix API have been developed to extract views from the Raytrix lightfield. The number of views per frame is configurable and can be set for instance to 3x3 or 9x9 according to the desired sparsity.

Synthetic content has been generated from the Sintel film (https://durian.blender.org/download/), which is a short computer animated film by the Blender institute, part of the Blender Foundation. A specific Blender add-on is used to extract views from a frame. As previously, the number of views is configurable. Synthetic contents present the advantage to provide a ground truth useful to evaluate how efficient our algorithms are to compute, for instance, the depth maps.

7. New Results

7.1. Analysis and modeling for compact representation

3D modelling, light-fields, 3D meshes, epitomes, image-based rendering, inpainting, view synthesis

7.1.1. Visual attention

Participant: 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 deal with the understanding and modeling of overt attention.

Saccadic model: Since 2015, we have worked on saccadic model, which predicts the visual scanpaths of an observer watching a scene displayed onscreen. In 2016, we proposed a first improvement consisting in using spatially-variant and context-dependent viewing biases. We showed that the joint distribution of saccade amplitudes and orientations is significantly dependent on the type of visual stimulus. In addition, the joint distribution turns out to be spatially variant within the scene frame. This model outperforms state-of-the-art saliency models, and provides scanpaths in close agreement with human behavior. In [19], [35], we went further by showing that saccadic models are a flexible framework that can be tailored to emulate observer's viewing tendencies. More specifically, we tailored the proposed model to simulate visual scanpaths of 5 age groups of observers (i.e. adults, 8-10 y.o., 6-8 y.o., 4-6 y.o. and 2 y.o.). The key point is that the joint distribution of saccade amplitude and orientation is a visual signature specific to each age group, and can be used to generate age-dependent scanpaths. Our age-dependent saccadic model does not only output human-like, age-specific visual scanpaths, but also significantly outperforms other state-of-the-art saliency models. We demonstrated that the computational modelling of visual attention, through the use of saccadic model, can be efficiently adapted to emulate the gaze behavior of a specific group of observers.

Effects on Comics by Clustering Gaze Data: Comics are a compelling communication medium conveying a visual storytelling. With a smart mixture of text or/and other visual information, artists tell a story by drawing the viewer attention on specific areas. With the digital comics revolution (e.g. mobile comic and webcomic), we are witnessed a resurgence of interest for this art form. This new form of comics allows not only to tackle a wider audience but also new consumption methods. An open question in this endeavor is identifying where in a comic panel the effects should be placed. We proposed a fast, semi-automatic technique to identify effects-worthy segments in a comic panel by utilizing gaze locations as a proxy for the importance of a region. We took advantage of the fact that comic artists influence viewer gaze towards narrative important regions. By capturing gaze locations from multiple viewers, we can identify important regions. The key contribution is to leverage a theoretical breakthrough in the computer networks community towards robust and meaningful clustering of gaze locations into semantic regions, without needing the user to specify the number of clusters. We have developed a method based on the concept of relative eigen quality that takes a scanned comic image and a set of gaze points and produces an image segmentation. A variety of effects such as defocus, recoloring, stereoscopy, and animations has been demonstrated. We also investigated the use of artificially generated gaze locations from saliency models in place of actual gaze locations.

Perceptual metric for perceptual transfer: Color transfer between input and target images has raised a lot of interest in the past decade. 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. Different color transfer methods often result in different output images. The process of determining the most plausible output image is difficult and requires, due to the lack of an objective metric, time-consuming and costly subjective experiments. To overcome this problem, we proposed a perceptual model for evaluating results from color transfer methods [31]. From a subjective experiment, involving several color transfer methods, we build a regression model with random forests to describe the relationship between a set of features (e.g. objective quality, saliency, etc.) and the subjective scores. An analysis and a cross-validation showed that the predictions of the proposed quality metric are highly accurate.

7.1.2. *Saliency-based navigation in omnidirectional image*

Participants: Olivier Le Meur, Thomas Maugey.

Omnidirectional images describe the color information at a given position from all directions. Affordable 360° cameras have recently been developed leading to an explosion of the 360 degrees data shared on the social networks. However, an omnidirectional image does not contain interesting content everywhere. Some part of the images are indeed more likely to be looked at by some users than others. Knowing these regions of interest might be useful for 360° image compression, streaming, retargeting or even editing. In the work published in [25], a new approach based on 2D image saliency is proposed both to model the user navigation within a 360° image, and to detect which parts of an omnidirectional content might draw users' attention. A double cube projection is first used to put the saliency estimation in the classical 2D image framework. Consecutively, the saliency map serves as a support for the navigation estimation algorithm.

7.1.3. *Context-aware Clustering and Assessment of Photo Collections*

Participants: Dmitry Kuzovkin, Olivier Le Meur.

To ensure that all important moments of an event are represented and that challenging scenes are correctly captured, both amateur and professional photographers often opt for taking large quantities of photographs. As such, they are faced with the tedious task of organizing large collections and selecting the best images among similar variants. Automatic methods assisting with this task are based on independent assessment approaches, evaluating each image apart from other images in the collection. However, the overall quality of

Figure 2. Example of the saliency map (left) and the estimated navigation (right).

photo collections can largely vary due to user skills and other factors. We explore the possibility of context-aware image quality assessment, where the photo context is defined using a clustering approach, and statistics of both the extracted context and the entire photo collection are used to guide identification of low-quality photos. We demonstrate that the proposed method is able to adapt flexibly to the nature of processed albums and to facilitate the task of image selection in diverse scenarios.

7.1.4. Light fields view extraction from lenslet images

Participants: Pierre David, Christine Guillemot, Mikael Le Pendu.

Practical systems have recently emerged for the capture of real light fields which go from cameras arrays to single cameras mounted on moving gantries and plenoptic cameras. While camera arrays capture the scene from different viewpoints, hence with a large baseline, plenoptic cameras use an array of micro-lenses placed in front of the photosensor to separate the light rays striking each microlens into a small image on the photosensors pixels, and this way capture dense angular information with a small baseline. Extracting views from the raw lenslet data captured by plenoptic cameras involves several processing steps: devignetting which, with white images, aims at compensating for the loss of illumination at the periphery of the micro-lenses, color demosaicing, alignment of the sensor data with the micro-lens array, and converting the hexagonal sampling grid into a rectangular sampling grid. These steps are quite critical as they have a strong impact on the quality of the extracted sub-aperture images (views).

