

RESEARCH CENTRE  
**Inria Centre**  
at Rennes University

2023  
**ACTIVITY REPORT**

Project-Team  
**SIROCCO**

**Analysis representation, compression and  
communication of visual data**

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

**DOMAIN**

**Perception, Cognition and Interaction**

**THEME**

**Vision, perception and multimedia  
interpretation**

*Inria*

# Contents

<b>Project-Team SIROCCO</b>	<b>1</b>
<b>1 Team members, visitors, external collaborators</b>	<b>2</b>
<b>2 Overall objectives</b>	<b>3</b>
2.1 Introduction	3
2.2 Visual Data Analysis	3
2.3 Signal processing and learning methods for visual data representation and compression	3
2.4 Algorithms for inverse problems in visual data processing	4
2.5 User-centric compression	4
<b>3 Research program</b>	<b>4</b>
3.1 Introduction	4
3.2 Data Dimensionality Reduction	4
3.3 Deep neural networks	5
3.4 Coding theory	6
<b>4 Application domains</b>	<b>6</b>
4.1 Overview	6
4.2 Compression of emerging imaging modalities	7
4.3 Networked visual applications	7
4.4 Editing, post-production and computational photography	7
<b>5 Highlights of the year</b>	<b>8</b>
5.1 Awards	8
<b>6 New software, platforms, open data</b>	<b>8</b>
6.1 New software	8
6.1.1 PnP-A	8
6.1.2 SIUPPA	8
6.1.3 DeepULFCam	8
6.1.4 OSLO: On-the-Sphere Learning for Omnidirectional images	9
6.1.5 JNR-MLF	9
6.1.6 PnP-Reg	10
6.1.7 UnrolledFDL	10
6.1.8 JointFDL-ViewSynthesis	10
<b>7 New results</b>	<b>11</b>
7.1 Visual Data Analysis	11
7.1.1 Light field reconstruction from a focal stack with unrolled optimization in the Fourier domain	11
7.1.2 Meta-surfaces and image reconstruction co-design for 2D and Light Field microscopy	11
7.1.3 Vanishing points aided Hash-Frequency Encoding for Neural Radiance Fields (NeRF) estimation from Sparse 360°Input	12
7.2 Signal processing and learning methods for visual data representation and compression	13
7.2.1 Joint compression, demosaicing and denoising of satellite images	13
7.2.2 Plenoptic Point Clouds modeling and compression	14
7.2.3 Joint Neural and Implicit Representations for Multiple Light Fields	14
7.2.4 Graph Neural Networks on Large Graphs	15
7.3 Algorithms for inverse problems in visual data processing	15
7.3.1 Plug-and-play optimization and learned priors	16
7.3.2 Unrolled deep equilibrium models for multi-tasks inverse problems	16
7.3.3 Low light image restoration	17
7.4 User centric compression	17
7.4.1 Compression with a helper: zero-error compression bounds	17

7.4.2	Coding for Machine: zero-error compression bounds	18
7.4.3	Coding for Machine: learning on Entropy Coded Images with CNN	18
7.4.4	Generative compression of images at extremely low bitrates	18
7.4.5	Image Embedding and User Multi-preference Modeling for Data Collection Sampling	19
7.4.6	A Water-filling Algorithm Maximizing the Volume of Submatrices Above the Rank	20
7.4.7	Towards digital sobriety: why improving the energy efficiency of video streaming is not enough	20
<b>8</b>	<b>Bilateral contracts and grants with industry</b>	<b>21</b>
8.1	Bilateral contracts with industry	21
<b>9</b>	<b>Partnerships and cooperations</b>	<b>22</b>
9.1	European initiatives	22
9.1.1	H2020 projects	22
9.2	National initiatives	23
9.2.1	IA Chair: DeepCIM- Deep learning for computational imaging with emerging image modalities	23
9.2.2	CominLabs Colearn project: Coding for Learning	24
9.2.3	ANR Young researcher grant: MAssive multimedia DATA collection REpurposing (MADARE)	24
<b>10</b>	<b>Dissemination</b>	<b>25</b>
10.1	Promoting scientific activities	25
10.1.1	Scientific events: organisation	25
10.1.2	Scientific events: selection	25
10.1.3	Journal	25
10.1.4	Leadership within the scientific community	25
10.1.5	Research administration	26
10.2	Teaching - Supervision - Juries	26
10.2.1	Teaching	26
10.2.2	Supervision	26
10.2.3	Juries	27
10.3	Popularization	27
10.3.1	Internal or external Inria responsibilities	27
<b>11</b>	<b>Scientific production</b>	<b>27</b>
11.1	Major publications	27
11.2	Publications of the year	28

## **Project-Team SIROCCO**

*Creation of the Project-Team: 2012 January 01*

### **Keywords**

#### **Computer sciences and digital sciences**

A5. – Interaction, multimedia and robotics

A5.3. – Image processing and analysis

A5.4. – Computer vision

A5.9. – Signal processing

A9. – Artificial intelligence

#### **Other research topics and application domains**

B6. – IT and telecom

## 1 Team members, visitors, external collaborators

### Research Scientists

- Christine Guillemot [Team leader, INRIA, Senior Researcher, HDR]
- Thomas Maugey [INRIA, Senior Researcher, from Oct 2023, HDR]
- Aline Roumy [INRIA, Senior Researcher, HDR]

### Post-Doctoral Fellow

- Rémy Leroy [INRIA, Post-Doctoral Fellow, from May 2023]

### PhD Students

- Ipek Anil Atalay Appak [CEA]
- Denis Bacchus [INRIA]
- Tom Bachard [UNIV RENNES I]
- Stephane Belemkoabga [TYNDAL FX, CIFRE, from Oct 2023]
- Tom Bordin [INRIA]
- Nicolas Charpenay [UNIV RENNES]
- Davi Rabbouni De Carvalho Freitas [INRIA]
- Rita Fermanian [INRIA]
- Kai Gu [INRIA]
- Reda Kaafarani [MEDIKIND]
- Brandon Le Bon [INRIA]
- Arthur Lecert [INRIA]
- Yiqun Liu [ATEME (Cifre)]
- Esteban Pesnel [MEDIKIND, from Nov 2023]
- Claude Petit [INRIA]
- Rémi Piau [INRIA]
- Soheib Takhtardeshir [UNIV MID SWEDEN]
- Samuel Willingham [INRIA]

### Technical Staff

- Sébastien Bellenous [INRIA, Engineer]

### Interns and Apprentices

- Antonin Joly [INRIA]

### Administrative Assistant

- Caroline Tanguy [INRIA]

## External Collaborator

- Nicolas Keriven [CNRS]

## 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, High Dynamic Range (HDR) imaging, multiview, plenoptic, light fields, 360° 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
- User-centric compression.

While aiming at generic approaches, some of the solutions developed are applied to practical problems in partnership with industry (InterDigital, Ateame, 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 necessary 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. View synthesis is also needed 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 User-centric compression

The ever-growing volume of image/video traffic motivates the search for new coding solutions suitable for band and energy limited networks but also space and energy limited storage devices. In particular, we investigate compression strategies that are adapted to the users needs and data access requests in order to meet all these transmission and/or storage constraints. Our first goal is to address theoretical issues such as the information theoretical bounds of these compression problems. This includes compression of a database with random access, compression with interactivity, and also data repurposing that takes into account the users needs and user data perception. A second goal is to construct practical coding for all these 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 modeling, 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 modeling 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

**Keywords:** 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 dictionary matrix  $A \in \mathbb{R}^{m \times n}$ , and a data  $\mathbf{b} \in \mathbb{R}^m$  with  $m \ll n$  and  $A$  is of full row rank, one seeks the solution of  $\min\{\|\mathbf{x}\|_0 : \mathbf{A}\mathbf{x} = \mathbf{b}\}$ , where  $\|\mathbf{x}\|_0$  denotes the  $\ell_0$  norm of  $\mathbf{x}$ , i.e. the number of non-zero components in  $\mathbf{x}$ .  $A$  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 nonzero components. In practice, one actually seeks an approximate and thus even sparser solution which satisfies  $\min\{\|\mathbf{x}\|_0 : \|\mathbf{Ax} - \mathbf{b}\|_p \leq \rho\}$ , for some  $\rho \geq 0$ , characterizing an admissible reconstruction error.

The recent theory of *compressed 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  $i$ , and  $\mathbf{A}$  the adjacency matrix. 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 particularly 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

**Keywords:** 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  $\theta$ , that computes a representation  $Y$  from the data  $X$ , and a decoder  $g_d$ , parametrized by  $\phi$ , that gives a reconstruction  $\hat{X}$  of  $X$ . Autoencoders can be used for dimensionality reduction, compression, denoising. When it is used for compression, the representation needs to be quantized, leading to a quantized representation  $\hat{Y} = Q(Y)$ . 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.

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 (GANs) 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, GANs, 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 modeling. 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

**Keywords:** 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 <sup>1</sup> 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 <sup>2</sup> (SW) and Wyner-Ziv <sup>3</sup> (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 side information (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.

<sup>1</sup>T. M. Cover and J. A. Thomas, Elements of Information Theory, Second Edition, July 2006.

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

<sup>3</sup>A. Wyner and J. Ziv, "The rate-distortion function for source coding with 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 (HDR) 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 activities 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 Awards

A. Roumy has received the 2023 IEEE outstanding editorial board award for her activities as associate editor for IEEE Trans. on Image Processing.

The research paper [16] was selected among the Top 3 % of the accepted papers at the 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).

## 6 New software, platforms, open data

### 6.1 New software

#### 6.1.1 PnP-A

**Name:** Plug-and-play algorithms

**Keywords:** Algorithm, Inverse problem, Deep learning, Optimization

**Functional Description:** The software is a framework for solving inverse problems using so-called "plug-and-play" algorithms in which the regularisation is performed with an external method such as a denoising neural network. The framework also includes the possibility to apply preconditioning to the algorithms for better performances. The code is developed in python, based on the pytorch deep learning framework, and is designed in a modular way in order to combine each inverse problem (e.g. image completion, interpolation, demosaicing, denoising, ...) with different algorithms (e.g. ADMM, HQS, gradient descent), different preconditioning methods, and different denoising neural networks.

