UMR IRISA

Activity Report 2019

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:

Other Research Topics and Application Domains:

1. Team, Visitors, External Collaborators

Research Scientists

Christine Guillemot [Team leader, Inria, Senior Researcher, HDR]
Xiaoran Jiang [Inria, Starting Research Position]
Maja Krivokuca [Inria, Starting Research Position, from Mar 2019]
Thomas Maugey [Inria, Researcher]
Aline Roumy [Inria, Researcher, HDR]

Post-Doctoral Fellows

Christian Galea [Inria, until Jan 2019]
Ehsan Miandji [Inria, from Nov 2019]
Mai Quyen Pham [Inria, until Mar 2019]

PhD Students

Pierre David [Inria]
Elian Dib [Inria]
Simon Evain [Inria]
Patrick Garus [Orange Labs]
Fatma Hawary [Technicolor, until Jan 2019]
Navid Mahmoudian Bidgoli [Inria, until Sep 2019]
Xuan Hien Pham [InterDigital]
Mira Rizkallah [Univ de Rennes I, until Jan 2019]
Jinglei Shi [Inria]

Technical staff

Pierre Allain [Inria, Engineer]
Laurent Guillo [CNRS]
Fatma Hawary [Inria, Engineer, from Mar 2019]
Navid Mahmoudian Bidgoli [Inria, Engineer, from Oct 2019]
Hoai Nam Nguyen [Inria, Engineer]
Mira Rizkallah [Inria, Engineer]
Alexander Sagel [Inria, Engineer, from Feb 2019]

Visiting Scientist

Zhaolin Xiao [Xian University, Prof., Dec. 2018 - Nov. 2019]

2. Overall Objectives

2.1. Introduction

Efficient processing, i.e., analysis, storage, access and transmission of visual content, with continuously increasing data rates, in environments which are more and more mobile and distributed, remains a key challenge of the signal and image processing community. New imaging modalities (HDR, multiview, plenoptic, light fields, 360$^{\text{o}}$ videos) generating very large volumes of data contribute to the sustained need for efficient algorithms for a variety of processing tasks.

Building upon a strong background on signal/image/video processing and information theory, the goal of the SIROCCO team is to design mathematically founded tools and algorithms for visual data analysis, modeling, representation, coding, and processing, with for the latter area an emphasis on inverse problems related to super-resolution, view synthesis, HDR recovery from multiple exposures, denoising and inpainting. Even if 2D imaging is still within our scope, the goal is to give a particular attention to HDR imaging, light fields, and 360° videos. The project-team activities are structured and organized around the following inter-dependent research axes:

- Visual data analysis
- Signal processing and learning methods for visual data representation and compression
- Algorithms for inverse problems in visual data processing
- Distributed coding for interactive communication.

While aiming at generic approaches, some of the solutions developed are applied to practical problems in partnership with industry (Technicolor, Ericsson, Orange) 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 network adaptation, of advanced functionalities such as navigation, interactive streaming and high quality rendering.

2.2. Visual Data Analysis

Most visual data processing problems require a prior step of data analysis, of discovery and modeling of correlation structures. This is a pre-requisite for the design of dimensionality reduction methods, of compact representations and of fast processing techniques. These correlation structures often depend on the scene and on the acquisition system. Scene analysis and modeling from the data at hand is hence also part of our activities. To give examples, scene depth and scene flow estimation is a cornerstone of many approaches in multi-view and light field processing. The information on scene geometry helps constructing representations of reduced dimension for efficient (e.g. in interactive time) processing of new imaging modalities (e.g. light fields or 360° videos).

2.3. Signal processing and learning methods for visual data representation and compression

Dimensionality reduction has been at the core of signal and image processing methods, for a number of years now, hence have obviously always been central to the research of Sirocco. These methods encompass sparse and low rank models, random low-dimensional projections in a compressive sensing framework, and graphs as a way of representing data dependencies and defining the support for learning and applying signal de-correlating transforms. The study of these models and signal processing tools is even more compelling for designing efficient algorithms for processing the large volumes of high-dimensionality data produced by novel imaging modalities. The models need to be adapted to the data at hand through learning of dictionaries or of neural networks. In order to define and learn local low-dimensional or sparse models, it is necessay to capture and understand the underlying data geometry, e.g. with the help of manifolds and manifold clustering tools. It also requires exploiting the scene geometry with the help of disparity or depth maps, or its variations in time via coarse or dense scene flows.

2.4. Algorithms for inverse problems in visual data processing

Based on the above models, besides compression, our goal is also to develop algorithms for solving a number of inverse problems in computer vision. Our emphasis is on methods to cope with limitations of sensors (e.g. enhancing spatial, angular or temporal resolution of captured data, or noise removal), to synthesize virtual views or to reconstruct (e.g. in a compressive sensing framework) light fields from a sparse set of input views, to recover HDR visual content from multiple exposures, and to enable content editing (we focus on color transfer, re-colorization, object removal and inpainting). Note that view synthesis is a key component of multiview and light field compression as well as to support user navigation and interactive streaming. It is also needed to avoid angular aliasing in some post-capture processing tasks, such as re-focusing, from a sparse light field. Learning models for the data at hand is key for solving the above problems.

2.5. Distributed coding for interactive communication

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 solutions suitable for band and energy limited networking environments. Our goal is to address theoretical issues such as the problem of modeling the correlation channel between sources, and to practical coding solutions for distributed processing and communication and for interactive streaming.

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 machine learning, of coding and information theory. 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. Data Dimensionality Reduction

Manifolds, graph-based transforms, compressive sensing

Dimensionality reduction encompasses a variety of methods for low-dimensional data embedding, such as sparse and low rank models, random low-dimensional projections in a compressive sensing framework, and sparsifying transforms including graph-based transforms. These methods are the cornerstones of many visual data processing tasks (compression, inverse problems).

Sparse representations, *compressive sensing*, and *dictionary learning* have been shown to be powerful tools for efficient processing of visual data. The objective of *sparse representations* is to find a sparse approximation of a given input data. In theory, given $A \in \mathbb{R}^{m \times n}$, and $\mathbf{b} \in \mathbb{R}^m$ with $m << n$ and A is of full row 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. A is known as the dictionary, its columns a_j are the atoms, they are assumed to be normalized in Euclidean norm. There exist many solutions x to $Ax = b$. The problem is to find the sparsest solution x, i.e. the one having 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 recent theory of *compressive sensing*, in the context of discrete signals, can be seen as an effective dimensionality reduction technique. The idea behind compressive sensing is that a signal can be accurately recovered from a small number of linear measurements, at a rate much smaller than what is commonly prescribed by the Shannon-Nyquist theorem, provided that it is sparse or compressible in a known basis. Compressed sensing has emerged as a powerful framework for signal acquisition and sensor design, with a number of open issues such as learning the basis in which the signal is sparse, with the help of dictionary learning methods, or the design and optimization of the sensing matrix. The problem is in particular investigated in the context of light fields acquisition, aiming at novel camera design with the goal of offering a good trade-off between spatial and angular resolution.

While most image and video processing methods have been developed for cartesian sampling grids, new imaging modalities (e.g. point clouds, light fields) call for representations on irregular supports that can be well represented by *graphs*. Reducing the dimensionality of such signals require designing novel transforms yielding compact signal representation. One example of transform is the Graph Fourier transform whose basis functions are given by the eigenvectors of the graph Laplacian matrix $\mathbf{L} = \mathbf{D} - \mathbf{A}$, where $\mathbf{D}$ is a diagonal degree matrix whose i^{th} diagonal element is equal to the sum of the weights of all edges incident to the node i. The eigenvectors of the Laplacian of the graph, also called laplacian eigenbases, are analogous to the Fourier bases in the Euclidean domain and allow representing the signal residing on the graph as a linear combination of eigenfunctions akin to Fourier Analysis. This transform is particular efficient for compacting smooth signals on the graph. The problems which therefore need to be addressed are (i) to define graph structures on which the corresponding signals are smooth for different imaging modalities and (ii) the design of transforms compacting well the signal energy with a tractable computational complexity.

3.3. Deep neural networks

Autoencoders, Neural Networks, Recurrent Neural Networks

From dictionary learning which we have investigated a lot in the past, our activity is now evolving towards deep learning techniques which we are considering for dimensionality reduction. We address the problem of unsupervised learning of transforms and prediction operators that would be optimal in terms of energy compaction, considering autoencoders and neural network architectures.