We have addressed two important steps of the view extraction from lenslet data: color demosaicing and alignment of the micro-lens array on the photosensor. We have developed a new method guided by a white lenslet image for color demosaicing of raw lenslet data [27](best paper award). The white lenslet image gives measures of confidence on the color values which are then used to weight the color samples interpolation (see Fig.3. Similarly, the white image is used to guide the interpolation performed in the alignment of the micro-len arrays on the photosensor. The method significantly decreases the crosstalk artefacts from which suffer existing methods.

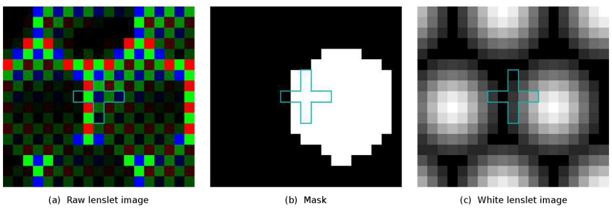

Figure 3. (a) is the raw image we want to demosaic, (b) is a mask which holds every pixel belonging to the same lenslet, (c) white image.

7.1.5. Super-rays for efficient Light fields processing

Participants: Matthieu Hog, Christine Guillemot.

Light field acquisition devices allow capturing scenes with unmatched post-processing possibilities. However, the huge amount of high dimensional data poses challenging problems to light field processing in interactive time. In order to enable light field processing with a tractable complexity, we have addressed, in collaboration with Neus Sabater (technicolor) the problem of light field over-segmentation [15]. We have introduced the concept of super-ray, which is a grouping of rays within and across views (see Fig.4), as a key component of a light field processing pipeline. The proposed approach is simple, fast, accurate, easily parallelisable, and does not need a dense depth estimation. We have demonstrated experimentally the efficiency of the proposed approach on real and synthetic datasets, for sparsely and densely sampled light fields. As super-rays capture a coarse scene geometry information, we have also shown how they can be used for real time light field segmentation and correcting refocusing angular aliasing.

Figure 4. Super-rays for the sparsely sampled light field in the Tsukuba dataset.

7.2. Representation and compression of large volumes of visual data

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

7.2.1. Cloud-based image and video compression

Participants: Jean Begaint, Christine Guillemot.

The emergence of cloud applications and web services has led to an increasing use of online resources for storing and exchanging images and videos. Billions of images are already stored in the cloud, and hundreds of millions are uploaded every day. Redundancy between images stored in the cloud can be leveraged to efficiently compress images by exploiting inter-images correlations. We have developed a region-based prediction scheme to exploit correlation between images in the cloud. In order to compensate the deformations between correlated images, the reference image of the cloud is first segmented into multiple regions determined from matched local features and aggregated super-pixels. We then estimate a photometric and geometric deformation model between the matched regions in the reference frame and frame to be coded. Multiple references are then generated, by applying the estimated deformation models to the reference frame, and organized in a pseudo-sequence to be differentially encoded with classic video coding tools. Experimental results demonstrate that the proposed approach yields significant rate-distortion performance improvements compared to current coding solutions such as HEVC.

7.2.2. *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. The goal of the collaboration with Ericsson is to develop novel compression algorithms that allow 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. In 2016, a scalable video compression was developed 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. In 2017 instead, we developed a backward compatible compression algorithm of HDR images, where only the LDR data are sent [14]. The novelty of the approach relies on the optimization of an invertible mapping called Tone Mapping Operator (TMO) that maps efficiently the HDR data to the LDR data. Two optimizations have been carried out in a rate-distortion sense: in the first problem, the distortion of the HDR data is minimized under the constraint of minimum LDR datarate, while in the second problem, a new constraint is added in the optimization problem to insure that LDR data are closed to some “aesthetic” a priori. Taking into account the aesthetic of the scene in video compression is indeed novel, since video compression is traditionally optimized to deliver the smallest distortion with the input data at the minimum datarate. Moreover, we provided new statistical models for estimating the distortions and the rate and showed their accuracy to the real data. Finally, a novel axis is currently carried out to efficiently exploit the temporal redundancy in HDR videos.

7.2.3. *Sparse image representation and deep learning for compression*

Participants: Thierry Dumas, Christine Guillemot, Aline Roumy.

Deep learning is a novel research area that attempts to extract high level abstractions from data by using a graph with multiple layers. One could therefore expect that deep learning might allow efficient image compression based on these high level features. However, there are many issues that make the learning task difficult in the context of image compression. First, learning a transform is equivalent to learning an autoencoder, which is of its essence unsupervised and therefore more difficult that classical supervised learning, where deep learning has shown tremendous results. Second, the learning has to be performed under a rate-distortion criterion, and not only a distortion criterion, as is classically done in machine learning. Last but not least, deep learning, as classical machine learning, consists in two phases: (i) build a graph that can make a good representation of the data (i.e. find an architecture usually made with neural nets), and (ii) learn the parameters of this architecture from large-scale data. As a consequence, neural nets are well suited for a specific task (text or image recognition) and require one training per task. The difficulty to apply machine learning approach to image compression is that it is important to deal with a large variety of patches, and with also various compression rates. Different architectures have been proposed to design a single neural network that can work efficiently at any coding rate either by a Winner Take all approach [28] or an adaptation to the quantization noise during the training [40].

7.2.4. *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.

In 2016, we have developed a graph-based representation for complex camera configurations. In particular, a generalized Graph-Based Representation has beend eveloped which handles two views with complex translations and rotations between them. The proposed approach uses the epipolar segments to have a row-wise description of the geometry that is as simple as for rectified views. In 2017, the Graph-based Representation has been extended to build a rate-distortion optimized description of the geometry of multi-view images [22]. This work brings two major novelties. First the graph can now handle multiple views (more than 2) thanks to a recursive construction of the geometry across the views. Second, the number of edges describing the geometry information is carefully chosen with respect to a rate-distortion criterion evaluated on the reconstructed views.

An adaptation of the graph-based representations (GBR) has been proposed to describe color and geometry information of light fields (LF) in [38]. Graph connections describing scene geometry capture inter-view dependencies. They are used as the support of a weighted Graph Fourier Transform (wGFT) to encode disoccluded pixels. The quality of the LF reconstructed from the graph is enhanced by adding extra color information to the representation for a sub-set of sub-aperture images. Experiments show that the proposed scheme yields rate-distortion gains compared with HEVC based compression (directly compressing the LF as a video sequence by HEVC).

7.2.5. *Light fields compression using sparse reconstruction*

Participants: Fatma Hawary, Christine Guillemot.