**Contact:** Christine Guillemot

**Participants:** Mikael Le Pendu, Christine Guillemot

#### 6.1.2 SIUPPA

**Name:** Stochastic Implicit Unrolled Proximal Point Algorithm

**Keywords:** Inverse problem, Optimization, Deep learning

**Functional Description:** This code implements a stochastic implicit unrolled proximal point method, where an optimization problem is defined for each iteration of the unrolled ADMM scheme, with a learned regularizer. The code is developed in python, based on the pytorch deep learning framework. The unrolled proximal gradient method couples an implicit model and a stochastic learning strategy. For each backpropagation step, the weights are updated from the last iteration as in the Jacobian-Free Backpropagation Implicit Networks, but also from a randomly selected set of unrolled iterations. The code includes several applications, namely denoising, super-resolution, deblurring and demosaicing.

**Contact:** Christine Guillemot

**Participants:** Brandon Le Bon, Mikael Le Pendu, Christine Guillemot

#### 6.1.3 DeepULFCam

**Name:** Deep Unrolling for Light Field Compressed Acquisition using Coded Masks

**Keywords:** Light fields, Optimization, Deep learning, Compressive sensing

**Functional Description:** This code implements an algorithm for dense light field reconstruction using a small number of simulated monochromatic projections of the light field. Those simulated projections consist of the composition of: 1) the filtering of a light field by a color coded mask placed between the aperture plane and the sensor plane, performing both angular and spectral multiplexing, 2) a color filtering array performing color multiplexing, 3) a monochromatic sensor. The composition of these processing steps is modeled by a linear projection operator. The light field is then reconstructed by performing the 'unrolling' of an iterative reconstruction algorithm (namely the HQS 'half-quadratic splitting' algorithm, a variant of the ADMM optimization algorithm) using a deep convolutional neural network as proximal operator of the regularizer. The algorithm makes use of the structure of the projection operator to efficiently solve the quadratic data-fidelity minimization sub-problem in closed form. This code is designed to be compatible with Python 3.7 and Tensorflow 2.3. Other required dependencies are: Pillow, PyYAML.

**Contact:** Guillaume Le Guludec

**Participants:** Guillaume Le Guludec, Christine Guillemot

#### 6.1.4 OSLO: On-the-Sphere Learning for Omnidirectional images

**Keywords:** Omnidirectional image, Machine learning, Neural networks

**Functional Description:** This code implements a deep convolutional neural network for omnidirectional images. The approach operates directly on the sphere, without the need to project the data on a 2D image. More specifically, from the sphere pixelization, called healpix, the code implements a convolution operation on the sphere that keeps the orientation (north/south/east/west). This convolution has the same complexity as a classical 2D planar convolution. Moreover, the code implements stride, iterative aggregation, and pixel shuffling in the spherical domain. Finally, image compression is implemented as an application of this on-the-sphere CNN.

**Authors:** Navid Mahmoudian Bidgoli, Thomas Maugey, Aline Roumy

**Contact:** Aline Roumy

#### 6.1.5 JNR-MLF

**Name:** Joint Neural Representation for Multiple Light Fields

**Keywords:** Neural networks, Light fields

**Functional Description:** The code learns a joint implicit neural representation of a collection of light fields. Implicit neural representations are neural networks that represent a signal as a mapping from the space of coordinates to the space of values. For traditional images, it is a mapping from a 2D space to a 3D space. In the case of light field, it is a mapping from a 4D to a 3D space.

The algorithm works by learning a factorisation of the weights and biases of a SIGNET based network. For each layer, a base of matrices is learned that serves as a representation shared between all light fields (a.k.a. scene) of the dataset, together with, for each scene in the dataset, a set of coefficients with respect to this base, which acts as individual representations. The matrices formed by taking the linear combinations of the base matrices with the coefficients corresponding to a given scene serve as the weight matrices of the network. In addition to the set of coefficients, we also learn an individual bias vector for each scene.

The code is therefore composed of two parts: 1)- The representation provider, which takes the index  $i$  of a scene, and outputs a weight matrix and bias vector for each layer. 2)- The synthesis network which, using the weights and biases, computes the values of the pixels by querying the network on the coordinates of all pixels. The network is learned using Adam. At each iteration, a light field index  $i$  is chosen at random, along with a batch of coordinates, and the corresponding values are predicted, using weights and biases from the representation provider and the synthesis network. The model's parameters  $U$ ,  $V$  and (the relevant column of)  $\Sigma$  are then updated by gradient backpropagation using a MSE minimization objective. The code uses TensorFlow 2.7.

**Contact:** Guillaume Le Guludec

**Participants:** Guillaume Le Guludec, Christine Guillemot

#### 6.1.6 PnP-Reg

**Keywords:** Inverse problem, Deep learning, Optimization

**Functional Description:** This is the pytorch implementation of a network directly modeling the gradient of a MAP regularizer. The training of the gradient of the regularizer is done jointly with the training of the corresponding MAP denoiser. This network representing the gradient of an image regularizer can be used in a plug-and-play framework together with gradient-based optimization methods, It can also be used to initialize the regularizer in an unrolled gradient-based optimization framework in which the regularizer can in addition be fine tuned end-to-end for a given task. The code can be used to solve typical inverse problems in imaging such as denoising, deblurring, superresolution or pixel-wise inpainting.

**URL:** <https://github.com/rita442/PnP-ReG>

**Contact:** Rita Fermanian

**Participants:** Rita Fermanian, Christine Guillemot, Mikael Le Pendu

#### 6.1.7 UnrolledFDL

**Name:** Unrolled Optimization of Fourier Disparity Layers for Light field representation and view synthesis

**Keywords:** Light fields, Deep learning, Optimization, View synthesis

**Functional Description:** This code implements a method which reconstruct the light field of a scene from a sparse set of input views taken at a particular viewpoint and at different focal distances, i.e. from a so-called focal stack. By using a focal stack as input, the code allows us to reconstruct the scene light field from images captured by a standard 2D camera. It first performs an unrolled optimization of a Fourier Disparity Layer (FDL) based representation of light fields from the input focal stack. The FDL representation samples the LF in the depth dimension by decomposing the scene as a discrete sum of layers. Once the layers are known, they can be used to produce different viewpoints of the scene while controlling the focus and simulating a camera aperture of arbitrary shape and size.

**URL:** [https://github.com/BrandonLeBon/Unrolled\\_Fourier\\_Disparity\\_Layer](https://github.com/BrandonLeBon/Unrolled_Fourier_Disparity_Layer)

**Contact:** Brandon Le Bon

**Participants:** Christine Guillemot, Brandon Le Bon, Mikael Le Pendu

#### 6.1.8 JointFDL-ViewSynthesis

**Name:** Joint unrolled optimization of FDL light field representation and view synthesis

**Keywords:** Light fields, Deep learning, Optimization, View synthesis

**Functional Description:** This code is a pytorch implementation of a method that allows reconstructing the light field of a scene from a focal stack, i.e. a set of images taken at different focus distances using a single traditional camera. The code estimates an FDL light field representation using an unrolled optimization approach and jointly optimizes a convolutional neural network enabling view synthesis. The joint optimization allows an efficient handling of the occlusions issues.

**URL:** [https://github.com/BrandonLeBon/Joint\\_Optimization\\_FDL](https://github.com/BrandonLeBon/Joint_Optimization_FDL)

**Contact:** Brandon Le Bon

**Participants:** Christine Guillemot, Brandon Le Bon, Mikael Le Pendu

## 7 New results

### 7.1 Visual Data Analysis

**Keywords:** 3D modeling, Light-fields, camera design, 3D point clouds, Plenoptic point clouds.

#### 7.1.1 Light field reconstruction from a focal stack with unrolled optimization in the Fourier domain

**Participants:** Christine Guillemot, Brandon Lebon.

We have addressed the problem of capturing a light field using a single traditional camera, by solving the inverse problem of dense light field reconstruction from a focal stack containing only very few images captured at different focus distances. An end-to-end joint optimization framework has been designed, where a novel unrolled optimization method is jointly optimized with a view synthesis deep neural network. The proposed unrolled optimization method constructs Fourier Disparity Layers (FDL), a compact representation of light fields which samples Lambertian non-occluded scenes in the depth dimension and from which all the light field viewpoints can be computed [24]. Solving the optimization problem in the FDL domain allows us to derive a closed-form expression of the data-fit term of the inverse problem. Furthermore, unrolling the FDL optimization allows us to learn a prior directly in the FDL domain [8]. In order to widen the FDL representation to more complex scenes, a Deep Convolutional Neural Network (DCNN) is trained to synthesize novel views from the optimized FDL. We show that this joint optimization framework reduces occlusion issues of the FDL model (see Figure 1 below), and outperforms recent state-of-the-art methods for light field reconstruction from focal stack measurements.

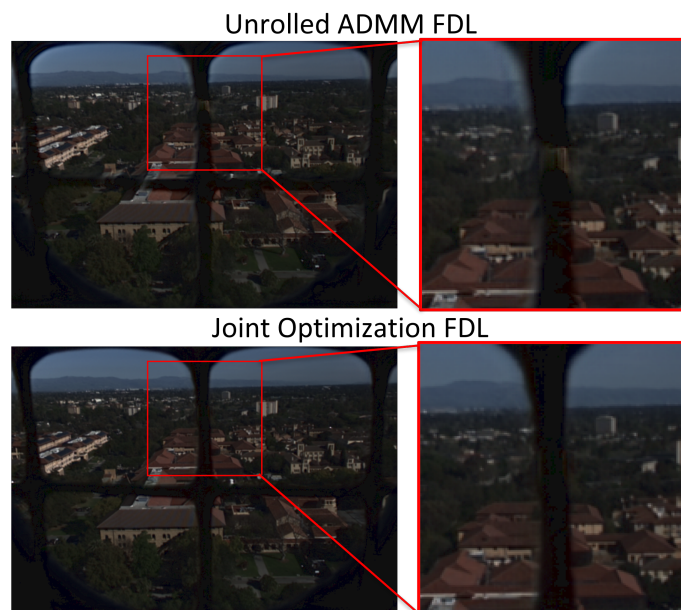


Figure 1: View reconstructed with the unrolled FDL method (top) and with the joint unrolled and view synthesis method (bottom).