An autoencoder is a neural network with an encoder g_e, parametrized by θ, that computes a representation Y from the data X, and a decoder g_d, parametrized by ϕ, that gives a reconstruction $\widehat{X}$ of X (see Figure below). Autoencoders can be used for dimensionality reduction, compression, denoising. When it is used for compression, the representation need to be quantized, leading to a quantized representation $\widehat{Y} = Q(Y)$ (see Figure below). If an autoencoder has fully-connected layers, the architecture, and the number of parameters to be learned, depends on the image size. Hence one autoencoder has to be trained per image size, which poses problems in terms of genericity.
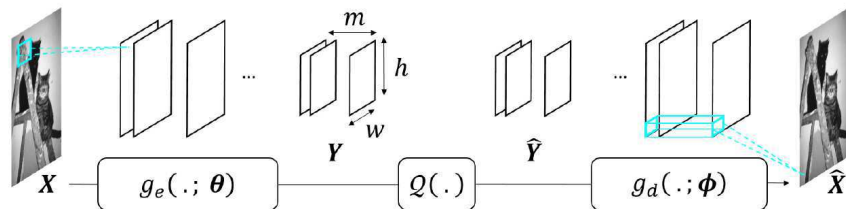
Figure 1. Illustration of an autoencoder.

To avoid this limitation, architectures without fully-connected layer and comprising instead convolutional layers and non-linear operators, forming convolutional neural networks (CNN) may be preferable. The obtained representation is thus a set of so-called feature maps.

The other problems that we address with the help of neural networks are scene geometry and scene flow estimation, view synthesis, prediction and interpolation with various imaging modalities. The problems are posed either as supervised or unsupervised learning tasks. Our scope of investigation includes autoencoders, convolutional networks, variational autoencoders and generative adversarial networks (GAN) but also recurrent networks and in particular Long Short Term Memory (LSTM) networks. Recurrent neural networks attempting to model time or sequence dependent behaviour, by feeding back the output of a neural network layer at time t to the input of the same network layer at time t+1, have been shown to be interesting tools for temporal

frame prediction. LSTMs are particular cases of recurrent networks made of cells composed of three types of neural layers called gates.

Deep neural networks have also been shown to be very promising for solving inverse problems (e.g. super-resolution, sparse recovery in a compressive sensing framework, inpainting) in image processing. Variational autoencoders, generative adversarial networks (GAN), learn, from a set of examples, the latent space or the manifold in which the images, that we search to recover, reside. The inverse problems can be re-formulated using a regularization in the latent space learned by the network. For the needs of the regularization, the learned latent space may need to verify certain properties such as preserving distances or neighborhood of the input space, or in terms of statistical modelling. GANs, trained to produce images that are plausible, are also useful tools for learning texture models, expressed via the filters of the network, that can be used for solving problems like inpainting or view synthesis.

3.4. 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 [1] 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 [2] (SW) and Wyner-Ziv [3] (WZ) theorems. Let us consider two binary correlated sources X and Y. If the two coders communicate, it is well known from Shannon's theory that the minimum lossless rate for X and Y is given by the joint entropy $H(X, Y)$. Slepian and Wolf have established in 1973 that this lossless compression rate bound can be approached with a vanishing error probability for long sequences, even if the two sources are coded separately, provided that they are decoded jointly and that their correlation is known to both the encoder and the decoder.

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

4. Application Domains

4.1. 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.

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

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

[3] 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.

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 imaging modalities, e.g. high dynamic range images and videos (higher bit depth, extended colorimetric space), light fields and omni-directional imaging.

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 (FVV). 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

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 and mobile access networks with different radio access technologies 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

- C. Guillemot has received the 2019 Eurasip Technical Achievement Award.

6. New Software and Platforms

6.1. M360CT

Multi-360 Calibration Toolkit

KEYWORDS: Omnidirectional camera - Calibration - FTV - 6DoF

FUNCTIONAL DESCRIPTION: Based on multiple synchronized sequences of a chessboard pattern moving in the scene, the algorithm computes the internal and external camera parameters of the different cameras under the unified spherical model. This software is composed of two executables, the first one for the individual calibration of each camera, the second one for the fusion of all the outputs of the first executable. The work has been submitted at APP with the number IDNN.FR.001.510008.S.P.2018.000.10800.

- Participants: Cédric Le Cam, Thomas Maugey and Laurent Guillo
- Contact: Thomas Maugey
- URL: http://project.inria.fr/ftv360

6.2. Context-adaptive neural network based prediction for image compression

KEYWORDS: Deep learning - Image compression - Intra prediction - Neural networks

FUNCTIONAL DESCRIPTION: This code implements (i) the learning of deep neural networks for intra-prediction of video compression and (ii) a video coder/decoder integrating the learned deep neural networks. This code allows to reproduce the results of the paper "Thierry Dumas, Aline Roumy and Christine Guillemot. Context-adaptive neural network based prediction for image compression, IEEE Transactions on Image Processing, 2019." To this end, the code implements the pre-processing of an image database (i.e. extraction of a set of pairs containing a block to be predicted and its context) to yield a training set. Then, from this learning set, the code implements the learning of one deep neural network per block size (full-connected network at size 8x8 and smaller, and convolutional network at bigger size). The code contains the parameters of all deep networks learned from the preprocessed ILSVRC2012 training database. Then, the code contains a modified version of the HEVC video compression test model (HM 16.9), which integrates the learned prediction functions and the signalling of the intra prediction mode (for both classical HEVC intra prediction modes and the learned neural networks).

- Contact: Aline Roumy
- URL: https://github.com/thierrydumas/context_adaptive_neural_network_based_prediction

6.3. LFDE-FLEX

LFDE-FLEX: A Framework for Learning Based Depth from a Flexible Subset of Dense and Sparse Light Field Views

KEYWORDS: Light fields - Depth estimation - Deep learning

FUNCTIONAL DESCRIPTION: The code implements a learning based depth estimation framework suitable for both densely and sparsely sampled light fields. The proposed framework consists of three processing steps: initial depth estimation, fusion with occlusion handling, and refinement. The estimation can be performed from a flexible subset of input views. The fusion of initial disparity estimates, relying on two warping error measures, allows us to have an accurate estimation in occluded regions and along the contours. In contrast with methods relying on the computation of cost volumes, the proposed approach does not need any prior information on the disparity range.

- Participants: Jinglei Shi, Xiaoran Jiang and Christine Guillemot
- Contact: Jinglei Shi

6.4. EPI-SENet

EPI-based light field view extrapolation network

KEYWORDS: Light fields - Deep learning - View synthesis

FUNCTIONAL DESCRIPTION: This code implements a learning based algorithm for light field view extrapolation. from axial volumes of sheared epipolar plane images (EPIs). The learned SENet network is based on tensorflow backend. The inputs of this network are multiple views in a row of a structured dense light field. The network predicts novel views in order to extend the light field baseline by a factor which can go up to 4 times the initial baseline. The code also performs digital refocusing with the original and extrapolated views. As with extended numerical aperture in classical imaging, the extrapolated light field gives refocused images with a shallower depth of field (DOF), leading to more accurate refocusing results.

- Participants: Zhaolin Xiao, Jinglei Shi, Xiaoran Jiang and Christine Guillemot
- Contact: Xiaoran Jiang

6.5. 4D-SFE

4D-SFE: 4D Scene Flow Estimator from Light Fields

KEYWORDS: Light fields - Scene Flow - Motion analysis - Depth estimation

FUNCTIONAL DESCRIPTION: This software implements a method for scene flow estimation from light fields by computing an optical flow, a disparity map and a disparity variation map for the whole light field. It takes as inputs two consecutive frames from a light field video, as well as optical flow and disparity maps estimated for each view of the light field (e.g. with a deep model like PWC-Net) and saved as .flo files. First the light field is divided into 4D clusters called superrays, then a neighboring weighted graph is built between the different clusters and finally a 4D affine model is fitted for every cluster, using the initial estimations from the optical flow and disparity estimations that are contained in the cluster and in the neighboring clusters.

- Participants: Pierre David and Christine Guillemot
- Contact: Christine Guillemot

6.6. LMVS-Net

LMVS-Net: Lightweight Neural Network for Monocular View Synthesis with Occlusion Handling

KEYWORDS: Light fields - View synthesis - Deep learning

FUNCTIONAL DESCRIPTION: This code implements the method described in A Lightweight Neural Network for Monocular View Synthesis with Occlusion Handling, allowing to perform monocular view synthesis in a stereo setting. From one input image, it computes a view laterally located, left-side or right-side, and with the required disparity range depending on user input. It is also able to retrieve a disparity map from the input image, as well as a confidence map to distinguish the occluded regions, as well as to evaluate the pixelwise accuracy of the prediction. The code was developed using Keras and TensorFlow.

- Participants: Simon Evain and Christine Guillemot
- Contact: Simon Evain

6.7. 4D-LFDE

KEYWORDS: Light fields - Depth estimation

FUNCTIONAL DESCRIPTION: This code implements a learning based solution for disparity estimation for either densely or sparsely sampled light fields from 4 corner input views. The code contains two parts "DispEstim" and "DispPropa". The DispEstim module (implemented in tensorflow) takes the 4 corner views of a light field, and estimates the disparity information for these input view positions. The DispPropa module (implemented in Matlab) then generates one disparity map per target view position by propagating corner disparity maps and by applying an occlusion-aware soft 3D reconstruction method. The final output is a .mat file which contains disparity maps for every view positions of a light field.