Light field data exhibits large amount of information, which poses challenging problems in terms of storage capacity, hence the need for efficient compression schemes. In collaboration with Technicolor (Dominique Thoreau and Guillaume Boisson), we have developed a scalable coding method for the light field data based on the sparsity of light fields in the angular (view) domain. A selected set of the light field sub-aperture images is encoded as a video sequence in a base layer and transmitted to the decoder. The remaining light field views are then reconstructed from the decoded subset of views, by exploiting the light field sparsity in the angular continuous Fourier domain. The reconstructed light field is enhanced using a patch-based restoration method which further exploits the light field angular redundancy.

7.2.6. *Light fields dimensionality reduction and compression*

Participants: Elian Dib, Christine Guillemot, Xiaoran Jiang, Mikael Le Pendu.

We have investigated low rank approximation methods exploiting data geometry for dimensionality reduction of light fields. We have developed an approximation method in which homographies and the rank approximation model are jointly optimized [32]. The homographies are searched in order to align linearly correlated sub-aperture images in such a way that the batch of views can be approximated by a low rank model. The light field views are aligned using either one global homography or multiple homographies depending on how much the disparity across views varies from one depth plane to the other. The rank constraint is expressed as a product of two matrices, where one matrix contains basis vectors and where the other one contains weighting coefficients. The basis vectors and weighting coefficients can be compressed separately exploiting their respective characteristics. The optimization hence proceeds by iteratively searching for the homographies and the factored model of the input set of sub-aperture images (views), which will minimize the approximation error.

A light field compression algorithm based on a low rank approximation exploiting scene and data geometry has then be developed [18]. The best pair of key parameters (approximation rank and quantization step size), in terms of rate-distortion performance, of the algorithm are predicted based on a model learned from a set of training light fields. The model is learned as a function of several input light field features: disparity indicators defined as a function of the decay rate of the SVD values of the original and registered view matrices, as well as texture indicators defined in terms of the decay rate of SVD values computed on the central view. The parameter prediction problem is cast as a multi-output classification problem solved using a Decision Tree ensemble method, namely the Random Forest method. The approximation method is currently being extended to local super-ray based low rank models.

7.3. Rendering, inpainting and super-resolution

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

7.3.1. Transformation of the Beta distribution for color transfer

Participants: Hristina Hristova, Olivier Le Meur.

After having investigated the use of multivariate generalized Gaussian distribution in color transfer, we propose a novel transformation between two Beta distributions. The key point is that performing a Gaussian-based transformation between bounded distributions may result in out-of-range values. Furthermore, as a symmetrical distribution, the Gaussian distribution cannot model asymmetric distributions. This reveals important limitations of the Gaussian model when applied to image processing tasks and, in particular, to color transfer. To tackle these limitations of the Gaussian-based transformations, we investigate the use of bounded distributions, and more specifically, the Beta distribution. The Beta distribution is a bounded two-parameter dependent distribution, which can admit different shapes and thus, fit various data, bounded in a discrete interval. Adopting the Beta distribution to model color and light distributions of images is our key idea and motivation. The proposed transformation progressively and accurately reshapes an input Beta distribution into a target Beta distribution using four intermediate statistical transformations. Experiments have shown that the proposed method obtains more natural and less saturated results than results of recent state-of-the-art color transfer methods. Moreover, the results portray better both the target color palette and the target contrast.

7.3.2. Light field inpainting and edit propagation

Participants: Oriel Frigo, Christine Guillemot, Mikael Le Pendu.

With the increasing popularity of computational photography brought by light field, simple and intuitive editing of light field images is becoming a feature of high interest for users. Light field editing can be combined with the traditional refocusing feature, allowing a user to include or remove objects from the scene, change its color, its contrast or other features.

A simple approach for editing a light field image can be obtained with an edit propagation, where first a particular subaperture view is edited (most likely the center one) and then a coherent propagation of this edit is performed through the other views. This problem is particularly challenging for the task of inpainting, as the disparity field is unknown under the occludding mask.

We have developed two methods which exploit two different light field priors, namely a low rank prior and a smoothness prior in epipolar plane images (EPI) to propagate a central view inpainting or edit to all the other views. In the first method, a set of warped versions of the inpainted central view with random homographies are vectorized and concatenated columnwise into a matrix together with the views of the light field to be inpainted. Because of the redundancy between the views, the matrix satisfies a low rank assumption enabling us to fill the region to inpaint with low rank matrix completion. To this end, a new matrix completion algorithm, better suited to the inpainting application than existing methods, has also been developed. In its simple form, our method does not require any depth prior, unlike most existing light field inpainting algorithms. The method has then been extended to better handle the case where the area to inpaint contains depth discontinuities.

In the second approach, the problem of propagating an edit from a single view to the remaining light field is solved by a structure tensor driven diffusion on the epipolar plane images [29]. Since EPIs are piecewise smooth and have no complex texture content, tensor driven diffusion is naturally suited for inpainting the EPIs as an efficient technique to obtain a coherent edit propagation. The proposed method has been shown to be useful for two applications: light field inpainting and recolorization. While the light field recolorization is obtained with a straightforward diffusion, the inpainting application is particularly challenging, as the structure tensors accounting for disparities are unknown under the occluding mask. This issue has been addressed with a disparity inpainting by means of an interpolation constrained by superpixel boundaries.

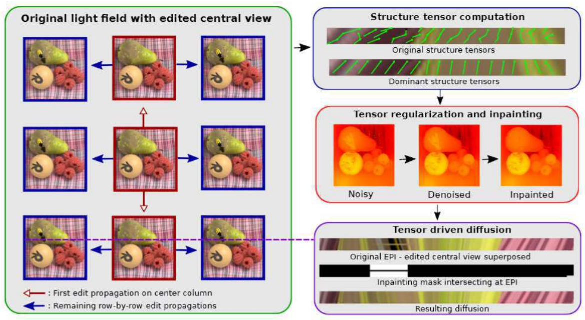

Figure 5. Overview of the proposed method. On the left, a light field with an inpainted central view and after a first edit propagation, which is performed through the center column of views (red arrows). Remaining edit propagations are performed row-by-row (blue arrows). The remaining boxes illustrate the steps of epipolar plane diffusion: Structure tensor computation where we obtain dominant structure tensors which are then spatially regularized and inpainted to estimate the structre tensors in the unknown part of the light field. Finally a tensor driven diffusion is performed on the EPIs.

7.3.3. Light fields super-resolution

Participants: Christine Guillemot, Lara Younes.