#### 7.1.2 Meta-surfaces and image reconstruction co-design for 2D and Light Field microscopy

**Participants:** Anil Ipek Atalay Appak, Christine Guillemot.

Conventional microscopy systems have limited depth of field, which often necessitates depth scanning techniques hindered by light scattering. Various techniques have been developed to address this challenge, but they have limited extended depth of field (EDOF) capabilities. Extended Depth of Field (EDOF) lenses can solve this problem, however there is a trade-off between the spatial resolution and the depth of field of the captured images.

To overcome this challenge, in collaboration with the university of Tampere (Prof. Humeyra Caglayan, Erdem Sahin) in the context of the Plenoptima Marie Curie project, we have addressed this problem of EDOF while preserving a high resolution by designing a pipeline for end-to-end learning and optimization of an optical layer together with a deep learning based image reconstruction method. In our design, instead of phase masks, we consider meta-surface materials. Meta-surfaces are periodic sub-wavelength structures that can change the amplitude, phase, and polarization of the incident wave. The meta-surface parameters are end-to-end optimized to have the highest depth of field with good spatial resolution and high reconstruction quality [7]. Our end-to-end optimization framework allows building a computational EDOF microscope that combines a 4f microscopy optical setup incorporating a learned optics at the Fourier plane and a post-processing deblurring neural network (see the illustration of the principle in the Figure 2 below). Utilizing the end-to-end differentiable model, we have introduced a systematic design methodology for computational EDOF microscopy based on the specific visualization requirements of the sample under examination. We have shown that the metasurface optics provides key advantages for extreme EDOF imaging conditions, where the extended DOF range is well beyond what is demonstrated in state of the art, achieving superior EDOF performance.

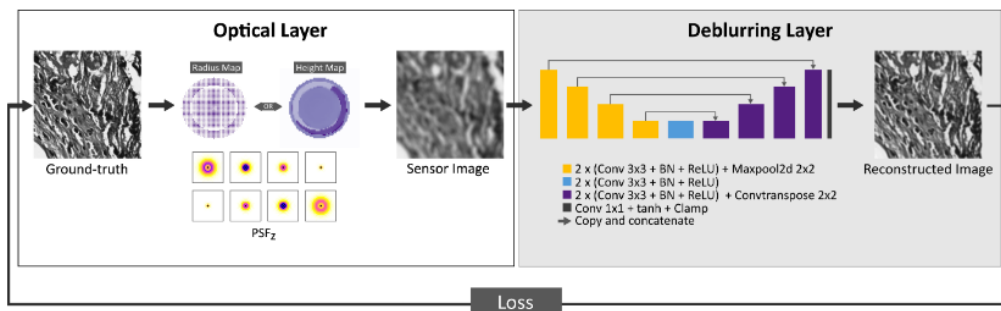


Figure 2: Overall representation of the end-to-end framework for joint optimization of the phase coding mask and the Deblurring CNN. The end-to-end framework consists of an optical layer that simulates image formation for the learned optics, and the deblurring layer employs U-Net architecture to reconstruct in-focus images within the desired DOF range.

### 7.1.3 Vanishing points aided Hash-Frequency Encoding for Neural Radiance Fields (NeRF) estimation from Sparse 360° Input

**Participants:** Kai Gu, Christine Guillemot, Thomas Mauey.

Neural Radiance Fields (NeRF) enable novel view synthesis of 3D scenes when trained with a set of 2D images. One of the key components of NeRF is the input encoding, i.e. mapping the coordinates to higher dimensions to learn high-frequency details, which has been proven to increase the quality. Among various input mappings, hash encoding is gaining increasing attention for its efficiency. However, its performance, when using sparse inputs, is limited. To address this limitation, we have proposed a new input encoding scheme that improves hash-based NeRF for sparse inputs, i.e. few and distant cameras, specifically for 360° view synthesis [21]. We have combined frequency encoding and hash encoding and shown that this combination can increase dramatically the quality of hash-based NeRF for sparse inputs. Additionally, we have explored scene geometry by estimating vanishing points in omnidirectional images (ODI) of indoor and city scenes in order to align frequency encoding with scene

structures (see Figure 3 below). We have shown that our vanishing point-aided scene alignment further improves deterministic and non-deterministic encodings on image regression (view reconstruction and rendering) and view synthesis tasks, where sharper textures and more accurate geometry of scene structures can be reconstructed.

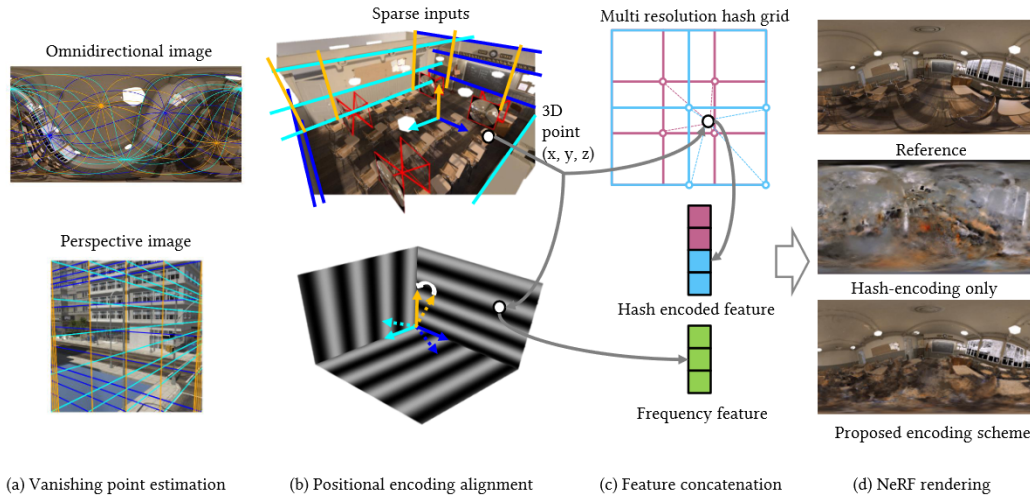


Figure 3: The proposed method improves the sparse input view synthesis of the hash-encoding-based NeRF method. (a) Vanishing point extraction from the perspective or omnidirectional images. (b) Scene-positional encoding alignment according to the estimated vanishing points. (c) Hash-frequency encoding compositing (d) NeRF rendering.

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

**Keywords:** Sparse representation, data dimensionality reduction, compression, scalability, rate-distortion theory.

### 7.2.1 Joint compression, demosaicing and denoising of satellite images

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

Today’s satellite images have a high resolution, thanks to high-performance sensors. This increase in the volume of data to be transmitted to Earth requires efficient compression methods. Designing efficient compression algorithms for satellite images must take into account several constraints. First, (i) it must be well *adapted to the raw data format*. In particular, we have considered the cameras for the Lion satellite constellation with ultra-high spatial resolution at the price of a lower spectral resolution. More precisely, the three spectral bands (RGB) are acquired with a single sensor with an in-built colour filter array. Second, (ii) it should be *adapted to the image statistics*. Indeed, satellite images contain very high-frequency details with small objects spread over very few pixels only. Finally, (iii) the compression must be quasi-lossless to allow accurate on-ground interpretation. In addition, the images are also affected by noise during the acquisition process, which hinders the compression process.

While existing compression solutions mostly assume that the captured raw data has been demosaicked prior to compression, we have designed an end-to-end trainable neural network for joint decoding, demosaicking and denoising satellite images on the ground at the decoder side, while encoding the raw data. We have first introduced a training loss combining a perceptual loss with the classical mean square error, which has been shown to better preserve the high-frequency details present in satellite images.



We then developed a multi-loss balancing strategy which significantly improves the performance of the proposed joint demosaicking-compression solution [16]. We also proposed a novel guidance branch which is used during training, to enforce intermediate features of the demosaicked and decoded images to be close to those that would be obtained with ground truth full colour (non-mosaicked) images during decoding. We have shown that our method achieves better rate-distortion trade-offs compared with current satellite baselines while reducing the overall complexity [42].

We also addressed the limitations of learned image compression neural networks, which have difficulties adapting to certain satellite image characteristics, especially high frequencies that disappear at a high bit-rate in the blur generated in the reconstruction. To answer this problem we described in [15] a joint end-to-end trainable neural network. It is separated into a general compression network and a smaller specialized network. We trained this specialized network to compress the residual part of the image to best preserve the high-frequency details present in the satellite images. The proposed model achieves higher rate-distortion performance than current lossy image compression standards and also manages to retrieve details previously poorly reconstructed.

### 7.2.2 Plenoptic Point Clouds modeling and compression

**Participants:** Davi Freitas, Christine Guillemot.

Modeling 3D scenes from image data to render novel photorealistic views has been a central topic in the field of computer graphics and vision. The most recent advances have come from the use of neural volumetric representations, such as the volumetric Neural Radiance Fields (NeRF). Although the implicit representation of NeRF is able to encapsulate the scene information, its rendering is slow. PlenOctree (Plenoptic Octrees) models have been proposed in order to cope with the rendering speed issue.

A scene is first encoded in a NeRF-like network through training. However, instead of predicting the RGB color of each pixel in each view directly, the network predicts spherical harmonic coefficients. The spherical harmonics are spherical basis functions which allow the encoding of the appearances and preserve view-dependent effects such as specularities. The color is then calculated by summing the weighted spherical harmonic bases evaluated at the corresponding ray direction. The spherical harmonics enable the representation to model view-dependent appearance. To build the PlenOctree model, the NeRF-SH model is densely sampled. This allows extracting the octree parameters (the density  $\sigma$ , and the Spherical Harmonics (SH) coefficients for each octant) and thus generate the octree-based radiance field. Then, this octree model is fine-tuned and compressed, e.g. by controlling the visibility threshold and the number of spherical harmonics. However, decreasing the number of harmonics affects the capability to represent view-dependent effects, such as specularity and non-Lambertian surfaces, and this proves to be the biggest bottleneck to reduce the bit rate. PlenOctree models have a large size, which is an issue for storage and transmission.