- Participants: Xiaoran Jiang, Christine Guillemot and Jinglei Shi
- Contact: Xiaoran Jiang

6.8. Compression of omnidirectional images on the sphere

KEYWORDS: Image compression - Omnidirectional image

FUNCTIONAL DESCRIPTION: This code implements an the compression of an omnidirectional directly on the sphere, without a need to project the data on a 2D image. More specifically, from the sphere pixelization, called healpix, the code implements the partitions of the set of pixels into blocks, a block scanning order, an intra prediction between blocks, and a Graph Fourier Transform for each block residual. Finally, the image to be displayed in the viewport is generated.

- Contact: Aline Roumy

6.9. Performance evaluation of interactive video compression schemes and navigation model for omnidirectional images

KEYWORDS: Image compression - Omnidirectional image

FUNCTIONAL DESCRIPTION: This code consists of two parts. First, the code generates typical navigation paths of users viewing omnidirectional images. This generation relies on Markov modeling of the user behavior (probability to choose a first viewing direction to start the navigation, probability to choose a head motion direction, probability of continuing the head motion in the same direction, probability to stop the head motion). The second part of the code implements various criteria to evaluate the compression performance. Three criteria are computed: the distortion averaged along a set of typical navigation paths, the transmission rate related to these navigation paths, and also the storage cost of the compressed image to be able to serve any possible image request. From these 3 criteria, weighted Bjontegaard metric, and iso values are computed.

- Contact: Aline Roumy

6.10. Interactive compression for omnidirectional images and texture maps of 3D models

KEYWORDS: Image compression - Random access

FUNCTIONAL DESCRIPTION: This code implements a new image compression algorithm that allows to navigate within a static scene. To do so, the code provides access in the compressed domain to any block and therefore allows extraction of any subpart of the image. This codec implements this interactive compression for two image modalities: omnidirectional images and texture maps of 3D models. For omnidirectional images the input is a 2D equirectangular projection of the 360 image. The output is the image seen in the viewport. For 3D models, the input is a texture map and the 3D mesh. The output is also the image seen in the viewport.

The code consists of three parts: (A) an offline encoder (B) an online bit extractor and (C) a decoder. The offline encoder (i) partitions the image into blocks, (ii) optimizes the positions of the access blocks, (iii) computes a set of geometry aware predictions for each block (to cover all possible navigation paths), (iv) implements transform quantization for all blocks and their predictions, and finally (v) evaluates the encoding rates. The online bit extractor (Part B) first computes the optimal and geometry aware scanning order. Then it extracts in the bitstream, the sufficient amount of information to allow the decoding of the requested blocks. The last part of the code is the decoder (Part C). The decoder reconstructs the same scanning order as the one computed at the online bit extractor. Then, the blocks are decoded (inverse transform, geometry aware predictions, ...) and reconstructed. Finally the image in the viewport is generated.

- Contact: Aline Roumy

6.11. Platforms

6.11.1. 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 the context of the project ADT ATeP 2016-2018 (funded by Inria), such datasets were 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 have been made available to the community via a public web site. In 2019, we have published and presented our dataset at the ACM MMSys conference [27].

6.11.2. CLIM processing toolbox

Participants: Pierre Allain, Christine Guillemot, Laurent Guillo.

As part of the ERC Clim project, the EPI Sirocco is developing a light field processing toolbox. The toolbox and libraries are developed in C++ and the graphical user interface relies on Qt. As input data, this tool accepts both sparse light fields acquired with High Density Camera Arrays (HDCA) and denser light fields captured with plenoptic cameras using microlens arrays (MLA). At the time of writing, in addition to some simple functionalities, such as re-focusing, change of viewpoints, with different forms of visualization, the toolbox integrates more advanced tools for scene depth estimation from sparse and dense light fields, for super-ray segmentation and scene flow estimation, and for light field denoising and angular interpolation using anisotropic diffusion in the 4D ray space. The toolbox is now being interfaced with the C/C++ API of the tensorflow platform, in order to execute deep models developed in the team for scene depth and scene flow estimation, view synthesis, and axial super-resolution.

7. New Results

7.1. Visual Data Analysis

Scene depth, Scene flows, 3D modelling, Light-fields, 3D point clouds

7.1.1. *Scene depth estimation from light fields*

Participants: Christine Guillemot, Xiaoran Jiang, Jinglei Shi.

While there exist scene depth estimation methods, these methods, mostly designed for stereo content or for pairs of rectified views, do not effectively apply to new imaging modalities such as light fields. We have focused on the problem of *scene depth estimation* for every viewpoint of a dense light field, exploiting information from only a sparse set of views [23]. This problem is particularly relevant for applications such as light field reconstruction from a subset of views, for view synthesis, for 3D modeling and for compression. Unlike most existing methods, the proposed algorithm computes disparity (or equivalently depth) for every viewpoint taking into account occlusions. In addition, it preserves the continuity of the depth space and does not require prior knowledge on the depth range.

We have then proposed a learning based depth estimation framework suitable for both densely and sparsely sampled light fields. The proposed framework consists of three processing steps: initial depth estimation, efficient fusion with occlusion handling and refinement. The estimation can be performed from a flexible subset of input views. The fusion of initial disparity estimates, relying on two warping errors measures, allows us to have an accurate estimation in occluded regions and along the contours. The use of trained neural networks has the advantage of a limited computational cost at estimation time. In contrast with methods relying on the computation of cost volumes, the proposed approach does not need any prior information on the disparity range. Experimental results show that the proposed method outperforms state-of-the-art light fields depth estimation methods for a large range of baselines [14].

The training of the proposed neural networks based architecture requires having ground truth disparity (or depth) maps. Although a few synthetic datasets exist for dense light fields with ground truth depth maps, no such dataset exists for sparse light fields with large baselines. This lack of training data with ground truth depth maps is a crucial issue for supervised learning of neural networks for depth estimation. We therefore created two datasets, namely SLFD and DLFD, containing respectively sparsely sampled and densely sampled synthetic light fields. To our knowledge, SLFD is the first available dataset providing sparse light field views and their corresponding ground truth depth and disparity maps. The created datasets have been made publicly available together with the code and the trained models.

7.1.2. *Scene flow estimation from light fields*

Participants: Pierre David, Christine Guillemot.

We have addressed the problem of scene flow estimation from sparsely sampled video light fields. Scene flows can be seen as 3D extensions of optical flows by also giving the variation in depth along time in addition to the optical flow. Scene flows are tools needed for temporal processing of light fields. Estimating dense scene flows in light fields pose obvious problems of complexity due to the very large number of rays or pixels. This is even more difficult when the light field is sparse, i.e., with large disparities, due to the problem of occlusions. The developments in this area are also made difficult due to the lack of test data, i.e., there is no publicly available synthetic video light fields with the corresponding ground truth scene flows. In order to be able to assess the performance of the proposed method, we have therefore created synthetic video light fields from the MPI Sintel dataset. This video light field data set has been produced with the Blender software by creating new production files placing multiple cameras in the scene, controlling the disparity between the set of views.

We have then developed a local 4D affine model to represent scene flows, taking into account light field epipolar geometry. The model parameters are estimated per cluster in the 4D ray space. We have first developed a sparse to dense estimation method that avoids the difficulty of computing matches in occluded areas [17], which we have further extended by developing a dense scene flow estimation method from light fields. The local 4D affine parameters are in this case derived by fitting the model on initial motion and disparity estimates obtained by using 2D dense optical flow estimation techniques.

We have shown that the model is very effective for estimating scene flows from 2D optical flows (see Fig.2). The model regularizes the optical flows and disparity maps, and interpolates disparity variation values in occluded regions. The proposed model allows us to benefit from deep learning-based 2D optical flow estimation methods while ensuring scene flow geometry consistency in the 4 dimensions of the light field.
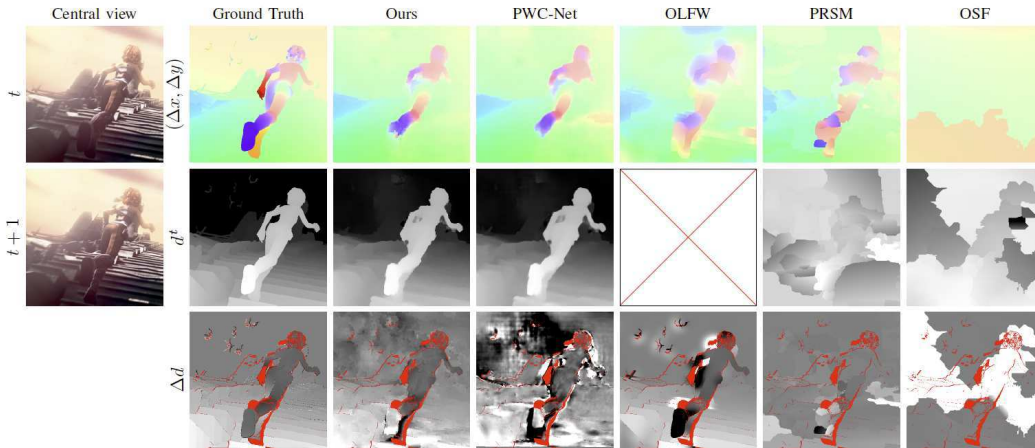
Figure 2. Visual comparison of our method with respect to reference methods (PWC-Net: deep learning method for optical flow estimation; oriented light field window (OLFW), Piece-wise Rigid Scene Model (PRSM), Object Scene Flow (OSF)). First row: optical flows; Second row: disparity maps; Third row: disparity variations. The red pixels are the occlusion mask where there is no ground truth disparity variation available.