Capturing high spatial resolution light fields remains technologically challenging, and the images rendered from real light fields have today a significantly lower spatial resolution compared to traditional 2D cameras. In collaboration with the University of Malta (Prof. Reuben Farrugia), we have developed an example-based super-resolution algorithm for light fields, which allows the increase of the spatial resolution of the different views in a consistent manner across all sub-aperture images of the light field [12]. To maintain consistency across all sub-aperture images of the light field, the algorithm operates on 3D stacks (called patch-volumes) of 2D-patches, extracted from the different sub-aperture images. The patches forming the 3D stack are best matches across subaperture images. A dictionary of examples is first constructed by extracting, from a training set of high- and low- resolution light fields, pairs of high- and low-resolution patch-volumes. These patch-volumes are of very high dimension. Nevertheless, they contain a lot of redundant information, hence actually lie on subspaces of lower dimension. The low- and high-resolution patch-volumes of each pair can therefore be projected on their respective low and high-resolution subspaces using e.g. Principal Component Analysis (PCA). The dictionary of pairs of projected patch-volumes (the examples) map locally the relation between the high-resolution patch volumes and their low-resolution (LR) counterparts. A linear mapping function is then learned, using Multivariate Ridge Regression (RR), between the subspaces of the low- and high- resolution patch-volumes. Each overlapping patch-volume of the low-resolution light field can then be super-resolved by a straight application of the learned mapping function (some results in Fig.6).

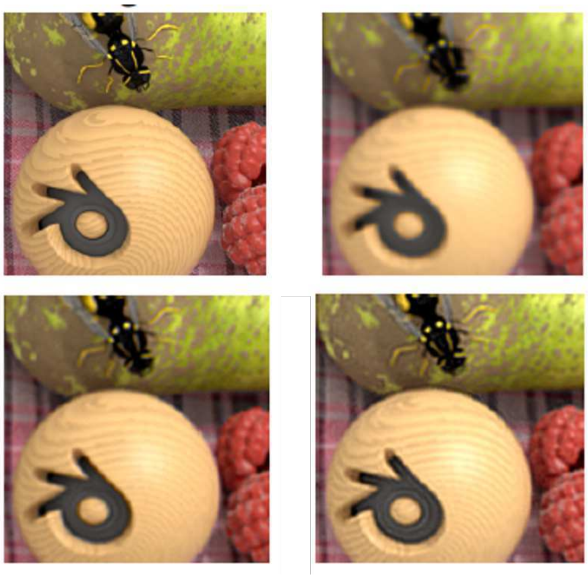

Figure 6. Illustration of some results on a crop of a central view with a magnification factor of 2; Top row: original (left), bicubic interpolation (right); Bottom row: super-resolved with a deep learning technique of the litterature (left), super-resolved with the proposed method (right).

This work is currently being extended on one hand by exploring how deep learning techniques can further benefit the scheme, and on the other hand by considering a hybrid system in which a 2D high resolution image, a priori not aligned with the light field views, can guide the light field super-resolution process.

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. Interactive Coding for Navigation in 3D scenes (ICON 3D)

Participants: Thomas Maugey, Aline Roumy.

In order to have performing FTV systems, the data transmission has to take into account the interactivity of the user, *i.e.,* the viewpoint that is requested. In other words, a FTV system transmits to the visualisation support only what needs to be updated when a user changes its viewpoint angle (*i.e.*, the new information appearing in its vision field).

In the context of the project ICON 3D funded by the GdR-Isis, we have developed new geometry prediction algorithms for surface meshes. Given a part of a mesh, the prediction algorithm is able to estimate a neighboring mesh subset corresponding to the one newly visible after user viewpoint angle change. For each mesh of a 3D model, we have generated all the predictions possible depending on the part of the model known by the decoder. Then we have characterized the prediction error.

The question of which data representation to use for Interactive Navigation has also been studied in [20]. More precisely, the navigation domain is split in small segments, each of them coded independently. This work has developed some optimal partitioning solution for different navigation scenario.

7.4.2. Correlation model selection for interactive video communication

Participants: Navid Mahmoudian Bidgoli, Thomas Maugey, Aline Roumy.

Interactive video communication has been recently proposed for multi-view videos. In this scheme, the server has to store the views as compactly as possible while allowing interactive navigation. Interactive navigation refers to the possibility for the user to select one view or a subset of views. To achieve this goal, the compression must be done using a model-based coding in which the correlation between the predicted view generated on the user side and the original view has to be modeled by a statistical distribution. In the context of the project Intercomm, the work published in [37] has proposed a framework for lossless fixed-length source coding to select a model among a candidate set of models that incurs the lowest extra rate cost to the system. Moreover, in cases where the depth image is available, we provide a method to estimate the correlation model.

7.4.3. Optimal selection of reference sensors for spatially correlated data storage

Participants: Thomas Maugey, Aline Roumy.

Highly instrumented Smart-cities, which are now common urban policies, are facing problems of management and storage of a large volume of data coming from an increasing number of sources. In the context of the project Intercom, we have proposed a data compression method by predictive coding of spatially correlated multi-source data. In a nutshell, some sensors are selected as references. They are used to predict the other sensor values, based on a Kriging prediction. We have proposed an algorithm to optimally select both the number and the position of the reference sensors among all the ones that are stored on a server and shared with a high number of users. This work has been done in collaboration with the Inria I4S project-team, IFFSTAR and the L2S.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

8.1.1. CIFRE contract with Envivio/ Ericsson on LDR compatible HDR video coding

Participants: Christine Guillemot, David Gommelet, Aline Roumy.

- Title : LDR-compatible coding of HDR video signals.
- Partners : Envivio.
- Funding : Cifre Envivio/Ericsson.
- 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.

8.1.2. CIFRE contract with Harmonic on image analysis for HDR video compression

Participants: Maxime Rousselot, Olivier Le Meur.

- Title : image and video analysis for HDR video compression
- Partners : Harmonic, Univ. Rennes 1
- Funding: Harmonic, ANRT
- Period: April 2016-April 2019

This project (in collaboration with Rémi Cozot, FRVSense) aims to investigate two main axes. First, we want to assess whether the representation of High Dynamic Range signal has an impact on the coding efficiency. We will focus mainly on the Hybrid Log-Gamma (HLG) and Perceptual Quantizer (PQ) OETF (Opto-Electronic Transfer Function)approaches. The former defines a nonlinear transfer function which is display-independent and able to produce high quality images without compromising the director's artistic intent. The latter approach is based on Just Noticeable Difference curve. If it turns out that this representation has an impact, the coding strategy should be adjusted with respect to the representation. In addition, specific preprocessing tools will be defined to deal with the limitations of PQ and HLG approaches.