We have addressed this weakness by proposing an efficient compression scheme for the plenOctree model which allows maintaining high rendering quality and speed. The proposed approach both improves the training of the plenOctree model with geometry and sparsity constrained regularization, reduces the number of occupied voxels by controlling the visibility threshold, and then efficiently encodes the spherical harmonics. Results over a set of test camera poses show that we can reduce about eight times the bit rates of the encoded models and still obtain a higher quality of the synthesized images when compared to the original PlenOctree models or, alternatively, a reduction of about 50 times while presenting minimal degradation for novel view synthesis [20].

### 7.2.3 Joint Neural and Implicit Representations for Multiple Light Fields

**Participants:** Christine Guillemot, Remy Leroy.

Neural implicit representations have appeared as promising techniques for representing a variety of signals, among which light fields, offering several advantages over traditional grid-based representations,

such as independence to the signal resolution. While it has been shown to be a good representation for a given type of signal, i.e. for a particular light field, we realized that exploiting the features shared between different scenes remains an understudied problem. We have therefore made one step towards this end by developing a method for sharing the representation between thousands of light fields, splitting the representation between a part that is shared between all light fields and a part which varies individually from one light field to another [25].

Assuming that the neural representations of the different scenes lie on a manifold, the algorithm learns a factorisation (in the same vein as a singular value decomposition) of the weights matrices forming the layers of a deep neural network, to learn the basis defining the common manifold. For each layer, a base of matrices  $U$  and  $V$  is thus learned that serves as a representation shared between all light fields (a.k.a. scene) of the dataset, together with, for each scene in the dataset, a set of coefficients  $\sigma$  with respect to this base, which acts as individual representations. The matrices formed by taking the linear combinations of the base matrices with the coefficients corresponding to a given scene serve as the weight matrices of the network. In addition to the set of coefficients, we also learn an individual bias vector for each scene. The architecture is therefore composed of two parts: 1)- a representation provider, which takes the index  $i$  of a scene, and outputs a weight matrix and bias vector for each layer. 2)- The synthesis network which, using the weights and biases, computes the values of the pixels by querying the network on the coordinates of all pixels. The model's parameters  $U$ ,  $V$  and (the relevant column of)  $\sigma$  are then updated by gradient backpropagation using a MSE minimization objective.

We have shown that this joint representation possesses good interpolation properties, and allows for a more light-weight storage of a whole database of light fields, exhibiting a ten-fold reduction in the size of the representation when compared to using a separate representation for each light field. Based on these results, we have then focused on the problem of generalizability of this learned joint representation to new scenes that have not been seen during training.

#### 7.2.4 Graph Neural Networks on Large Graphs

**Participants:** Nicolas Keriven.

Graph Neural Networks (GNN) have become the de-facto deep architectures for most learning problems on data structured as graphs, such as graphs and nodes classification or regression, learning-based graph sampling or graph compression, and so on. Nevertheless, GNN act on different irregular domains represented by one or several graphs that may vary between the training phase and the testing phase, which severely constrain the manner in which they can be built compared to traditional Convolutional Neural Network, as they must respect rules of permutation invariance and equivariance. Moreover, in the absence of traditional convolution, computational complexity quickly becomes an issue, as modern graphs can comprise several millions of nodes. As a result, GNNs encounter many unique challenges limiting their performance, and a better understanding of their theoretical and empirical properties is required to address them.

We have produced a new analysis of the properties of GNNs on *large* graphs by leveraging random graphs models, which are traditionally used to represent graphs sampled from manifolds, or real-world large graphs such as relational databases. In [22], we completely characterize the function space that such GNN can produce in the infinite-node limit, which emphasize GNNs' strength but also limitations. We outline the role of so-called graph Positional Encodings, a recent methodology inspired by Transformer models. In [35], we examine the effects of a crucial part in every GNN architecture on this infinite-node limit, the so-called *aggregation* function. We outline how different functions used in practice yield extremely different rates of convergence to the limit. Finally, in [36], we characterize the generalization properties of simple estimators related to the message-passing process at the core of GNNs on random graphs, a first step toward a more complete characterization.

### 7.3 Algorithms for inverse problems in visual data processing

**Keywords:** Inpainting, denoising, view synthesis, super-resolution.

### 7.3.1 Plug-and-play optimization and learned priors

**Participants:** Rita Fermanian, Christine Guillemot.

Recent optimization methods for inverse problems have been introduced with the goal of combining the advantages of well understood iterative optimization techniques with those of learnable complex image priors. A first category of methods, referred to as "Plug-and-play" methods, has been introduced where a learned network-based prior is plugged in an iterative optimization algorithm. These learnable priors can take several forms, the most common ones being: a projection operator on a learned image subspace, a proximal operator of a regularizer or a denoiser. Plug-and-Play optimization methods have in particular been introduced for solving inverse problems in imaging by plugging a denoiser into the ADMM (Alternating Direction Multiplier Method) optimization algorithm. The denoiser accounts for the regularization and therefore implicitly determines the prior knowledge on the data, hence replacing typical handcrafted priors.

In the context of the AI chair DeepCIM, we have shown that it is possible to train a network directly modeling the gradient of a MAP regularizer while jointly training the corresponding MAP denoiser. We have used this network in gradient-based optimization methods and obtain better results comparing to other generic Plug-and-Play approaches. We have also shown that the regularizer can be used as a pre-trained network for unrolled gradient descent. Lastly, we have shown that the resulting denoiser allows for a better convergence of the Plug-and-Play ADMM.

Although the design of learned denoisers has garnered significant research attention in the context of traditional 2D images, omnidirectional image denoising has received relatively limited attention in the literature. Furthermore, extending processing models and tools designed for 2D images to the sphere presents many challenges due to the inherent distortions and non-uniform pixel distributions associated with spherical representations and their underlying projections. We have addressed the problem of omnidirectional image denoising and studied the advantage of denoising the spherical image directly rather than its mapping [49]. We have introduced a novel network called SphereDRUNet to denoise spherical images using deep learning tools on a spherical sampling. We have shown that denoising directly the sphere using our network gives better performance, compared to denoising the projected equirectangular images with a similarly learned model.

### 7.3.2 Unrolled deep equilibrium models for multi-tasks inverse problems

**Participants:** Christine Guillemot, Samuel Willingham.

One advantage of plug-and-play methods with their learned priors is their genericity in the sense that they can be used for any inverse problem, and do not need to be re-trained for each new problem, in contrast with deep models learned as a regression function for a specific task. However, priors learned independently of the targeted problem may not yield the best solution. Unrolling a fixed number of iterations of optimization algorithms is another way of coupling optimization and deep learning techniques. The learnable network is trained end-to-end within the iterative algorithm so that performing a fixed number of iterations yields optimized results for a given inverse problem. Several optimization algorithms (Iterative Shrinkage Thresholding Algorithm (ISTA), Half S Quadratic Splitting (HQS), and Alternating Direction Method of Multipliers (ADMM)) have been unrolled in the literature, where a learned regularization network is used at each iteration of the optimization algorithm. While usual iterative methods iterate until idempotence, i.e. until the difference between the input and the output is sufficiently small, the number of iterations in unrolled optimization methods is set to a small value, to cope with memory issues.

Another approach to cope with the memory issue of unrolled optimization methods is the Deep equilibrium (DEQ) model which extends unrolled methods to have a theoretically infinite amount of iterations by leveraging fixed-point properties. This allows for simpler back-propagation, which can even be done in a Jacobian-free manner. The resulting back-propagation has a memory-footprint independent

of the number of iterations used. Deep equilibrium models, indeed avoid large memory consumption by back-propagating through the architecture's fixed point. This results in an architecture that is intended to converge and allows for the prior to be more general. However, unrolled methods train a network end-to-end for a specific degradation and necessitate retraining for each specific degradation.

We have leveraged deep equilibrium models to train a plug-and-play style prior that can be used to solve inverse problems for a range of tasks [32]. Deep equilibrium models (DEQs), iterate an unrolled method until convergence and thereby enable end-to-end training on the reconstruction error with simplified back-propagation. We have investigated to what extent a solution for several inverse problems can be found by employing a multi-task DEQ. This MT-DEQ is used to train a prior on the actual estimation error, in contrast to a theoretical noise model used for plug-and-play methods. This has the advantage that the resulting prior is trained for a range of degradations beyond pure Gaussian denoising.

### 7.3.3 Low light image restoration

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

The Retinex theory is particularly well suited to the restoration of nighttime images, as it can be used to decompose an image into the illumination of the scene and its reflectance. It is however not well suited for outdoor scenes where we can have colored light. In addition, it does not take into account the non linearities present in the image acquisition pipeline. We have therefore revisited the retinex model to better take into account the physics of light, i.e. by including a colored illumination, and by taking into account some non linearities of the acquisition pipeline.

Based on this new decomposition model, we have proposed a deep learning-based architecture inspired by the style transfer methods, with two generative adversarial networks (GAN) trained in an unsupervised manner [11]. This deep neural network is composed of two branches, one for each of the components (illumination and reflectance) and includes instance normalization, as done in style transfer, to enforce a better separation between the style (illumination) and the content (reflectance), and thus better handle the component separation problem. The network is trained with additional loss terms corresponding to physical priors, the reflectance being the shared information between the night and daylight image distributions.

## 7.4 User centric compression

**Keywords:** Information theory, interactive communication, coding for machines, generative compression, database sampling.

### 7.4.1 Compression with a helper: zero-error compression bounds

**Participants:** Nicolas Charpenay, Aline Roumy.

Zero-error source coding is at the core of practical compression algorithms. Here, zero error refers to the case where the data is decompressed without any error for any block-length. This must be differentiated from the vanishing error setup, where the probability of error of the decompressed data decreases to zero as the length of the processed data tends to infinity. In point-to point communication (one sender/one receiver), zero-error and vanishing-error information theoretic compression bounds coincides. This is however not the case in networks, with possibly many senders and/or receivers. For instance, in the source coding problem with a helper at the decoder, the compression bounds under both hypotheses do not coincide. More precisely, there exist examples, where both bounds differ. But more importantly, the zero-error compression characterization for the general case is still an open problem. This is due to the fact that the compression bounds are given by the so-called complementary graph entropy, which requires to compute the coloring of an infinite product of graphs. Therefore, this

complementary graph entropy has no single-letter formula, except for some particular cases such as: perfect graphs, the pentagon graph, and their disjoint union. In [18], we derived a structural result that equates the complementary graph entropies of AND product and disjoint unions (up to a multiplicative constant), which has two consequences. First, we prove that the cases where the complementary graph entropies of these two compositions (AND product and disjoint union) can be linearized coincide. Second, we determine the complementary graph entropies in cases where it was unknown: AND products of perfect graphs; product of perfect graph and the pentagon graph, which are not perfect in general.