7.1.3. *Depth estimation at the decoder in the MPEG-I standard*

Participants: Patrick Garus, Christine Guillemot, Thomas Maugey.

This study, in collaboration with Orange labs., addresses several downsides of the system under development in MPEG-I for coding and transmission of immersive media. We study a solution, which enables Depth-Image-Based Rendering for immersive video applications, while lifting the requirement of transmitting depth information. Instead, we estimate the depth information on the client-side from the transmitted views. We have observed that doing this leads to a significant rate saving (37.3% in average). Preserving perceptual quality in terms of MS-SSIM of synthesized views, it yields to 24.6% rate reduction for the same quality of reconstructed views after residue transmission under the MPEG-I common test conditions. Simultaneously, the required pixel rate, i.e. the number of pixels processed per second by the decoder, is reduced by 50% for any test sequence [21].

7.1.4. *Spherical feature extraction for 360 light field reconstruction from omni-directional fish-eye camera captures*

Participants: Christine Guillemot, Fatma Hawary, Thomas Maugey.

With the increasing interest in wide-angle or 360° scene captures, the extraction of descriptors well suited to the geometry of this content is a key problem for a variety of processing tasks. Algorithms designed for feature extraction in 2D images are hardly applicable to 360° images or videos as they do not well take into account their specific spherical geometry. To cope with this difficulty, it is quite common to perform an equirectangular projection of the spherical content, and to compute spherical features on projected and stitched content. However, this process introduces geometrical distortions with implications on the accuracy of applications such as angle estimation, depth calculation and 3D scene reconstruction. We adapt a spherical feature descriptor to the geometry of fish-eye cameras that avoids equirectangular projection. The captured image is directly mapped onto a spherical model of the 360° camera. In order to evaluate the interest of the proposed fish-eye adapted descriptor, we consider the angular coordinates of feature points on the sphere. We assess the stability of the corresponding angles when capturing the scene by a moving fish-eye camera.

Experimental results show that the proposed fish-eye adapted descriptor allows a more stable angle estimation, hence a more robust feature detection, compared to spherical features on projected and stitched contents.

7.2. Signal processing and learning methods for visual data representation and compression

Sparse representation, data dimensionality reduction, compression, scalability, rate-distortion theory

7.2.1. Single sensor light field acquisition using coded masks

Participants: Christine Guillemot, Ehsan Miandji, Hoai Nam Nguyen.

We developed a simple variational approach for reconstructing color light fields in the compressed sensing framework with very low sampling ratio, using both coded masks and color filter arrays (CFA). A coded mask is placed in front of the camera sensor to optically modulate incoming rays, while a color filter array is assumed to be implemented at the sensor level to compress color information. Hence, the light field coded projections, operated by a combination of the coded mask and the CFA, measure incomplete color samples with a three times lower sampling ratio than reference methods that assume full color (channel-by-channel) acquisition. We then derived adaptive algorithms to directly reconstruct the light field from raw sensor measurements by minimizing a convex energy composed of two terms. The first one is the data fidelity term which takes into account the use of CFAs in the imaging model, and the second one is a regularization term which favors the sparse representation of light fields in a specific transform domain. Experimental results show that the proposed approach produces a better reconstruction both in terms of visual quality and quantitative performance when compared to reference reconstruction methods that implicitly assume prior color interpolation of coded projections.

We then pursued this study by developing a unifying image formation model that abstracts the architecture of most existing compressive-sensing light-field cameras, equipped with single lens and coded masks, as an equivalent multi-mask camera. It allows to compare different designs with a number of criteria: compression rate, light efficiency, measurement incoherence, as well as acquisition quality. Moreover, the underlying multi-mask camera can be flexibly adapted for various applications, such as single and multiple acquisitions, spatial super-resolution, parallax reconstruction, and color restoration. We also derived a generic variational algorithm solving all these concrete problems by considering appropriate sampling operators.

7.2.2. 3D point cloud processing and plenoptic point cloud compression

Participants: Christian Galea, Christine Guillemot, Maja Krivocuca.

Light fields, by capturing light rays emitted by a 3D scene along different orientations, give a very rich description of the scene enabling a variety of computer vision applications. The recorded 4D light field gives in particular information about the parallax and depth of the scene. The estimated depth can then be used to construct 3D models of the scene, e.g. in the form of a 3D point cloud, The constructed 3D point clouds, however, generally contain distortions and artefacts primarily caused by inaccuracies in the depth maps. We have developed a method for noise removal in 3D point clouds constructed from light fields [20]. While existing methods discard outliers, the proposed approach instead attempts to correct the positions of points, and thus reduce noise without removing any points, by exploiting the consistency among views in a light-field. The proposed 3D point cloud construction and denoising method exploits uncertainty measures on depth values.

Beyond classical 3D point clouds, plenoptic point clouds can be seen as natural extensions of 3D point clouds to Surface Light Fields (SLF). While the concept of surface light field (SLF) has been introduced as a function that assigns a color to each ray originating on a surface, plenoptic point clouds represent in each voxel illuminationa and color seen from different camera viewpoints. In other words, instead of each point being associated with a single colour value, there can be multiple values to represent the colour at that point as perceived from different viewpoints. This concept aims at combining the best of light fields and computer graphics modeling, for photo-realistic rendering from arbitrary points of view. However, this representation

leads to color maps per voxel, hence to large volumes of data. We have addressed the problem of efficient compression of this data based on the Region-Adaptive Hierarchical Transform (RAHT) method in which we have introduced clustering and speculat/diffuse components separation showing better adapted plenoptic point cloud color maps transforms.

7.2.3. *Low rank models and representations for light fields*

Participants: Elian Dib, Christine Guillemot, Xiaoran Jiang.

We have addressed the problem of light field dimensionality reduction. We have introduced a local low rank approximation method using a parameteric disparity model. The local support of the approximation is defined by super-rays. Superrays can be seen as a set of super-pixels that are coherent across all light field views. The light field low rank assumption depends on how much the views are correlated, i.e. on how well they can be aligned by disparity compensation. We have therefore introduced a disparity estimation method using a low rank prior. We have considered a parametric model describing the local variations of disparity within each super-ray, and alternatively search for the best parameters of the disparity model and of the low rank approximation. We have assessed the proposed disparity parametric model, by considering an affine disparity model. We have shown that using the proposed disparity parametric model and estimation algorithm gives an alignment of superpixels across views that favours the low rank approximation compared with using disparity estimated with classical computer vision methods. The low rank matrix approximation is ten computed on the disparity compensated super-rays using a singular value decomposition (SVD). A coding algorithm has been developed for the different components of the proposed disparity-compensated low rank approximation [19].

We have also, in collaboration with Trinity College Dublin, introduced a new Light Field representation for efficient Light Field processing and rendering called Fourier Disparity Layers (FDL) [11]. The proposed FDL representation samples the Light Field in the depth (or equivalently the disparity) dimension by decomposing the scene as a discrete sum of layers. The layers can be constructed from various types of Light Field inputs including a set of sub-aperture images, a focal stack, or even a combination of both. From our derivations in the Fourier domain, the layers are simply obtained by a regularized least square regression performed independently at each spatial frequency, which is efficiently parallelized in a GPU implementation. Our model is also used to derive a gradient descent based calibration step that estimates the input view positions and an optimal set of disparity values required for the layer construction. Once the layers are known, they can be simply shifted and filtered to produce different viewpoints of the scene while controlling the focus and simulating a camera aperture of arbitrary shape and size. A direct implementation in the Fourier domain allows real time Light Field rendering. Finally, direct applications such as view interpolation or extrapolation and denoising have also been evaluated [11]. The use of this representation for view synthesis based compression has also been assessed in [18].

7.2.4. *Graph-based transforms and prediction for light fields*

Participants: Christine Guillemot, Thomas Maugey, Mira Rizkallah.