8.1.3. CIFRE contract with Technicolor on image collection analysis

Participants: Dmitry Kuzovkin, Olivier Le Meur.

- Title : Spatiotemporal retargeting and recomposition based on artistic rules
- Partners : Technicolor, Univ. Rennes 1
- Funding: Technicolor, ANRT
- Period: Nov. 2015 – Nov. 2018

The goal of the project (in collaboration with Rémi Cozot, FRVSense) is to take advantage of the huge quantities of image and video data currently available - captured by both amateur and professional users - as well as the multiple copies of each scene that users often capture, to improve the aesthetic appeal of content. Additionally, given Technicolor's unique position, we propose to take advantage of insights as well as content from professional artists and colorists to learn how different content types can be enhanced.

8.1.4. CIFRE contract with Technicolor on light fields editing

Participants: Christine Guillemot, Matthieu Hog.

- Title : Light fields editing
- Research axis : 7.1.5
- 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 is to develop methods for light-field editing, and in 2017 we have introduced the the concept of super-ray which is a grouping of rays within and across views, and developed a fast algorithm for super-ray construction(see section 7.1.5).

8.1.5. CIFRE contract with Technicolor on light fields compressed representation

Participants: Christine Guillemot, Fatma Hawary.

- Title : Light fields compressed representation
- Research axis : 7.2.5
- Partners : Technicolor, Inria-Rennes.
- Funding : Technicolor, ANRT.
- Period : Feb.2016-Jan.2019.

The goal of this PhD is to study reconstruction algorithms from compressed measurements based on the assumption of sparsity in the Fourier domain. The goal is to apply these algorithms to scalable compression of light fields.

8.1.6. CIFRE contract with Technicolor on cloud-based image compression

Participants: Jean Begaint, Christine Guillemot.

- Title : Cloud-based image compression
- Research axis : 7.2.1
- 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

9.1.1. CominLabs/InterCom project

Participants: Aline Roumy, Thomas Maugey.

- Title : Interactive Communication (INTERCOM): Massive random access to subsets of compressed correlated data .
- Research axis : 7.4.2
- Partners : Inria-Rennes (Sirocco team and i4S team); LabSTICC, Telecom Bretagne, Signal & Communications Department; External partner: Kieffer L2S, CentraleSupelec, Univ. Paris Sud.
- Funding : Labex CominLabs.
- Period : Oct. 2016 - Nov. 2019.

This project aims to develop novel compression techniques allowing massive random access to large databases. Indeed, we consider a database that is so large that, to be stored on a single server, the data have to be compressed efficiently, meaning that the redundancy/correlation between the data have to be exploited. The dataset is then stored on a server and made available to users that may want to access only a subset of the data. Such a request for a subset of the data is indeed random, since the choice of the subset is user-dependent. Finally, massive requests are made, meaning that, upon request, the server can only perform low complexity operations (such as bit extraction but no decompression/compression). Algorithms for two emerging applications of this problem will be developed: Free-viewpoint Television (FTV) and massive requests to a database collecting data from a large-scale sensor network (such as Smart Cities).

9.2. European Initiatives

9.2.1. FP7 & H2020 Projects

9.2.1.1. ERC-CLIM

Participants: Pierre David, Elian Dib, Simon Evain, Christine Guillemot, Laurent Guillo, Mikael Le Pendu, Xiaoran Jiang, Jinglei Shi, Xin Su, Lara Younes.

Light fields yield a rich description of the scene ideally suited for advanced image creation capabilities from a single capture, such as simulating a capture with a different focus and a different depth of field, simulating lenses with different apertures, for creating images with different artistic intents or for producing 3D views. Light fields technology holds great promises for a number of application sectors, such as photography, augmented reality, light field microscopy, but also surveillance, to name only a few.

The goal of the ERC-CLIM project is to develop algorithms for the entire static and video light fields processing chain, going from compact sparse and low rank representations and compression to restoration, high quality rendering and editing.

9.3. International Initiatives

9.3.1. Inria Associate Teams Not Involved in an Inria International Labs

Title: Graph-based Omnidirectional video Processing

International Partner (Institution - Laboratory - Researcher):

Ecole Polytechnique Fédérale de Lausanne (Switzerland) - LTS4 - Pascal Frossard

Start year: 2017

See also: http://people.rennes.inria.fr/Thomas.Maugey/wp/projects/gop/

Due to new camera types, the format of the video data has become more complex than simple 2D images or videos as it was the case a few years ago. In particular, the omnidirectional cameras provide pixels on a whole sphere around a center point and enable a vision in 360°. In addition to the fact that the data size explodes with such cameras, the inherent structure of the acquired signal fundamentally differs from the 2D images, which makes the traditional video codec obsolete. In parallel of that, an important effort of research has been lead recently, especially at EPFL, to develop new processing tools for signals lying on irregular structures (graphs). It enables in particular to build efficient coding tools for new types of signals. The proposed research project will actually study how graphs can be built for defining a suitable structure on one or several 360° videos and then used for compression.

The collaboration between SIROCCO (Inria) and LTS4 (EPFL) has been very active in the recent years. However, only one-to-one collaboration was involved. When opening these new ambitious research direction, the project GOP will involve more than two or three researchers, and build a bidirectional collaboration between different people of the SIROCCO and LTS4 teams.

9.3.2. Inria International Partners

9.3.2.1. Informal International Partners

We have international collaborations with:

- Reuben Farrugia, Prof. at the University of Malta, with whom we continue collaborating on light field super-resolution.The collaboration strated during the sabbatical year (Sept. 2015-Aug. 2016) he spent within the team.
- Ehsan Miandji and Prof.Jonas Unger from Linkoping Univ. with whom we collaborate on compressive sampling of light fields. Ehsan Miandji has spent 1.5 month (June- July 2017) within the team.
- Chiara Galdi and Jean Luc Dugelay, prof. at Eurecom, with whom we collaborate on the application of light fields to biometry. Chiara Galdi has spent one month in the team (April 2017).
- Ole Johanssen and Prof. Bastian Goldluecke, from Univ. of Konstanz, with whom we collaborate on scene flow estimation with deep learning. Ole Johanssen has spent one month (Nov. 20- Dec. 20, 2017) in the team.
- The study on guided image inpainting is carried out in collaboration with Prof. Pascal Frossard from EPFL (Ecole Polytechique Federal de Lausanne).