#### 7.4.2 Coding for Machine: zero-error compression bounds

**Participants:** Nicolas Charpenay, Aline Roumy.

Due to the ever-growing amount of visual data produced and stored, there are now evidences that these data will not only be viewed by humans but also processed by machines. In this context, practical implementations aim to provide better compression efficiency when the primary consumers are machines that perform certain computer vision tasks. One difficulty arises from the fact that these tasks are not known upon compression. This problem has been studied in Information theory and is called the coding for computing problem. However, no single letter formula exists for this problem, due to the fact that the optimal coding performance depends on an infinite product of characteristic graphs, that represents the joint distribution of the data and the tasks. In [19], we introduced a generalization of this problem, where the encoder has some partial information about the request, for instance the type of the request that will be made at the decoder. We proposed a condition, which allows to have a single letter characterization of the optimal compression performance.

#### 7.4.3 Coding for Machine: learning on Entropy Coded Images with CNN

**Participants:** Rémi Piau, Thomas Maugey, Aline Roumy.

Learning on entropy coded data has many benefits. First, it avoids decoding, but also it allows one to process compact data. However, this type of learning has been overlooked due to the chaos introduced by entropy coding functions.

In [29], we proposed an empirical study to see whether learning with convolutional neural networks (CNNs) on entropy coded data is possible. First, we define spatial and semantic closeness, two key properties that we experimentally show to be necessary to guarantee the efficiency of the convolution. Then, we show that these properties are not satisfied by the data processed by an entropy coder. Despite this, our experimental results show that learning in such difficult conditions is still possible, and that the performances are far from a random guess. These results have been obtained thanks to the construction of CNN architectures designed for 1D data (one based on VGG, the other on ResNet). Finally, we propose some experiments that explain why CNN are still performing reasonably well on entropy coded data.

In [30], we introduced a new metric that measures the chaos in the data representation. This measure is easy to compute as it depends on the encoded data only. Moreover, for a family of entropy coders, this measure allows one to predict the accuracy of the learning algorithm that process entropy coded data.

#### 7.4.4 Generative compression of images at extremely low bitrates

**Participants:** Tom Bordin, Thomas Maugey.

An image is traditionally compressed with the aim of minimizing the error made in the reconstruction. The Mean Squared Error (MSE) naturally comes as a simple and efficient criterion to evaluate the fidelity of the output in terms of distortion. This metric remains widely used to evaluate methods compressing

images ranging from high to low bitrates. But what happens when compression is pushed to the extreme ( $\sim 0.02$  bpp)? While excessive compression of sensitive data such as movies or vacations photos is not desirable, a drastic reduction of storage could be welcomed for the so-called "cold data". This massive amount of data that is stored but almost never accessed is estimated to represent 60% of today's storage while projected to become 80% by 2025. In that case, compression at extremely low bitrates could be an alternative to deleting potentially useful data or keeping a huge amount of data that might not be used. However, when targeting such bitrates, the relevance of the MSE as an evaluation criterion drops. Indeed, there exists a tradeoff between the perceived quality of the output and the fidelity in terms of pixels. This is called the perception-distortion tradeoff. Moreover, it is shown that decreasing the bitrate exacerbates the opposing goals of the two metrics. Optimization of the MSE then leads to the apparition of numerous compression artifacts, which makes the use of such bitrates really unattractive.

In that context, we proposed in [17] a framework for image compression in which the fidelity criterion is replaced by a semantic and quality preservation objective. Encoding the image thus becomes a simple extraction of semantic, enabling to reach drastic compression ratios. The decoding side is handled by a generative model relying on the diffusion process for the reconstruction of images. We first propose to describe the semantic using low resolution segmentation maps as guide. We further improve the generation, introducing color maps guidance without retraining the generative decoder. We show that it is possible to produce images of high visual quality with preserved semantic at extremely low bitrates when compared with classical codecs (see Figure below).

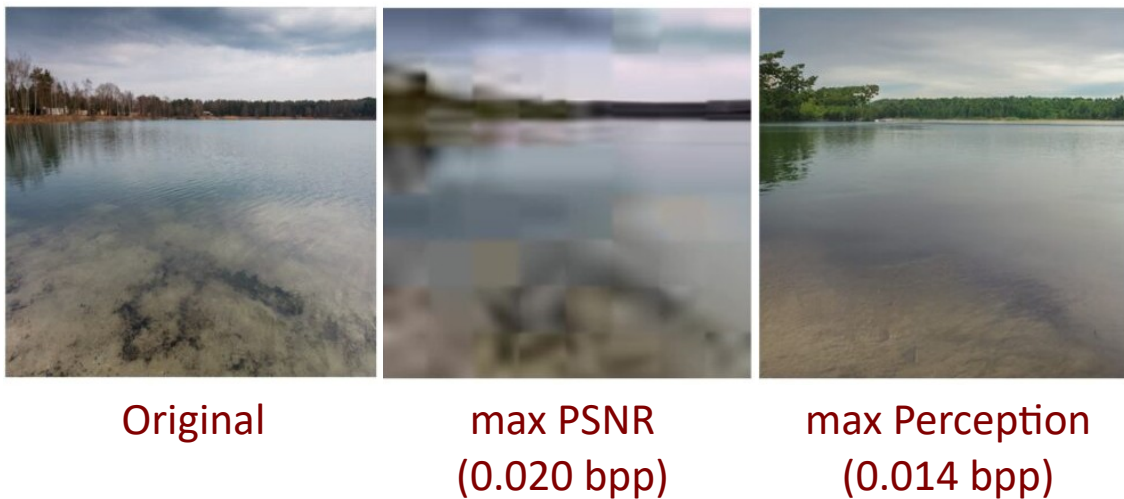


Figure 4: Illustration of the perception-distortion tradeoff. MSE-based (middle) and perception-based (right) compression methods side-by-side at extremely low bitrates.

#### 7.4.5 Image Embedding and User Multi-preference Modeling for Data Collection Sampling

**Participants:** Thomas Maugey.

Image sampling based on user perception has recently increased interest given the unstoppable increase of stored datasets. For example, let's think of the large-scale image collections associated with Instagram and other social media accounts, online retailers, or your own personal photo Gallery saved on the cloud. Storing such large-scale datasets has become a big burden, from an economical as well as sustainability perspective. Traditional image data sampling methods can be understood as a way to alleviate such burden, by retaining only pictures from a dataset with the ideal goal of providing a summary (in the sense of a brief description) of the database. The challenging aspect is however to understand which is the best set of images to select (not to delete), especially given the user's preferences. This is highly challenging as the user will perceive the deleted images as lost information.

In [14], we proposed an end-to-end user-centric sampling method aimed at selecting the images from an image collection that are able to maximize the information perceived by a given user. As main contributions, we first introduce novel metrics that assess the amount of *perceived* information retained by the user when experiencing a set of images. Given the actual information present in a set of images, which is the volume spanned by the set in the corresponding latent space, we show how to take into account the user's preferences in such a volume calculation to build a user-centric metric for the perceived information. Finally, we propose a sampling strategy seeking the minimum set of images that maximize the information perceived by a given user. Experiments using the coco dataset show the ability of the proposed approach to accurately integrate user preference while keeping a reasonable diversity in the sampled image set.

#### 7.4.6 A Water-filling Algorithm Maximizing the Volume of Submatrices Above the Rank

**Participants:** Claude Petit, Aline Roumy, Thomas Maugey.

In [28], we proposed an algorithm to extract, from a given rectangular matrix, a submatrix with maximum volume, whose number of extracted columns is greater than the initial matrix rank. This problem arises in compression and summarization of databases, recommender systems, learning, numerical analysis or applied linear algebra. We use a continuous relaxation of the maximum volume matrix extraction problem, which admits a simple and closed form solution: the nonzero singular values of the extracted matrix must be equal. The proposed algorithm extracts matrices with singular values, which are close to be equal. It is inspired by a water-filling technique, traditionally dedicated to equalization strategies in communication channels. Simulations show that the proposed algorithm performs better than sampling methods based on determinantal point processes (DPPs) and achieves similar performance as the best known algorithm, but with a lower complexity.

#### 7.4.7 Towards digital sobriety: why improving the energy efficiency of video streaming is not enough

**Participants:** Thomas Maugey.

In 2018, online video streaming constituted 60% of the global data flow, and, by itself, generated 1% of the global emissions (as much as a country like Spain emits). And this quantity explodes: video traffic is estimated to grow by 79% between 2021 and 2027. At the same time, the sixth report of the Intergovernmental Panel on Climate Change (IPCC) states that if we want to keep the global warming under 1.5°C (Paris agreement), one should target a global emissions decrease of 50% when compared to those of 2019. This corresponds to a decrease of 7.6% per year. They also state that *this is not the path that is currently taken*. Hence, every part of our society must urgently aim sobriety. This is for example the case of online video streaming. So the question is simple: what are the solutions that should be envisaged to halve by 2030 the GHG emissions due to video streaming ?

As these emissions are correlated (not necessarily proportional) to the video data flow, one could rely on the progress of video compression algorithms to decrease the size of every individual video. For example, the last H.266/VVC standard reaches 50% of gain when compared with the previous codec H.265/HEVC published 7 years before. Even though similar gains can be expected for the future codec, this cannot, by itself, enable a drastic emissions reduction, because of at least three reasons. First, the size reduction is achieved on individual videos and not on the global flow. In other words, video compression does not fight against the number of videos that is created or consumed. Yet, this amount of video streamed will explode in the coming years. Second, compression gains are usually not exploited to decrease the required bandwidth but rather to increase the video resolution or to create new video usages (IoT, Virtual Reality, etc.). Last, compression gains are achieved only at the expense of huge algorithm complexity. Encoding a video becomes more and more complex and thus requires, for each codec, more energy. Another cause of video streaming gas emissions is precisely the complexity of the operations all

along the transmission chain (including the video coding algorithms as mentioned just above). Intensive research efforts have been made to decrease this complexity and thus the consumed energy.