We have investigated Graph-based transforms for low dimensional embedding of light field data. Both non separable and separable transforms have been considered. The low-dimensional embedding can be learned with a few eigen vectors of the graph Laplacian. However, the dimension of the data (e.g. light fields) has obvious implications on the storage footprint of the Laplacian matrix and on the eigenvectors computation complexity, making graph-based non separable transforms impractical for such data. To cope with this difficulty, we have developed local super-rays based non separable and separable (spatial followed by angular) weighted and unweighted transforms to jointly capture light fields correlation spatially and across views [13]. Despite the local support of limited size defined by the super-rays, the Laplacian matrix of the non separable graph remains of high dimension and its diagonalization to compute the transform eigen vectors remains computationally expensive. To solve this problem, we have then performed the local spatio-angular transform in a separable manner.

While separable transforms on super-rays allow us to significantly decrease the eigenvector computation complexity. The basis functions of the spatial graph transforms to be applied on the super-ray pixels of each view are often not compatible. We have indeed shown that when the shape of corresponding super-pixels in the different views is not isometric, the basis functions of the spatial transforms are not coherent, resulting in decreased correlation between spatial transform coefficients, hence in a loss of performance of the angular transform, compared to the non-separable case. We have therefore developed a graph construction optimization procedure which seeks to find the eigen-vectors which align the best with those of a reference one while still approximately diagonalizing their respective Laplacians [13]. The proposed optimization method aims at preserving angular correlation even when the shapes of the super-pixels are not isometric. Experimental results show the benefit of the approach in terms of energy compaction. A coding scheme has also been developed to assess the rate-distortion perfomances of the proposed transforms

The use of local transforms with limited supports is a way to cope with the computational difficulty. Unfortunately, the locality of the support may not allow us to fully exploit long term signal dependencies present in both the spatial and angular dimensions in the case of light fields. We have therefore introduced sampling and prediction schemes, based on graph sampling theory, with local graph-based transforms enabling to efficiently compact the signal energy and exploit dependencies beyond the local graph support [30], [12]. The proposed approach has been shown to be very efficient in the context of spatio-angular transforms for quasi-lossless compression of light fields.

7.2.5. *Intra-coding of 360-degree images on the sphere*
Participants: Navid Mahmoudian Bidgoli, Thomas Maugey, Aline Roumy.

Omni-directional images are characterized by their high resolution (usually 8K) and therefore require high compression efficiency. Existing methods project the spherical content onto one or multiple planes and process the mapped content with classical 2D video coding algorithms. However, this projection induces sub-optimality. Indeed, after projection, the statistical properties of the pixels are modified, the connectivity between neighboring pixels on the sphere might be lost, and finally, the sampling is not uniform. Therefore, we propose to process uniformly distributed pixels directly on the sphere to achieve high compression efficiency. In particular, a scanning order and a prediction scheme are proposed to exploit, directly on the sphere, the statistical dependencies between the pixels. A Graph Fourier Transform is also applied to exploit local dependencies while taking into account the 3D geometry. Experimental results demonstrate that the proposed method provides up to 5.6% bitrate reduction and on average around 2% bitrate reduction over state-of-the-art methods. This work has led to a publication in the PCS conference 2019 [25].

7.3. Algorithms for inverse problems in visual data processing

Inpainting, view synthesis, super-resolution

7.3.1. *View synthesis in light fields and stereo set-ups*
Participants: Simon Evain, Christine Guillemot, Xiaoran Jiang, Jinglei Shi.

We have developed a learning-based framework for light field view synthesis from a subset of input views. Building upon a light-weight optical flow estimation network to obtain depth maps, our method employs two reconstruction modules in pixel and feature domains respectively. For the pixel-wise reconstruction, occlusions are explicitly handled by a disparity-dependent interpolation filter, whereas inpainting on disoccluded areas is learned by convolutional layers. Due to disparity inconsistencies, the pixel-based reconstruction may lead to blurriness in highly textured areas as well as on object contours. On the contrary, the feature-based reconstruction well performs on high frequencies, making the reconstruction in the two domains complementary. End-to-end learning is finally performed including a fusion module merging pixel and feature-based reconstructions. Experimental results show that our method achieves state-of-the-art performance on both synthetic and real-world datasets, moreover, it is even able to extend light fields baseline by extrapolating high quality views without additional training.

We have also designed a very lightweight neural network architecture, trained on stereo data pairs, which performs view synthesis from one single image. With the growing success of multi-view formats, this problem is indeed increasingly relevant. The network returns a prediction built from disparity estimation, which fills in wrongly predicted regions using a occlusion handling technique. To do so, during training, the network learns to estimate the left-right consistency structural constraint on the pair of stereo input images, to be able to replicate it at test time from one single image. The method is built upon the idea of blending two predictions: a prediction based on disparity estimation, and a prediction based on direct minimization in occluded regions. The network is also able to identify these occluded areas at training and at test time by checking the pixelwise left-right consistency of the produced disparity maps. At test time, the approach can thus generate a left-side and a right-side view from one input image, as well as a depth map and a pixelwise confidence measure in the prediction. The work outperforms visually and metric-wise state-of-the-art approaches on the challenging KITTI dataset, all while reducing by a very significant order of magnitude (5 or 10 times) the required number of parameters (6.5 M).

7.3.2. *Inverse problems in light field imaging with 4D anisotropic diffusion and neural networks*

Participants: Pierre Allain, Christine Guillemot, Laurent Guillo.

We have addressed inverse problems in ligt field imaging by following two methodological directions. We first introduced a 4D anisotropic diffusion framework based on PDEs [4]. The proposed regularization method operated in the 4D ray space and, unlike the methods operating on epipolar plane images, does not require prior estimation of disparity maps. The method performs a PDE-based diffusion with anisotropy steered by a tensor field based on local structures in the 4D ray space that we extract using a 4D tensor structure. To enhance coherent structures, the smoothing along directions, surfaces, or volumes in the 4D ray space is performed along the eigenvectors directions. Although anisotropic diffusion is well understood for 2D imaging, its interpretation and understanding in the 4D space is far from being straightforward. We have analysed the behaviour of the diffusion process on a light field toy example, i.e. a tesseract (a 4D cube). This simple light field example allows an in-depth analysis of how each eigenvector influences the diffusion process. The proposed ray space regularizer is a tool that has enabled us to tackle a variety of inverse problems (denoising, angular and spatial interpolation, regularization for enhancing disparity estimation as well as inpainting) in the ray space.

In collaboration with the university of Malta (Pr. Reuben Farrugia), we have explored the benefit of low rank prios in light field super-resolution with deep neural networks. This led us to design a learning-based spatial light field super-resolution method that allows the restoration of the entire light field with consistency across all sub-aperture images [7]. The algorithm first uses optical flows to align the light field views and then reduces its angular dimension using low-rank approximation. We then consider the linearly independent columns of the resulting low-rank model as an embedding, which is restored using a deep convolutional neural network. The super-resolved embedding is then used to reconstruct the remaining sub-aperture images. The original disparities are restored using inverse warping where missing pixels are approximated using a novel light field inpainting algorithm. We pursued this study by designing an approach that, thanks to a low rank approximation model, can leverage models learned for 2D image super-resolution [8]. This approach avoids the need for a large amount of light field training data which is, unlike 2D images, not available. It also allows us to reduce the dimension, hence the number of parameters, of the network to be learned.

7.3.3. *Neural networks for axial light field super-resolution*

Participants: Christine Guillemot, Zhaolin Xiao.

Axial light field resolution refers to the ability to distinguish features at different depths by refocusing. The axial refocusing precision corresponds to the minimum distance in the axial direction between two distinguishable refocusing planes. High refocusing precision can be essential for some light field applications like microscopy. We first introduced a refocusing precision model based on a geometrical analysis of the flow of rays within the virtual camera. The model establishes the relationship between the feature distinguishability by refocusing and different camera settings. We have then developed a learning-based method to extrapolate

novel views from axial volumes of sheared epipolar plane images (EPIs (see an example of extrapolated views in Fig.3)). As extended numerical aperture (NA) in classical imaging, the extrapolated light field gives re-focused images with a shallower depth of field (DOF), leading to more accurate refocusing results. Most importantly, the proposed approach does not need accurate depth estimation. Experimental results with both synthetic and real light fields, including with microscopic data, demonstrate that our approach can effectively enhance the light field axial refocusing precision.
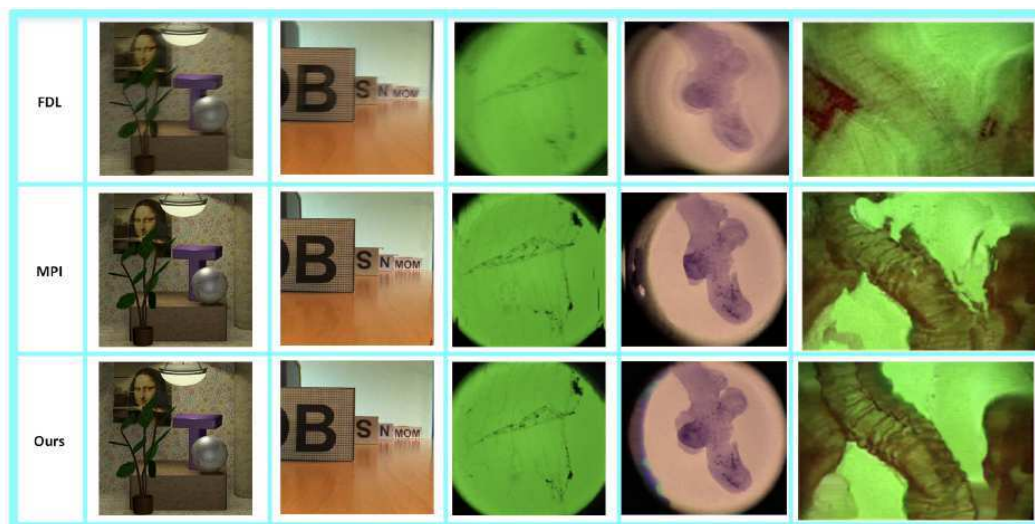

Figure 3. Extrapolation results with a 4X larger baseline, in comparison with reference methods using multiple plane images (MPI) and Fourier disparity layers (FDL).