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. Member of the Conference Program Committees

- C. Guillemot has been a member of the VISAPP 2017 international program committee.
- C. Guillemot has been a member of the international program committee of the CVPR 2017 workshop on light fields for computer vision.
- O. Le Meur has been a member of technical program committees of international conferences: QoMEX 2017, Mutual Benefits of Cognitive and Computer Vision (MBCC) 2017 workshop in conjunction with ICCV 2017, IPTA 2017, ICMR 2017.
- A. Roumy has been a member of the technical program committee of the CVPR 2017 workshop on New Trends in Image Restoration and Enhancement.
- A. Roumy has been a member of the technical program committee of the (Groupement de Recherche en Traitement du Signal et des Images) GRETSI 2017 workshop.

10.1.2. Journal

10.1.2.1. Member of the Editorial Boards

- C. Guillemot is senior area editor of the IEEE Trans. on Image Processing.
- C. Guillemot is associate editor of the International Journal on Mathematical Imaging and Vision.
- O. Le Meur is member of the editorial board of the IET Image Processing Journal
- A. Roumy is associate editor of the Springer Annals of Telecommunications.

10.1.3. Invited Talks

- C. Guillemot gave a seminar at IST-Lisboa on Light field image processing (10th May 2017).
- C. Guillemot gave an invited talk in the context of the IEEE distinguished lecturer program, at UPM Madrid on Light field image processing (29th June 2017).
- C. Guillemot gave an invited talk in the context of the IEEE distinguished lecturer program University Autonoma de Barcelona on Light field image processing (30th June 2017).
- C. Guillemot gave an invited talk in the context of the IEEE distinguished lecturer program on "Light field image processing: analysis, representation, compression and editing", at the VISVA 2017 summer school, Singapore (8th July 2017).
- C. Guillemot gave a Keynote (“Towards dense multi-view imaging”) at the Hot3D workshop jointly held with IEEE-ICME, Hong Kong, 10 July 2017.
- C. Guilemot gave a seminar at MidSweden University in light field image processing (15th Nov. 2017).
- O. Le Meur gave an invited talk at the AFIG (Association Française d’informatique Graphique). The title of the presentation was: "un nouveau regard sur l’attention visuelle".
- O. Le Meur gave an invited talk at the NEUROSTIC workshop (GDR-ISI). The title of the presentation was: "Modèles de saillance visuelle et mouvements oculaires. Un nouveau départ?".
- O. Le Meur was a captsone speaker at EuroRV3 workshop. The title of the presentation was: "The computational modelling of visual attention : saliency model vs saccadic model"
- O. Le Meur gave an invited talk at ENSTA. The title of the presentation was: "Saccadic model, a promising framework for modelling overt visual attention".

- T. Maugey gave an invited talk at EPFL (Lausanne, Switzerland) on “Massive Random Access », May 2017.
- T. Maugey gave an invited talk at the scientific day of the L2TI (University Paris XIII), December 2017.
- A. Roumy gave an introductory talk on Interactive video compression at the Comité des projets, Rennes, May 2017.
- A. Roumy gave a talk on Challenges in data compression for the IoT, at the IoTsustain Workshop, UCL, London, July 2017.

10.1.4. Leadership within the Scientific Community

- C. Guillemot is member of the IEEE IVMSP technical committee
- C. Guillemot is senior member of the steering committee of IEEE Trans. on Multimedia (2016-2018).
- C. Guillemot has been appointed member of the IEEE Signal Processing Society Nominations and Appointments Committee for a two year term (2018-2019).
- A. Roumy is a member of the Executive board of the National Research group in Image and Signal Processing (GRETSI).

10.1.5. Scientific Expertise

- C. Guillemot served in the HCERES panel which evaluated the CEDRIC laboratory at CNAM (Oct. 2017).

10.1.6. Research Administration

- C. Guillemot is a member of Inria’s evaluation committee.
- C. Guillemot is member of the “bureau du Comité des Projets”.
- A. Roumy served as a member of Board of Examiners (Comité de sélection) for an Associate Professor position (Maitre de Conférences) at ENSSAT-IRISA, Lannion (61MCF1758).
- A. Roumy served as a member of Board of Examiners (Comité de sélection) for an Associate Professor position (Maitre de Conférences) in Digital Communication at CentraleSupelec, Rennes.
- A. Roumy served as a jury member for the selection of Inria CR (researcher) candidates, March and May 2017, Inria, Rennes, France
- A. Roumy is a member of the technological development and technology transfer committee at Inria Rennes
- A. Roumy was a member of Inria’s PhD grant committee (CORDI-S).

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

- Master: C. Guillemot, Image and video compression, 10 hours, M2 SISEA, Univ. of Rennes 1, France.
- Undergraduate: L. Guillo, "Scientific programming ", 20 hours, University Rennes 1, "Physique, chimie et structure de la matière (PCSTM), 1st year", France.
- Master: O. Le Meur, Selective visual attention, 6 hours, M2, Univ. of Paris 8, France.
- Master: O. Le Meur, Acquisition/Image Processing/Compression, 22 hours, M2 MITIC, Univ. of Rennes 1, 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.

- Professional training: O. Le Meur, Image Processing and OpenCV, 42 hours, Technicolor Rennes.
- Master: T. Maugey, course on 3D models in a module on "advanced video", 4 hours, M2 SISEA, Univ. of Rennes 1, France.
- Engineering degree: A. Roumy, Sparse methods in Image and Signal processing, 13 hours, INSA Rennes, 5th year, Mathematical engineering, France.
- Engineering degree: A. Roumy, Image processing, 28 hours, ECAM Rennes, 4th year, France.
- Master: A. Roumy, Foundations of Smart Sensing, 18 hours, ENSAI Bruz, Master of Science in Statistics for Smart Data, France.
- Master: A. Roumy, High dimensional statistical learning, 9 hours, University Rennes 1, Research in Computer Science (SIF) master, France.
- Master: A. Roumy, Information Theory, 15 hours, University Rennes 1, Research in Computer Science (SIF) master, France.