Even though these research results are crucial and make the energy expenditure more optimized, we demonstrate in [27] that this is insufficient to achieve the goal set by the IPCC. We build a simple model derived from the Kaya identity. This model enables to compare the order of magnitude of the energy reduction on one side and the digital affluence on the other side. Results show that even drastic energy reduction cannot cope with the forecasted video data flow explosion. Said differently, to achieve Paris agreement objectives (half emissions in 2030 when compared to 2019), the academic, industrial and politic actors must also consider laws and regulation to limit the video data consumption.

## 8 Bilateral contracts and grants with industry

### 8.1 Bilateral contracts with industry

#### CIFRE contract with Ateame on neural networks for video compression

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

- Title : Neural networks for video compression of reduced complexity
- Partners : Ateame (T. Guionnet, M. Abdoli), Inria-Rennes.
- Funding : Ateame, ANRT.
- Period : Aug.2020-Jul.2023.

The goal of this Cifre contract is to investigate deep learning architectures for the inference of coding modes in video compression algorithms with the ultimate goal of reducing the encoder complexity. The first step addresses the problem of Intra coding modes and quad-tree partitioning inference. The next step will consider Inter coding modes taking into account motion and temporal information.

#### CIFRE contract with TyndallFX on Radiance fields representation for dynamic scene reconstruction

**Participants:** Stephane Belemkoabga, Christine Guillemot, Thomas Maugey.

- Title : Radiance fields representation for dynamic scene reconstruction
- Partners : TyndallFX (M. Hudon, R. Mallart), Inria-Rennes.
- Funding : TyndallFX, ANRT.
- Period : Oct.2023-Sept.2026.

The goal of this project is to design novel methods for modeling and compact representation of radiance fields for scene reconstruction and view synthesis. The problems that are addressed are those of fast and efficient estimation of the camera pose parameters and of the 3D model of the scene based on Gaussian splatting, and as well as the one of tracking and modeling the deformation of the model due to the global camera motion and to the motion of the different objects in the scene.



### Contract LITCHIE with Airbus on deep learning for satellite imaging

**Participants:** Denis Bacchus, Christine Guillemot, Arthur Lecert, Aline Roumy.

- Title : Deep learning methods for low light vision
- Partners : Airbus (R. Fraisse), Inria-Rennes.
- Funding : BPI.
- Period : Sept.2020-Aug.2023.

The goal of this contract is to investigate deep learning methods for low light vision with satellite imaging. The SIROCCO team focuses on two complementary problems: compression of low light images and restoration under conditions of low illumination, and hazing. The problem of low light image enhancement implies handling various factors simultaneously including brightness, contrast, artifacts and noise. We investigate solutions coupling the retinex theory, assuming that observed images can be decomposed into reflectance and illumination, with machine learning methods. We address the compression problem taking into account the processing tasks considered on the ground such as the restoration task, leading to an end-to-end optimization approach.

### Cifre contract with MediaKind on Multiple profile encoding optimization

**Participants:** Reda Kaafarani, Thomas Maugey, Aline Roumy.

- Title : Multiple profile encoding optimization
- Partners : MediaKind, Inria-Rennes.
- Funding : MediaKind, ANRT.
- Period : April 2021-April 2024.

The goal of this Cifre contract is to optimize a streaming solution taking into the whole process, namely the encoding, the long-term and the short term storages (in particular for replay, taking into the popularity of the videos), the multiple copies of a video (to adapt to both the resolution and the bandwidth of the user), and the transmissions (between all entities: encoder, back-end and front-end server, and the user). This optimization will be with several objectives as well. In particular, the goals will be to maximize the user experience but also to save energy and/or the deployment cost of a streaming solution.

## 9 Partnerships and cooperations

### 9.1 European initiatives

#### 9.1.1 H2020 projects

**PLENOPTIMA: Marie Skłodowska-Curie Innovative Training Network** [PLENOPTIMA project on cordis.europa.eu](https://cordis.europa.eu/plenoptima)

**Participants:** Anil Atalay Appak, Davi Freitas, Christine Guillemot, Kai Gu, Thomas Maugey, Soheib Takhtardeshir, Samuel Willingham.

**Title:** Plenoptic Imaging

**Duration:** From January 1, 2021 to December 31, 2024

**Partners:**

- Institut National de Recherche en Informatique et Automatique (INRIA), France
- Mittuniversitetet (MIUN), Sweden
- Technische Universität Berlin (TUB), Germany
- Tampereen Korkeakoulu (TAMPERE UNIVERSITY), Finland
- Institute of Optical Material and Technologies, Bulgarian Academy of Sciences (IOMT), Bulgaria

**Inria contact:** Christine Guillemot

**Coordinator:**

**Summary:** Plenoptic Imaging aims at studying the phenomena of light field formation, propagation, sensing and perception along with the computational methods for extracting, processing and rendering the visual information.

The PLENOPTIMA ultimate project goal is to establish new cross-sectorial, international, multi-university sustainable doctoral degree programmes in the area of plenoptic imaging and to train the first fifteen future researchers and creative professionals within these programmes for the benefit of a variety of application sectors. PLENOPTIMA develops a cross-disciplinary approach to imaging, which includes the physics of light, new optical materials and sensing principles, signal processing methods, new computing architectures, and vision science modelling. With this aim, PLENOPTIMA joins five of the strongest research groups in nanophotonics, imaging and machine learning in Europe with twelve innovative companies, research institutes and a pre-competitive business ecosystem developing and marketing plenoptic imaging devices and services.

PLENOPTIMA advances the plenoptic imaging theory to set the foundations for developing future imaging systems that handle visual information in fundamentally new ways, augmenting the human perceptual, creative, and cognitive capabilities. More specifically, it develops 1) Full computational plenoptic imaging acquisition systems; 2) Pioneering models and methods for plenoptic data processing, with a focus on dimensionality reduction, compression, and inverse problems; 3) Efficient rendering and interactive visualization on immersive displays reproducing all physiological visual depth cues and enabling realistic interaction.

All ESRs are registered in Joint/Double degree doctoral programmes at academic institutions in Bulgaria, Finland, France, Germany and Sweden.

## 9.2 National initiatives

### 9.2.1 IA Chair: DeepCIM- Deep learning for computational imaging with emerging image modalities

**Participants:** Rita Fermanian, Christine Guillemot, Brandon Lebon, Remy Leroy.

**Funding :** ANR (Agence Nationale de la Recherche).

**Period :** Sept. 2020 - Aug. 2024.

**Inria contact :** Christine Guillemot

**Summary:** The project aims at leveraging recent advances in three fields: image processing, computer vision and machine (deep) learning. It will focus on the design of models and algorithms for data dimensionality reduction and inverse problems with emerging image modalities. The first research challenge will concern the design of learning methods for data representation and dimensionality reduction. These methods encompass the learning of sparse and low rank models, of signal priors or representations in latent spaces of reduced dimensions. This also includes the learning of efficient and, if possible, lightweight architectures for data recovery from the representations of reduced dimension. Modeling joint distributions of pixels constituting a natural image is also a fundamental requirement for a variety of processing tasks. This is one of the major challenges in generative image modeling, field conquered in recent years by deep learning. Based on the above models, our goal is also to develop algorithms for solving a number of inverse problems with novel imaging modalities. Solving inverse problems to retrieve a good representation of the scene from the captured data requires prior knowledge on the structure of the image space. Deep learning techniques designed to learn signal priors, can be used as regularization models.

### 9.2.2 CominLabs Colearn project: Coding for Learning

**Participants:** Thomas Maugey, Rémi Piau, Aline Roumy.

- Partners: Inria-Rennes (Sirocco team); LabSTICC, IMT Atlantique, (team Code and SD); IETR, INSA Rennes (Syscom team).
- Funding: Labex CominLabs.
- Period: Sept. 2021 - Dec. 2024.

The amount of data available online is growing so fast that it is essential to rely on advanced Machine Learning techniques so as to automatically analyze, sort, and organize the content uploaded by e.g. sensors or users. The conventional data transmission framework assumes that the data should be completely reconstructed, even with some distortions, by the server. Instead, this project aims to develop a novel communication framework in which the server may also apply a learning task over the coded data. The project will therefore develop an Information Theoretic analysis so as to understand the fundamental limits of such systems, and develop novel coding techniques allowing for both learning and data reconstruction from the coded data.

### 9.2.3 ANR Young researcher grant: MAssive multimedia DAta collection REpurposing (MADARE)

**Participants:** Tom Bordin, Thomas Maugey.

- Funding: ANR (Agence Nationale de la Recherche).
- Period: Apr. 2022 - Oct. 2025.

Compression algorithms are nowadays overwhelmed by the tsunami of visual data created everyday. Despite a growing efficiency, they are always constrained to minimize the compression error, computed in the pixel domain. The Data Repurposing framework, proposed in the MADARE project, will tear down this barrier, by allowing the compression algorithm to “reinvent” part of the data at the decoding phase, and thus saving a lot of bit-rate by not coding it. Concretely, a data collection is only encoded to a compact description that is used to guarantee that the regenerated content is semantically coherent with the initial one.

In practice, it opens several research directions: how to organise the latent space (in which the coded descriptions lie) such that the information is efficiently and intelligibly represented ? How to regenerate a

synthesized content from this compact description (based for example on guided diffusion algorithms) ?  
Finally, how to extend this idea to video ?

By revisiting the compression problem, the MADARE project aims gigantic compression ratios enabling, among other benefits, to reduce the impact of exploding data creation on the cloud servers' energy consumption.

## 10 Dissemination

**Participants:** Christine Guillemot, Nicolas Keriven, Thomas Maugey, Aline Roumy.

---

### 10.1 Promoting scientific activities

#### 10.1.1 Scientific events: organisation

##### General chair, scientific chair

- C. Guillemot has organized the Plenoptima training school on plenoptic processing and compression as well as the Plenoptima workshop on innovation and project management (26-30 June 2023).