7.3.4. *Neural networks for inverse problems in 2D imaging*

Participants: Christine Guillemot, Aline Roumy, Alexander Sagel.

The Deep Image Prior has been recently introduced to solve inverse problems in image processing with no need for training data other than the image itself. However, the original training algorithm of the Deep Image Prior constrains the reconstructed image to be on a manifold described by a convolutional neural network. For some problems, this neglects prior knowledge and can render certain regularizers ineffective. We have developed an alternative approach that relaxes this constraint and fully exploits all prior knowledge. We have evaluated our algorithm on the problem of reconstructing a high-resolution image from a downsampled version and observed a significant improvement over the original Deep Image Prior algorithm.

7.4. Distributed coding for interactive 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 compression scheme for interactive media*

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

We propose a new interactive compression scheme for omnidirectional images and 3D model. This requires two characteristics: efficient compression of data, to lower the storage cost, and random access ability to extract part of the compressed stream requested by the user (for reducing the transmission rate). For efficient compression, data needs to be predicted by a series of references that have been pre-defined and compressed. This contrasts with the spirit of random accessibility. We propose a solution for this problem based on incremental codes implemented by rate adaptive channel codes. This scheme encodes the image while adapting to any user request and leads to an efficient coding that is flexible in extracting data depending on the available information at the decoder. Therefore, only the information which is needed to be displayed at the user's side is transmitted during the user's request as if the request was already known at the encoder ((see Fig. 4). The experimental results demonstrate that our coder obtains better transmission rate than the state-of-the-art tile-based methods at a small cost in storage. Moreover, the transmission cost grows gradually with the size of the request and avoids a staircase effect, which shows the perfect suitability of our coder for interactive transmission. This work has led to a journal submission and several conference publications. In [24], we have proposed a new framework for evaluating the compression performance of interactive schemes. Indeed, interactive compression schemes can be characterized by tree criteria: the storage cost, the transmission rate and distortion. This contrasts with classical compression scheme, where only transmission rate and distortion are used. 3D-performance evaluation criteria are proposed. In [28], we have proposed to use the geometry to efficiently compress the 3D mesh texture. An interactive coding extension has been presented in [26].

Figure 4. A spherical image and several viewports corresponding to different user's requests.

7.4.2. *Reference source positioning for interactive compression*

Participants: Thomas Maugey, Mai Quyen Pham, Aline Roumy.

Large databases containing many HD videos or records from sensors over long time intervals, have to be efficiently compressed, to reduce their size. The compression has also to allow efficient access to random parts of the databases upon request from the users. Efficient compression is usually achieved with prediction between data points. However, this creates dependencies between the compressed representations, which is contrary to the idea of random access. Prediction methods rely in particular on reference data points, used to

predict other data points, and the placement of these references balances compression efficiency and random access. Existing solutions to position the references use ad hoc methods. We study this joint problem of compression efficiency and random access. We introduce the storage cost as a measure of the compression efficiency and the transmission cost for the random access ability. We show that the reference placement problem that trades off storage with transmission cost is an integer linear programming problem, that can be solved by standard optimizer. Moreover, we show that the classical periodic placement of the references is only optimal in a very restrictive case: namely, when the encoding costs of each data point are equal and when requests of successive data points are made.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

8.1.1. CIFRE contract with InterDigital on neural networks for video compression

Participants: Hien Pham, Christine Guillemot.

- Title : Neural networks for video compression
- Partners : InterDigital (Ph. Bordes, F. Galpin), Inria-Rennes.
- Funding : InterDigital, ANRT.
- Period : Jan.2019-Oct.2021.

The goal of this Cifre contract is to first investigate novel optical flow estimation methods using deep neural networks. Based on the optical flow methods, the next step will be to design temporal prediction schemes based on convolutional neural networks (CNN) for video compression. The methods will be assessed in the context of the VVC (Versatile Video Coding) standard.

8.1.2. CIFRE contract with Orange labs. on compression of immersive content

Participants: Patrick Garus, Christine Guillemot, Thomas Maugey.

- Title : Compression of immersive content
- Research axis : 7.1.3
- Partners : Orange labs. (J. Jung), Inria-Rennes.
- Funding : InterDigital, ANRT.
- Period : Jan.2019-Dec.2021.

The goal of this Cifre contract is to develop novel compression methods for 6 DoF immersive video content. This implies investigating depth estimation and view synthesis methods that would be robust to quantization noise. This also implies developing the corresponding coding mode decisions based on rate-distortion criteria.

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 : 7.4.1
- Partners : Inria-Rennes (Sirocco team and i4S team); LabSTICC, IMT Atlantique, Signal & Communications Department; External partners: L2S, CentraleSupelec, Univ. Paris Sud; EPFL, Signal Processing Laboratory (LTS4).
- Funding : Labex CominLabs.
- Period : Oct. 2016 - Dec. 2020.

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 are being 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 Allain, Pierre David, Elian Dib, Simon Evain, Christian Galea, Christine Guillemot, Laurent Guillo, Fatma Hawary, Xiaoran Jiang, Maja Krivokuca, Ehsan Miandji, Hoai Nam Nguyen, Mira Rizkallah, Alex Sagel, Jinglei Shi.

- Title : Computational Light field Imaging.
- Research axis : 7.1.1, 7.1.2, 7.1.4, 7.2.1, 7.2.3, 7.2.4, 7.2.2, 7.3.1, 7.3.2, 7.3.3, 7.3.4
- Partners : Inria-Rennes
- Funding : European Research Council (ERC) advanced grant
- Period : Sept. 2016 - Aug. 2021.

All imaging systems, when capturing a view, record different combinations of light rays emitted by the environment. In a conventional camera, each sensor element sums all the light rays emitted by one point over the lens aperture. Light field cameras instead measure the light along each ray reaching the camera sensors and not only the sum of rays striking each point in the image. In one single exposure, they capture the geometric distribution of light passing through the lens. This process can be seen as sampling the plenoptic function that describes the intensity of the light rays interacting with the scene and received by an observer at every point in space, along any direction of gaze, for all times and every wavelength.

The recorded flow of rays (the light field) is in the form of high-dimensional data (4D or 5D for static and dynamic light fields). The 4D/5D light field yields a very rich description of the scene enabling advanced creation of novel images from a single capture, e.g. for computational photography by simulating a capture with a different focus and a different depth of field, by simulating lenses with different apertures, by creating images with different artistic intents. It also enables advanced scene analysis with depth and scene flow estimation and 3D modelling. The goal of the ERC-CLIM project is to develop algorithms for the entire static and video light fields processing chain. The planned research includes the development of:

- novel low rank or graph-based models for dimensionality reduction and compression
- deep learning methods for scene analysis (e.g. scene depth and scene flow estimation)
- learning methods for solving a range of inverse problems: denoising, super-resolution, axial super-resolution, view synthesis.

9.3. International Initiatives

9.3.1. Inria International Labs

EPFL-Inria: Associate Team involved in the International Lab: Graph-based Omnidirectional video Processing (GOP)

- Participant: Thomas Maugey
- International Partner (Institution - Laboratory - Researcher): Ecole Polytechnique Fédérale de Lausanne (Switzerland) - LTS4 - Pascal Frossard
- period: 2017-2019

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 led 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. During this project, we study how graphs can be built for defining a suitable structure on one or several 360 videos and then used for compression.

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 started 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.

- Mikael Le Pendu and Prof. Aljosa Smolic from Trinity College Dublin on HDR light field recovery from multiple exposures.

- Pascal Frossard, Prof. at EPFL, in the context of the Comin Lab/Intercom project and in the context of the EPFL-Inria associated team.

9.4. International Research Visitors

9.4.1. *Visits of International Scientists*

9.4.1.1. Internships

- Zhaolin Xiao, Prof. Xian University, Dec. 2018-Nov. 2019.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. *Scientific Events: Organisation*

10.1.1.1. General Chair, Scientific Chair

- C. Guillemot has organized, as chair, the ERC-CLIM workshop, May 2019.