10.2.2. Juries

- C. Guillemot has been member (rapporteur) of the PhD committee of:
  - A. Purica, Telecom ParisTech, June 2017
  - G. Nieto, Univ. Grenoble-Alpes, Oct. 2017
- C. Guillemot has been member of the PhD jury of:
  - A. Aldahdooh, Univ. Nantes, Aug. 2017
  - C. Peyrard, Univ. Lyon, Sept. 2017
  - S. Puri, Univ. Nantes, Nov. 2017
  - B. Pasdeloup, IMT-Atlantique, Dec. 2017
- O. Le Meur has been member (rapporteur) of the jury of the PhD committee of:
  - Benoit Arbelot, Univ. Grenoble Alpes, Mar. 2017
  - Celine Craye, Univ. Paris-Saclay, Apr. 2017
  - Yashas Rai, Univ. of Nantes, Aug. 2017
- Thomas Maugey has been member of the jury of the PhD committee of:
  - A. Dricot TelecomParisTech, Mar. 2017
- A. Roumy has been member of the PhD jury of:
  - V. Gouldieff, CentraleSupelec, Rennes, October 2017.

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] G. COLUCCIA, A. ROUMY, E. MAGLI. *Operational Rate-Distortion Performance of Single-source and Distributed Compressed Sensing*, in "Communications, IEEE Transactions on", June 2014, vol. 62, n^{o} 6, pp. 2022 - 2033, IEEE Transactions on Communications [*DOI :* 10.1109/TCOMM.2014.2316176], https://hal.inria.fr/hal-00996698

[4] 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

[5] C. GUILLEMOT, A. ROUMY. *Toward constructive Slepian-Wolf coding schemes*, in "Distributed Source Coding: Theory, Algorithms, and Applications", P. L. DRAGOTTI, M. GASTPAR (editors), Elsevier, January 2009, https://hal.inria.fr/hal-01677745

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

[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] 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

Articles in International Peer-Reviewed Journals

[10] M. ALAIN, C. GUILLEMOT, D. THOREAU, P. GUILLOTEL. *Scalable Image Coding based on Epitomes*, in "IEEE Transactions on Image Processing", May 2017, vol. 26, n^{o} 8, pp. 3624-3635 [*DOI :* 10.1109/TIP.2017.2702396], https://hal.inria.fr/hal-01591504

[11] P. BUYSSENS, O. LE MEUR, M. DAISY, D. TSCHUMPERLÉ, O. LÉZORAY. *Depth-guided disocclusion inpainting of synthesized RGB-D images*, in "IEEE Transactions on Image Processing", 2017, vol. 26, n^{o} 2, pp. 525-538 [*DOI :* 10.1109/TIP.2016.2619263], https://hal.archives-ouvertes.fr/hal-01391065

[12] R. A. FARRUGIA, C. GALEA, C. GUILLEMOT. *Super Resolution of Light Field Images using Linear Subspace Projection of Patch-Volumes*, in "IEEE Journal of Selected Topics in Signal Processing", October 2017, pp. 1-14 [*DOI :* 10.1109/JSTSP.2017.2747127], https://hal.archives-ouvertes.fr/hal-01591488

[13] R. A. FARRUGIA, C. GUILLEMOT. *Face Hallucination Using Linear Models of Coupled Sparse Support*, in "IEEE Transactions on Image Processing", 2017, vol. 26, n^{o} 9, pp. 4562-4577 [*DOI :* 10.1109/TIP.2017.2717181], https://hal.archives-ouvertes.fr/hal-01591517

[14] D. GOMMELET, A. ROUMY, C. GUILLEMOT, M. ROPERT, J. LE TANOU. *Gradient-Based Tone Mapping for Rate-Distortion Optimized Backward-Compatible High Dynamic Range Compression*, in "IEEE Transactions on Image Processing", August 2017 [*DOI :* 10.1109/TIP.2017.2740159], https://hal.inria.fr/hal-01590258

[15] M. HOG, N. SABATER, C. GUILLEMOT. *Super-rays for Efficient Light Field Processing*, in "IEEE Journal of Selected Topics in Signal Processing", 2017 [*DOI :* 10.1109/JSTSP.2017.2738619], https://hal.archives-ouvertes.fr/hal-01407852

[16] H. HRISTOVA, O. LE MEUR, R. COZOT, K. BOUATOUCH. *High-dynamic-range image recovery from flash and non-flash image pairs*, in "Visual Computer", June 2017, vol. 33, n^{o} 6-8, pp. 725 - 735 [*DOI :* 10.1007/s00371-017-1399-0], https://hal.inria.fr/hal-01650464

[17] H. HRISTOVA, O. LE MEUR, R. COZOT, K. BOUATOUCH. *Transformation of the Multivariate Generalized Gaussian Distribution for Image Editing*, in "IEEE Transactions on Visualization and Computer Graphics", November 2017, 15 p. [*DOI :* 10.1109/TVCG.2017.2769050], https://hal.inria.fr/hal-01650368

[18] X. JIANG, M. LE PENDU, R. A. FARRUGIA, C. GUILLEMOT. *Light Field Compression with Homography-based Low Rank Approximation*, in "IEEE Journal of Selected Topics in Signal Processing", 2017 [*DOI :* 10.1109/JSTSP.2017.2747078], https://hal.archives-ouvertes.fr/hal-01591349

[19] O. LE MEUR, A. COUTROT, Z. LIU, P. RÄMÄ, A. LE ROCH, A. HELO. *Visual Attention Saccadic Models Learn to Emulate Gaze Patterns From Childhood to Adulthood*, in "IEEE Transactions on Image Processing", October 2017, vol. 26, n^{o} 10, pp. 4777 - 4789 [*DOI :* 10.1109/TIP.2017.2722238], https://hal.inria.fr/hal-01650322

[20] R. MA, T. MAUGEY, P. FROSSARD. *Optimized Data Representation for Interactive Multiview Navigation*, in "IEEE Transactions on Multimedia", 2017, pp. 1-15, forthcoming, https://hal.inria.fr/hal-01633660

[21] H. SONG, Z. LIU, H. DU, G. SUN, O. LE MEUR, T. REN. *Depth-Aware Salient Object Detection and Segmentation via Multiscale Discriminative Saliency Fusion and Bootstrap Learning*, in "IEEE Transactions on Image Processing", September 2017, vol. 26, n^{o} 9, pp. 4204 - 4216 [*DOI :* 10.1109/TIP.2017.2711277], https://hal.inria.fr/hal-01650409

[22] X. SU, T. MAUGEY, C. GUILLEMOT. *Rate-Distortion Optimized Graph-Based Representation for Multiview Images with Complex Camera Configurations*, in "IEEE Transactions on Image Processing", 2017 [*DOI :* 10.1109/TIP.2017.2685340], https://hal.archives-ouvertes.fr/hal-01492850