#### 10.1.2 Scientific events: selection

##### Member of the conference program committees

- A. Roumy was a member of the technical program committee of the (Conference on Computer Vision and Pattern Recognition) CVPR 2023 workshop on New Trends in Image Restoration and Enhancement (NTIRE).
- A. Roumy was a member of the technical program committee of the 2023 National Signal Processing workshop (colloque GRETSI).

#### 10.1.3 Journal

##### Member of the editorial boards

- C. Guillemot has been associate editor of the International Journal on Mathematical Imaging and Vision (2016-2023).
- T. Maugey is associate editor of the EURASIP Journal on Advances in Signal Processing.
- T. Maugey is associate editor of the IEEE Signal Processing Letter.
- A. Roumy was Associate Editor of the IEEE Transactions on Image Processing until Nov. 2023.
- A. Roumy is Senior Associate Editor of the IEEE Transactions on Image Processing since Nov. 2023.

#### 10.1.4 Leadership within the scientific community

- A. Roumy is member of the IEEE IVMS technical committee.
- A. Roumy is a member of the Executive board of the National Research group in Image and Signal Processing (GRETSI).
- T. Maugey is member of the EURASIP Technical Area Committee (Visual Information Processing) since Jan 2023 for a four-year term.

### 10.1.5 Research administration

- C. Guillemot was a member of the "Bureau du Comité des Projets" (until Dec. 2023)
- C. Guillemot is a member of the research commission of the University of Rennes, as Inria representative
- A. Roumy has been a member of Inria evaluation committee (until June 2023)
- A. Roumy is a member of the research commission and of the academic board of the University of Rennes 2, as Inria representative
- A. Roumy has been a member of the jury for the recruitment of Inria Junior researcher (CRCN/ISFP) in Lyon, and of the jury of the recruitment of Inria senior researchers
- A. Roumy has been a member of the Inria Delegation committee of the Inria center at Rennes University
- T. Maugey has been a member of the jury for the recruitment of Inria Junior researcher (CRCN/ISFP) in Rennes

## 10.2 Teaching - Supervision - Juries

### 10.2.1 Teaching

- T. Maugey has given a course on Graph Image Processing, 10 hours, M2 SiVOS, Univ. of Rennes 1, France.
- T. Maugey has given a course on Representation, editing and perception of digital images, 21 hours, M2 SIF, Univ. of Rennes 1, France.
- A. Roumy has given a Engineering degree course on Image and Video compression, 10 hours, University Rennes 1, ESIR, France.
- N. Keriven has given a course on Machine Learning, 24 hours, ENSTA Paris, France.
- N. Keriven has given a course on Signal Processing, 12 hours, ENS Rennes, France.
- N. Keriven has given a course on Sparsity and COMpressed Sensing, 12 hours, INSA Rennes, France.

### 10.2.2 Supervision

- C. Guillemot and A. Roumy have been co-supervising the PhD thesis of D. Bacchus and A. Lecert in the context of the Litchie project with Airbus, and the PhD thesis of Y. liu in the context of the Cifre contract with Ateame.
- C. Guillemot has been supervising the PhD thesis of B. Lebon and R. Fermanian in the context of the AI chair DeepCIM funded by the ANR.
- C. Guillemot and T. Maugey are co-supervising the PhD thesis of S. Belemkoabga in the context of the Cifre contract with TyndallFX.
- C. Guillemot is co-supervising the double degree PhD thesis of S. Willingham (together with M. Sojstrom from MidSeden University), of S. Takhtardeshir (together with M. Sojstrom and R. Olsson from MidSeden University), of Davi Freitas (together with I. Tabus from Tampere Univ.), of K. Gu (together with T. Maugey from Inria and S. Knorr from Technical Univ. Berlin), and of A. Atalay Appak (together with H. Caglayan and E. Sahim from Tampere University). These PhDs are double degrees in the framework of the European Marie Curie project Plenoptima.
- T. Maugey is supervising the PhD thesis of Tom Bachard and Tom Bordin (University of Rennes),

- T. Maugey and A. Roumy are co-supervising the PhD thesis of Reda Kaafarani, and Esteban Pesnel in the context of the Cifre contract with MediaKind, and the PhD thesis of Remi Piau in the context of the CominLabs CoLearn contract.
- A. Roumy has been supervising the PhD thesis of N. Charpenay (ENS-Paris Saclay doctoral grant), together with M. Le Treust (Ermine team, Irisa/Inria Center of the University of Rennes).
- A. Roumy is supervising the PhD thesis of C. Petit in collaboration with N. Keriven and T. Maugey.
- N. Keriven has supervised the PhD thesis of H. Ghanem with S. Vaiter (CNRS, LJAD, Nice) and J. Salmon (PU Univ. Montpellier).
- N. Keriven is supervising the PhD thesis of M. Cordonnier with S. Vaiter (CNRS, LJAD, Nice) and N. Tremblay (CNRS, GIPSA-lab, Grenoble).
- N. Keriven is supervising the PhD thesis of M. Gjorgjevski with S. Barthelmé (CNRS, GIPSA-lab, Grenoble) and Y. de Castro (Pr Centrale Lyon).
- N. Keriven has supervised the internship of A. Joly.

### 10.2.3 Juries

A. Roumy was rapporteur of the PhDs of

- M. Milovanovic, Institut Polytechnique de Paris, July 2023.
- Y. Sun, Univ. Paris-Saclay, Dec. 2023.

and chair of the PhD jury of

- T. Dardouri, Univ. Paris-Sorbonne, Feb. 2023.

T. Maugey was member of the PhD jury of S. Lingadahalli Ravi, IETR, Rennes, Dec. 2023.

## 10.3 Popularization

### 10.3.1 Internal or external Inria responsibilities

- T. Maugey is Scientific mediation officer in the scientific mediation team of Inria centre at Rennes University.

## 11 Scientific production

### 11.1 Major publications

- [1] R. Farrugia and 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://doi.org/10.1109/TPAMI.2019.2893666). URL: <https://hal.archives-ouvertes.fr/hal-01984843>.
- [2] X. Jiang, M. Le Pendu, R. A. Farrugia and 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://doi.org/10.1109/JSTSP.2017.2747078). URL: <https://hal.archives-ouvertes.fr/hal-01591349>.
- [3] M. Le Pendu, C. Guillemot and 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://doi.org/10.1109/TIP.2019.2922099). URL: <https://hal.archives-ouvertes.fr/hal-02130555>.
- [4] N. Mahmoudian Bidgoli, T. Maugey and A. Roumy. ‘Fine granularity access in interactive compression of 360-degree images based on rate-adaptive channel codes’. In: *IEEE Transactions on Multimedia* (2020). DOI: [10.1109/TMM.2020.3017890](https://doi.org/10.1109/TMM.2020.3017890). URL: <https://hal.inria.fr/hal-02946795>.

- [5] M. Q. Pham, A. Roumy, T. Maugey, E. Dupraz and M. Kieffer. ‘Optimal Reference Selection for Random Access in Predictive Coding Schemes’. In: *IEEE Transactions on Communications* 68.9 (2020), pp. 5819–5833. DOI: [10.1109/TCOMM.2020.3002937](https://doi.org/10.1109/TCOMM.2020.3002937). URL: <https://hal-imt-atlantique.archives-ouvertes.fr/hal-02925113>.
- [6] M. Rizkallah, X. Su, T. Maugey and C. Guillemot. ‘Geometry-Aware Graph Transforms for Light Field Compact Representation’. In: *IEEE Transactions on Image Processing* (Aug. 2019), pp. 1–15. DOI: [10.1109/TIP.2019.2928873](https://doi.org/10.1109/TIP.2019.2928873). URL: <https://hal.archives-ouvertes.fr/hal-02199839>.

## 11.2 Publications of the year

### International journals

- [7] I. A. Atalay Appak, E. Sahin, C. Guillemot and H. Caglayan. ‘Learning flat optics for extended depth of field microscopy imaging’. In: *Nanophotonics* (2023), pp. 1–10. DOI: [10.1515/nanoph-2023-0321](https://doi.org/10.1515/nanoph-2023-0321). URL: <https://hal.science/hal-04184173>.
- [8] B. L. Bon, M. Le Pendu and C. Guillemot. ‘Joint Fourier Disparity Layers unrolling with learned view synthesis for light field reconstruction from few-shots focal stacks’. In: *IEEE Access* 11 (1st Nov. 2023), pp. 123350–123360. DOI: [10.1109/access.2023.3329328](https://doi.org/10.1109/access.2023.3329328). URL: <https://hal.science/hal-04292432>.
- [9] H. Ghanem, J. Salmon, N. Keriven and S. Vaiter. ‘Supervised learning of analysis-sparsity priors with automatic differentiation’. In: *IEEE Signal Processing Letters* 30 (2023), pp. 339–343. DOI: [10.1109/LSP.2023.3244511](https://doi.org/10.1109/LSP.2023.3244511). URL: <https://hal.science/hal-03518852>.
- [10] N. Keriven. ‘Entropic Optimal Transport in Random Graphs’. In: *SIAM Journal on Mathematics of Data Science* 5.4 (2023), pp. 1028–1050. DOI: [10.1137/22M1518281](https://doi.org/10.1137/22M1518281). URL: <https://hal.science/hal-03576738>.
- [11] A. Lecert, A. Roumy, R. Fraise and C. Guillemot. ‘GAN architecture leveraging a Retinex model with colored illumination for low-light image restoration’. In: *IEEE Access* 11 (2023), pp. 84574–84588. DOI: [10.1109/ACCESS.2023.3301614](https://doi.org/10.1109/ACCESS.2023.3301614). URL: <https://hal.science/hal-04184198>.
- [12] F. G. Lohesara, D. R. Freitas, C. Guillemot, K. Eguiazarian and S. Knorr. ‘HEADSET: Human Emotion Awareness under Partial Occlusions Multimodal DataSET’. In: *IEEE Transactions on Visualization and Computer Graphics* (2023), pp. 1–11. URL: <https://hal.science/hal-04198563>.
- [13] C. Poon, N. Keriven and G. Peyré. ‘The geometry of off-the-grid compressed sensing’. In: *Foundations of Computational Mathematics* 23 (Feb. 2023), pp. 241–327. DOI: [10.1007/s10208-021-09545-5](https://doi.org/10.1007/s10208-021-09545-5). URL: <https://hal.science/hal-02484957>.
- [14] A. J. Tom, L. Toni and T. Maugey. ‘Image embedding and user multi-preference modeling for data collection sampling’. In: *EURASIP Journal on Advances in Signal Processing* 2023.1 (Oct. 2023), pp. 1–16. DOI: [10.1186/s13634-023-01069-0](https://doi.org/10.1186/s13634-023-01069-0). URL: <https://inria.hal.science/hal-04255807>.