10.1.2. *Scientific Events: Selection*

10.1.2.1. Member of the Conference Program Committees

- C. Guillemot has been member of the international steering committee of the Picture Coding Symposium (PCS), 2019.

- C. Guillemot has been member of the technical programme committee of IEEE-ICASSP 2019.

- C. Guillemot has been a member of the technical program committee of the CVPR 2019 workshop on New Trends in Image Restoration and Enhancement (NTIRE), 2019.

- A. Roumy has been a member of the technical program committee of the CVPR 2019 workshop on New Trends in Image Restoration and Enhancement (NTIRE).

- A. Roumy has been a member of the technical program committee of the IEEE International Conference on Communications (ICC) 2019, Workshop on Machine Learning in Wireless Communications (ML4COM).
- A. Roumy has been a member of the technical program committee of the IEEE Wireless Communications and Networking Conference 2019 (IEEE WCNC).
- A. Roumy has been a member of the technical program committee of the International conference on Telecommunications (ICT) 2019.
- A. Roumy has been a member of the technical program committee of the Gretsi 2019 conference, organized by the National Research group in Image and Signal Processing.

10.1.3. Journal

10.1.3.1. 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.
- T. Maugey is associate editor of the EURASIP Journal on Advances in Signal Processing (since Dec 2019).
- A. Roumy is associate editor of the Springer Annals of Telecommunications.
- A. Roumy is associate editor of the IEEE Trans. on Image Processing.

10.1.4. Invited Talks

- C. Guillemot gave a keynote talk on Light field image processing, at the Digital Optical Technologies International Conference, Munich, June 24-27, 2019.
- Thomas Maugey gave a seminar at IRISA's days on Art, Culture and Heritage, Rennes, France, "Data acquisition and compression for user immersion in a virtual scene" (Jan. 2019).
- Thomas Maugey gave a seminar at Orange Labs. PhD days, Rennes, France, "Light Field Acquisition and Depth estimation" (Mar. 2019).
- Thomas Maugey gave a seminar at GdR-Isis, Rennes, France, "Data acquisition and compression for user immersion in a 3D scene" (Mar. 2019).
- A. Roumy gave a talk on "Learning transforms for image compression", workshop on Learning for unstructured data, Univ. Paris 13, Nov. 2019.

10.1.5. Leadership within the Scientific Community

- C. Guillemot is member of the IEEE Signal Processing Society Nominations and Appointments Committee for a two-year term.
- A. Roumy is member of the IEEE IVMSP technical committee.
- A. Roumy is a Local Liaison Officer for the European Association for Signal Processing (EURASIP).
- A. Roumy is a member of the Executive board of the National Research group in Image and Signal Processing (GRETSI).

10.1.6. Scientific Expertise

- C. Guillemot has served as expert in a research proposal selection committee of the Irish Research Council, Feb. 2019.
- C. Guillemot is responsible of the theme « compression et protection des données images » for the realization of the encyclopedia SCIENCES of the publisher ISTE/Wiley (2019-2021).

10.1.7. Research Administration

- C. Guillemot is member of the "bureau du Comité des Projets".

- C. Guillemot has served as a member of the selection committee of a Prof. at the National University of Ireland (NUI), Galway, Ireland, Feb. 2019.
- C. Guillemot has served as a member of the selection committee of a Prof. at the Technical Univ. of Denmark, May 2019.
- A. Roumy is a member of Inria evaluation committee.
- A. Roumy has served as a member of the selection committee of an Assistant Prof. at University Paris 13 (May 2019).

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

- Master: C. Guillemot, Image and video compression, 10 hours (2018-2019), and advanced video processing, 4 hours (2018-2019), M2 SISEA, Univ. of Rennes 1, France.
- Undergraduate: L. Guillo, course of 35 hours on Functional immutable programming, 1st year of Mathematiques, Informatique, Electronique, mathematique-Economie" (MIEE), Univ. of Rennes 1
- Master: T. Maugey, course on 3D models in a module on advanced video, 4 hours (2018-2019), 12 hours (2019-2020), M2 SISEA, Univ. of Rennes 1, France.
- Master: T. Maugey, course on Image compression, 10 hours (2019-2020), M2 SISEA, Univ. of Rennes 1, France.
- Master: T. Maugey, course on Representation, editing and perception of digital images, 12 hours, M2 SIF, 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.
- Master: A. Roumy, Foundations of smart sensing, 18 hours, ENSAI, Master of Science in Statistics for Smart Data, France.
- Master: A. Roumy, Information theory, 15 hours, University Rennes 1, SIF master, France.

10.2.2. Juries

- C. Guillemot has been reviewer of the PhDs of:
    - B. Ray, UNiv. Poitiers, March 2019
    - T. Trinh Le, Univ. Paris Saclay, June 2019
    - Z. Chen, Chinese Univ. of Hong Kong, June 2019
    - D. Tsai, Queensland Univ. of Technology (QUT), Nov. 2019
    - M. De Pinheiro Carvalho, Univ. Paris Saclay, Nov. 2019
- C. Guillemot has been member of the PhD committees of:
    - M. Bichon, Univ. Rennes 1, March 2019
    - C.A. Noury, Univ. Clermond Ferrand, Nov. 2019
    - S. Manandhar, Univ. Rennes 1, Nov. 2019
- Aline Roumy has been reviewer of the PhDs of:
    - Shuo Zheng, Univ. Paris Saclay, prepared at Telecom ParisTech, Feb. 2019
    - Mohsen Abdoli , Univ. Paris Saclay, prepared at Univ. Paris Sud, June 2019
    - David Kibloff , Univ. Lyon, prepared at Insa Lyon, Sept. 2019
- Aline Roumy has been member of the PhD committees of:
    - Yigit Ugur, Univ. Paris Est, Nov. 2019
    - Fangping Ye, IMT Atlantique, Brest, Dec. 2019

10.3. Popularization

10.3.1. Internal or external Inria responsibilities

- C. Guillemot is responsible of the theme « compression et protection des données images » for the realization of the encyclopedia SCIENCES of the publisher ISTE/Wiley (2019-2021).

11. Bibliography

Publications of the year

Doctoral Dissertations and Habilitation Theses

[1] F. HAWARY. *Light Field Image Compression and Compressive Acquisition*, Université de Rennes 1, France ; Inria, May 2019, https://hal.archives-ouvertes.fr/tel-02378409

[2] N. MAHMOUDIAN BIDGOLI. *Compression for interactive communication of visual contents*, Université de Rennes 1, November 2019, https://hal.inria.fr/tel-02410190

[3] M. RIZKALLAH. *Graph based transforms for compression of new imaging modalities*, Université Rennes 1, April 2019, https://tel.archives-ouvertes.fr/tel-02285386

Articles in International Peer-Reviewed Journals

[4] P. ALLAIN, L. GUILLO, C. GUILLEMOT. *4D Anisotropic Diffusion Framework with PDEs for Light Field Regularization and Inverse Problems*, in "IEEE Transactions on Computational Imaging", May 2019, pp. 1-16 [*DOI :* 10.1109/TCI.2019.2919229], https://hal.archives-ouvertes.fr/hal-02135541

[5] T. DUMAS, A. ROUMY, C. GUILLEMOT. *Context-adaptive neural network based prediction for image compression*, in "IEEE Transactions on Image Processing", August 2019, vol. 29, pp. 679 - 693, https://arxiv.org/abs/1807.06244 , https://hal.archives-ouvertes.fr/hal-01841034

[6] E. DUPRAZ, A. ROUMY, T. MAUGEY, M. KIEFFER. *Rate-storage regions for Extractable Source Coding with side information*, in "Physical Communication", 2019, vol. 37, pp. 1-24 [*DOI :* 10.1016/J.PHYCOM.2019.100845], https://hal.inria.fr/hal-02390517

[7] R. A. FARRUGIA, C. GUILLEMOT. *Light Field Super-Resolution using a Low-Rank Prior and Deep Convolutional Neural Networks*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", 2019, pp. 1-15 [*DOI :* 10.1109/TPAMI.2019.2893666], https://hal.archives-ouvertes.fr/hal-01984843

[8] R. FARRUGIA, C. GUILLEMOT. *A Simple Framework to Leverage State-Of-The-Art Single-Image Super-Resolution Methods to Restore Light Fields*, in "Signal Processing: Image Communication", September 2019, vol. 80, pp. 1-11 [*DOI :* 10.1016/J.IMAGE.2019.115638], https://hal.archives-ouvertes.fr/hal-02378049

[9] C. GALDI, V. CHIESA, C. BUSCH, P. L. CORREIA, J.-L. DUGELAY, C. GUILLEMOT. *Light Fields for Face Analysis*, in "Sensors", June 2019, vol. 19, n° 12, pp. 1-27 [*DOI :* 10.3390/S19122687], https://hal.archives-ouvertes.fr/hal-02157348