[23] E. VIDAL, N. STURMEL, C. GUILLEMOT, P. CORLAY, F.-X. COUDOUX. *New adaptive filters as perceptual preprocessing for rate-quality performance optimization of video coding*, in "Signal Processing: Image Communication", 2017, vol. 52, pp. 124 - 137 [*DOI :* 10.1016/J.IMAGE.2016.12.003], https://hal.archives-ouvertes.fr/hal-01591984

[24] C. VERLEYSEN, T. MAUGEY, C. DE VLEESCHOUWER, P. FROSSARD. *Wide baseline image-based rendering based on shape prior regularisation*, in "IEEE Transactions on Image Processing", 2017, https://hal.inria.fr/hal-01575559

Invited Conferences

[25] T. MAUGEY, O. LE MEUR, Z. LIU. *Saliency-based navigation in omnidirectional image*, in "MMSP 2017 - IEEE International Workshop on Multimedia Signal Processing", Luton, United Kingdom, October 2017, https://hal.inria.fr/hal-01575566

International Conferences with Proceedings

[26] G. COLUCCIA, A. ROUMY, E. MAGLI. *Mismatched sparse denoiser requires overestimating the support length*, in "IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)", New Orleans, United States, March 2017 [*DOI :* 10.1109/ICASSP.2017.7953058], https://hal.inria.fr/hal-01589633

[27] *Best Paper*
P. DAVID, M. LE PENDU, C. GUILLEMOT. *White Lenslet Image Guided Demosaicing for Plenoptic Cameras*, in "MMSP 2017 - IEEE 19th International Workshop on Multimedia Signal Processing", Luton, United Kingdom, Multimedia Signal Processing (MMSP), 2017 IEEE 19th International Workshop on, October 2017, https://hal.archives-ouvertes.fr/hal-01590345.

[28] T. DUMAS, A. ROUMY, C. GUILLEMOT. *Image Compression with Stochastic Winner-Take-All Auto-Encoder*, in "2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2017)", New Orleans, United States, March 2017, https://hal.archives-ouvertes.fr/hal-01493137

[29] O. FRIGO, C. GUILLEMOT. *Epipolar Plane Diffusion: An Efficient Approach for Light Field Editing*, in "British Machine Vision Conference (BMVC)", London, France, September 2017, https://hal.archives-ouvertes.fr/hal-01591525

[30] F. HAWARY, C. GUILLEMOT, D. THOREAU, G. BOISSON. *Scalable Light Field Compression Scheme Using Sparse Reconstruction and Restoration*, in "ICIP 2017 - IEEE International Conference on Image Processing", Beijing, China, September 2017, https://hal.archives-ouvertes.fr/hal-01597536

[31] H. HRISTOVA, O. LE MEUR, R. COZOT, K. BOUATOUCH. *Perceptual metric for color transfer methods*, in "IEEE International Conference on Image Processing - ICIP 2017", Shanghai, China, September 2017, https://hal.inria.fr/hal-01651141

[32] X. JIANG, M. LE PENDU, R. A. FARRUGIA, S. S. HEMAMI, C. GUILLEMOT. *Homography-Based Low Rank Approximation of Light Fields for Compression*, in "IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)", New Orleans, United States, March 2017 [*DOI :* 10.1109/ICASSP.2017.7952369], https://hal.archives-ouvertes.fr/hal-01591315

[33] X. JIANG, M. LE PENDU, C. GUILLEMOT. *Light field compression using depth image based view synthesis*, in "IEEE International Conference on Multimedia & Expo Workshops (ICMEW)", Hong Kong, China, July 2017 [*DOI :* 10.1109/ICMEW.2017.8026313], https://hal.archives-ouvertes.fr/hal-01591329

[34] D. KUZOVKIN, T. POULI, R. COZOT, O. LE MEUR, J. KERVEC, K. BOUATOUCH. *Context-aware clustering and assessment of photo collections*, in "Proceedings of the symposium on Computational Aesthetics", Los Angeles, United States, July 2017 [*DOI :* 10.1145/3092912.3092916], https://hal.inria.fr/hal-01651171

[35] O. LE MEUR, A. COUTROT, A. LE ROCH, A. HELO, P. RÄMÄ, Z. LIU. *Age-dependent saccadic models for predicting eye movements*, in "IEEE International Conference on Image Processing - ICIP 2017", Shanghai, China, September 2017, https://hal.inria.fr/hal-01651151

[36] O. LE MEUR, A. COUTROT, Z. LIU, P. RÄMÄ, A. LE ROCH, A. HELO. *Your gaze betrays your age*, in "25th European Signal Processing Conference - EUSIPCO", Kos, Greece, August 2017, https://hal.inria.fr/hal-01651125

[37] N. MAHMOUDIAN BIDGOLI, T. MAUGEY, A. ROUMY. *Correlation model selection for interactive video communication*, in "ICIP 2017 - IEEE International Conference on Image Processing", Beijing, China, September 2017, https://hal.inria.fr/hal-01590627

[38] X. SU, M. RIZKALLAH, T. MAUGEY, C. GUILLEMOT. *Graph-based light fields representation and coding using geometry information*, in "IEEE International Conference on Image Processing (ICIP)", Beijing, China, IEEE , September 2017, https://hal.inria.fr/hal-01589626

National Conferences with Proceedings

[39] A. CRINIÈRE, A. ROUMY, T. MAUGEY, M. KIEFFER, J. DUMOULIN. *Sélection optimale de capteurs de référence pour le stockage de données spatialement corrélées*, in "GRETSI 2017 - XXVIème Colloque", Juan-les-Pins, France, GRETSI 2017, September 2017, https://hal.inria.fr/hal-01575610

[40] T. DUMAS, A. ROUMY, C. GUILLEMOT. *Auto-encodeur optimisé au sens débit-distorsion : indépendant de la quantification?*, in "GRETSI 2017", Juan-les-Pins, France, September 2017, https://hal.archives-ouvertes.fr/hal-01579257

[41] T. MAUGEY, C. LE CAM, L. GUILLO. *Télévision à point de vue libre et système de capture à plusieurs caméra omnidirectionnelles*, in "GRETSI 2017", Juan-les-Pins, France, September 2017, https://hal.inria.fr/hal-01575606

Other Publications

[42] J. BÉGAINT, D. THOREAU, P. GUILLOTEL, C. GUILLEMOT. *Region-based prediction for image compression in the cloud*, December 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01662639

[43] M. HOG, N. SABATER, C. GUILLEMOT. *Dynamic Super-Rays for Efficient Light Field Video Processing*, November 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01649342