[10] D. GOMMELET, J. LE TANOU, A. ROUMY, M. ROPERT, C. GUILLEMOT. *Optical-Flow Based Nonlinear Weighted Prediction for SDR and Backward Compatible HDR Video Coding*, in "IEEE Transactions on Image Processing", 2019, pp. 1-16 [*DOI* : 10.1109/TIP.2019.2945685], https://hal.archives-ouvertes.fr/hal-02378053

[11] M. LE PENDU, C. GUILLEMOT, A. SMOLIC. *A Fourier Disparity Layer representation for Light Fields*, in "IEEE Transactions on Image Processing", May 2019, pp. 5740 - 5753 [*DOI* : 10.1109/TIP.2019.2922099], https://hal.archives-ouvertes.fr/hal-02130555

[12] M. RIZKALLAH, T. MAUGEY, C. GUILLEMOT. *Prediction and Sampling with Local Graph Transforms for Quasi-Lossless Light Field Compression*, in "IEEE Transactions on Image Processing", 2019, pp. 1-13, https://hal.inria.fr/hal-02373664

[13] M. RIZKALLAH, X. SU, T. MAUGEY, C. GUILLEMOT. *Geometry-Aware Graph Transforms for Light Field Compact Representation*, in "IEEE Transactions on Image Processing", August 2019, pp. 1-15 [*DOI* : 10.1109/TIP.2019.2928873], https://hal.archives-ouvertes.fr/hal-02199839

[14] J. SHI, X. JIANG, C. GUILLEMOT. *A framework for learning depth from a flexible subset of dense and sparse light field views*, in "IEEE Transactions on Image Processing", June 2019, pp. 5867-5880 [*DOI* : 10.1109/TIP.2019.2923323], https://hal.archives-ouvertes.fr/hal-02155040

International Conferences with Proceedings

[15] P. ALLAIN, L. GUILLO, C. GUILLEMOT. *Light field denoising using 4D anisotropic diffusion*, in "ICASSP 2019 - IEEE International Conference on Acoustics, Speech, and Signal Processing", Brighton, United Kingdom, IEEE, May 2019, pp. 1692-1696 [*DOI* : 10.1109/ICASSP.2019.8682709], https://hal.archives-ouvertes.fr/hal-02116372

[16] J. BÉGAINT, F. GALPIN, P. GUILLOTEL, C. GUILLEMOT. *Deep frame interpolation for video compression*, in "DCC 2019 - Data Compression Conference", Snowbird, United States, IEEE, March 2019, pp. 1-10 [*DOI* : 10.1109/DCC.2019.00068], https://hal.archives-ouvertes.fr/hal-02202172

[17] P. DAVID, M. LE PENDU, C. GUILLEMOT. *Sparse to Dense Scene Flow Estimation from Light Fields*, in "ICIP 2019 - IEEE International Conference on Image Processing", Taïpei, Taiwan, IEEE, September 2019, pp. 1-5 [*DOI* : 10.1109/ICIP.2019.8803520], https://hal.archives-ouvertes.fr/hal-02123544

[18] E. DIB, M. LE PENDU, C. GUILLEMOT. *Light Field Compression using Fourier Disparity Layers*, in "ICIP 2019 - IEEE International Conference on Image Processing", Taipei, Taiwan, IEEE, September 2019, pp. 1-5 [*DOI* : 10.1109/ICIP.2019.8803756], https://hal.archives-ouvertes.fr/hal-02130487

[19] E. DIB, M. LE PENDU, X. JIANG, C. GUILLEMOT. *Super-rays based low rank approximation for light fields compression*, in "DCC 2019 - Data Compression Conference", Snowbird, United States, IEEE, March 2019, pp. 369-378 [*DOI* : 10.1109/DCC.2019.00045], https://hal.archives-ouvertes.fr/hal-02116365

[20] C. GALEA, C. GUILLEMOT. *Denoising of 3D point clouds constructed from light fields*, in "ICASSP 2019 - IEEE International Conference on Acoustics, Speech, and Signal Processing", Brighton, United Kingdom, IEEE, May 2019, pp. 1882-1886 [*DOI* : 10.1109/ICASSP.2019.8683548], https://hal.archives-ouvertes.fr/hal-02116377

[21] P. GARUS, J. JOEL, T. MAUGEY, C. GUILLEMOT. *Bypassing Depth Maps Transmission For Immersive Video Coding*, in "PCS 2019 - Picture Coding Symposium", Ningbo, China, IEEE, November 2019, pp. 1-5, https://hal.inria.fr/hal-02397800

[22] M. HOG, N. SABATER, C. GUILLEMOT. *Long Short Term Memory Networks For Light Field View Synthesis*, in "ICIP 2019 - IEEE International Conference on Image Processing", Taipei, Taiwan, IEEE, September 2019, pp. 1-5 [*DOI :* 10.1109/ICIP.2019.8803790], https://hal.archives-ouvertes.fr/hal-02202120

[23] X. JIANG, J. SHI, C. GUILLEMOT. *A learning based depth estimation framework for 4D densely and sparsely sampled light fields*, in "ICASSP 2019 - IEEE International Conference on Acoustics, Speech, and Signal Processing", Brighton, United Kingdom, IEEE, May 2019, pp. 2257-2261 [*DOI :* 10.1109/ICASSP.2019.8683773], https://hal.archives-ouvertes.fr/hal-02116375

[24] N. MAHMOUDIAN BIDGOLI, T. MAUGEY, A. ROUMY. *Evaluation framework for 360-degree visual content compression with user view-dependent transmission*, in "ICIP 2019 - IEEE International Conference on Image Processing", Taipei, Taiwan, IEEE, September 2019, pp. 1-5 [*DOI :* 10.1109/ICIP.2019.8803606], https://hal.archives-ouvertes.fr/hal-02202794

[25] N. MAHMOUDIAN BIDGOLI, T. MAUGEY, A. ROUMY. *Intra-coding of 360-degree images on the sphere*, in "PCS 2019 - Picture Coding Symposium", Ningbo, France, November 2019, pp. 1-5, https://hal.archives-ouvertes.fr/hal-02381193

[26] N. MAHMOUDIAN BIDGOLI, T. MAUGEY, A. ROUMY, F. NASIRI, F. PAYAN. *A geometry-aware compression of 3D mesh texture with random access*, in "PCS 2019 - Picture Coding Symposium", Ningbo, China, November 2019, pp. 1-5, https://hal.archives-ouvertes.fr/hal-02380798

[27] T. MAUGEY, L. GUILLO, C. LE CAM. *FTV360: a Multiview 360°Video Dataset with Calibration Parameters*, in "MMSys 2019 - 10th ACM Multimedia Systems Conference", Amherst, United States, ACM Press, June 2019, pp. 291-295 [*DOI :* 10.1145/3304109.3325815], https://hal.inria.fr/hal-02398005

[28] F. NASIRI, N. MAHMOUDIAN BIDGOLI, F. PAYAN, T. MAUGEY. *A Geometry-aware framework for compressing 3D mesh textures*, in "ICASSP 2019 - IEEE International Conference on Acoustics, Speech, and Signal Processing", Brighton, United Kingdom, IEEE, May 2019, pp. 4015-4019 [*DOI :* 10.1109/ICASSP.2019.8683258], https://hal.archives-ouvertes.fr/hal-02010328

[29] C. PETIT, A. ROUMY, G. COLUCCIA, E. MAGLI. *Density evolution of Orthogonal Matching Pursuit*, in "SPARS 2019 - Signal Processing with Adaptive Sparse Structured Representations - Workshop", Toulouse, France, 2019, 1 p. , https://hal.inria.fr/hal-02390560

[30] M. RIZKALLAH, T. MAUGEY, C. GUILLEMOT. *Graph-based Spatio-angular Prediction for Quasi-Lossless Compression of Light Fields*, in "DCC 2019 - Data Compression Conference", Snowbird, United States, IEEE, March 2019, pp. 379-388 [*DOI :* 10.1109/DCC.2019.00046], https://hal.archives-ouvertes.fr/hal-02116369

National Conferences with Proceedings

[31] N. M. BIDGOLI, T. MAUGEY, A. ROUMY. *Compression de contenus 360 et transmission adaptée à la navigation de l'utilisateur*, in "GRETSI 2019 - XXVIIème Colloque francophone de traitement du signal et des images", Lille, France, August 2019, https://hal.inria.fr/hal-02390575

[32] F. NASIRI, N. M. BIDGOLI, F. PAYAN, T. MAUGEY. *Codage d'atlas de textures avec prédiction guidée par la topologie des maillages*, in "GRETSI 2019 - XXVIIème Colloque francophone de traitement du signal et des images", Lille, France, August 2019, pp. 1-4, https://hal.archives-ouvertes.fr/hal-02159557

Scientific Popularization

[33] C. GUILLEMOT, A. ROUMY. *De Fourier à la compression d'images et de vidéos*, in "Interstices", October 2019, https://hal.inria.fr/hal-02336180