### International peer-reviewed conferences

- [15] P. Bacchus, R. Fraise, C. Guillemot and A. Roumy. ‘Filtered Residual Compression for Satellite Images’. In: *IGARSS 2023 - International Geoscience and Remote Sensing Symposium*. Pasadena, CA, United States, 16th July 2023, pp. 1–1. URL: <https://hal.science/hal-04125811>.
- [16] P. Bacchus, R. Fraise, A. Roumy and C. Guillemot. ‘Joint Compression and Demosaicking for Satellite Images’. In: *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. ICASSP 2023 - International Conference on Acoustics, Speech, and Signal Processing. Rhodes (Grèce), Greece: IEEE, 4th June 2023, pp. 1–5. DOI: [10.1109/ICASSP49357.2023.10096011](https://doi.org/10.1109/ICASSP49357.2023.10096011). URL: <https://hal.science/hal-04056997>.
- [17] T. Bordin and T. Maugey. ‘Semantic based generative compression of images for extremely low bitrates’. In: *MMSP 2023 - IEEE 25th International Workshop on MultiMedia Signal Processing*. Poitiers, France: IEEE, 2023, pp. 1–6. URL: <https://hal.science/hal-04231421>.

- [18] N. Charpenay, M. Le Treust and A. Roumy. ‘Complementary Graph Entropy, AND Product, and Disjoint Union of Graphs’. In: ISIT 2023 - International Symposium on Information Theory. Taipei (Taiwan), Taiwan, 25th June 2023, pp. 1–6. URL: <https://hal.science/hal-04133013>.
- [19] N. Charpenay, M. Le Treust and A. Roumy. ‘Optimal Zero-Error Coding for Computing under Pairwise Shared Side Information’. In: ITW 2023 - Information Theory Workshop. Saint Malo, France: IEEE, 2023, pp. 1–5. URL: <https://hal.science/hal-03901563>.
- [20] D. R. Freitas, C. Guillemot and I. Tabus. ‘Compression of PlenOctree Model Attributes Enabling Fast Communication and Rendering of Neural Radiance Fields’. In: *Proceedings European Signal Processing Conference. EUSIPCO 2023 - European Signal Processing Conference*. Helsinki, Finland, 4th Sept. 2023, pp. 1–5. URL: <https://hal.science/hal-04195520>.
- [21] K. Gu, T. Maugey, S. Knorr and C. Guillemot. ‘Vanishing Point Aided Hash-Frequency Encoding for Neural Radiance Fields (NeRF) from Sparse 360°Input’. In: *Proceedings ISMAR. ISMAR 2023 - 22nd IEEE International Symposium on Mixed and Augmented Reality*. Sydney, Australia: IEEE, 16th Oct. 2023, pp. 1–10. URL: <https://hal.science/hal-04197185>.
- [22] N. Keriven and S. Vaiter. ‘What functions can Graph Neural Networks compute on random graphs? The role of Positional Encoding’. In: *Proceedings NeurIPS. NeurIPS 2023 - 37th Annual Conference on Neural Information Processing Systems. Proceedings NeurIPS*. New-Orleans, United States, 23rd May 2023, pp. 1–28. URL: <https://hal.science/hal-04103771>.
- [23] B. Le Bon, M. Le Pendu and C. Guillemot. ‘Stochastic Unrolled Proximal Point Algorithm for linear image inverse problems’. In: EUSIPCO 2023 - 31st European Signal Processing Conference. Helsinki, Finland, 2023, pp. 1–5. URL: <https://hal.science/hal-04109191>.
- [24] B. Le Bon, M. Le Pendu and C. Guillemot. ‘Unrolled Fourier Disparity Layer Optimization For Scene Reconstruction From Few-Shots Focal Stacks’. In: ICASSP 2023 - IEEE International Conference on Acoustics, Speech and Signal Processing. Rhodes, Greece: IEEE, 4th June 2023, pp. 1–5. URL: <https://hal.science/hal-04054360>.
- [25] G. Le Guludec and C. Guillemot. ‘Joint Neural Representation For Multiple Light Fields’. In: ICASSP 2023 - IEEE International Conference on Acoustics, Speech and Signal Processing. Rhodes, Greece: IEEE, 4th June 2023, pp. 1–5. URL: <https://hal.science/hal-04054325>.
- [26] Y. Liu, H. Amirpour, M. Abdoli, C. Timmerer and T. Guionnet. ‘Preparing VVC for Streaming: A Fast Multi-Rate Encoding Approach’. In: VCIP 2023 - IEEE International Conference on Visual Communications and Image Processing. Jeju, North Korea: IEEE, 2023, pp. 1–5. DOI: [10.48550/arXiv.2312.08330](https://doi.org/10.48550/arXiv.2312.08330). URL: <https://hal.science/hal-04348313>.
- [27] T. Maugey. ‘Towards digital sobriety: why improving the energy efficiency of video streaming is not enough’. In: MMSP 2023 - 25th IEEE International Workshop on MultiMedia Signal Processing. Poitiers, France: IEEE, 2023, pp. 1–4. URL: <https://inria.hal.science/hal-04231363>.
- [28] C. Petit, A. Roumy and T. Maugey. ‘A Water-filling Algorithm Maximizing the Volume of Submatrices Above the Rank’. In: EUSIPCO 2023 - 31st European Signal Processing Conference. Helsinki, Finland, 4th Sept. 2023, pp. 1–5. URL: <https://inria.hal.science/hal-04132343>.
- [29] R. Piau, T. Maugey and A. Roumy. ‘Learning on entropy coded images with CNN’. In: *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. ICASSP 2023 - IEEE International Conference on Acoustics, Speech and Signal Processing. Rhodes, Greece: IEEE, 2023, pp. 1–5. DOI: [10.1109/ICASSP49357.2023.10095647](https://doi.org/10.1109/ICASSP49357.2023.10095647). URL: <https://hal.science/hal-04133662>.
- [30] R. Piau, T. Maugey and A. Roumy. ‘Predicting cnn learning accuracy using chaos measurement’. In: ICASSP 2023 - IEEE International Conference on Acoustics, Speech, and Signal Processing. Rhodes, Greece: IEEE, 2023, pp. 1–4. URL: <https://hal.science/hal-04133669>.
- [31] J. Shi and C. Guillemot. ‘Light Field Compression via Compact Neural Scene Representation’. In: ICASSP 2023 - IEEE International Conference on Acoustics, Speech, and Signal Processing. Rhodes island, Greece, 4th June 2023, pp. 1–5. URL: <https://inria.hal.science/hal-04017645>.



- [32] S. Willingham, M. Sjöström and C. Guillemot. ‘Prior for Multi-Task Inverse Problems in Image Reconstruction Using Deep Equilibrium Models’. In: *Proceedings European Conference on Signal Processing*. EUSIPCO 2023 - 31st European Signal Processing Conference. Helsinki (Aalto University School of Economics), Finland, 4th Sept. 2023, pp. 1–5. URL: <https://hal.science/hal-04195526>.

#### National peer-reviewed Conferences

- [33] P. Bacchus, A. Roumy, R. Fraisse and C. Guillemot. ‘Compression quasi sans perte d’images satellites par filtrage de résidus’. In: *Actes du GretsI 2023*. GRETSI 2023 - XXIXème Colloque Francophone de Traitement du Signal et des Images. Grenoble, France, 28th Aug. 2023, pp. 1–5. URL: <https://hal.science/hal-04156801>.
- [34] N. Charpenay, M. Le Treust and A. Roumy. ‘Codage optimal zéro-erreur pour le calcul de fonction avec information adjacente partagée deux à deux’. In: GRETSI 2023 - XXIXème Colloque Francophone de Traitement du Signal et des Images. GRENOBLE, France, 2023, pp. 1–4. URL: <https://hal.science/hal-04251973>.
- [35] M. Cordonnier, N. Keriven, N. Tremblay and S. Vaiter. ‘Convergence of Graph Neural Networks with generic aggregation functions on random graphs’. In: GRETSI 2023 - XXIXème Colloque Francophone de Traitement du Signal et des Images. Grenoble, France, 2023, pp. 1–4. URL: <https://hal.science/hal-04373554>.
- [36] M. Gjorgjevski, N. Keriven, S. Barthelme and Y. de Castro. ‘The Graphical Nadaraya-Watson Estimator in Latent Position Models’. In: *Actes GRETSI 2023*. GRETSI 2023 - XXIXème Colloque Francophone de Traitement du Signal et des Images. Grenoble, France, 2023, pp. 1–4. URL: <https://hal.science/hal-04241086>.
- [37] C. Petit, A. Roumy and T. Maugey. ‘Algorithme de type « Water-Filling » maximisant le volume de matrices rectangulaires au-delà du rang’. In: GRETSI 2023 - XXIXe colloque francophone de traitement du signal et des images. Grenoble, France, 2023, pp. 1–4. URL: <https://inria.hal.science/hal-04156155>.
- [38] M. Theveneau and N. Keriven. ‘Stability of Entropic Wasserstein Barycenters and application to random geometric graphs’. In: GRETSI 2023 - XXIXème Colloque Francophone de Traitement du Signal et des Images. Grenoble, France, 2023, pp. 1–7. URL: <https://hal.science/hal-03852485>.

#### Conferences without proceedings

- [39] M. Cordonnier, N. Keriven, N. Tremblay and S. Vaiter. ‘Convergence of Message Passing Graph Neural Networks with Generic Aggregation On Random Graphs’. In: GSP 2023 - 6th Graph Signal Processing workshop. Oxford, United Kingdom, 2023, pp. 1–3. URL: <https://hal.science/hal-04106511>.
- [40] A. Lecert, A. Roumy, R. Fraisse and C. Guillemot. ‘Illumination colorée et information commune pour le modèle Horn-Retinex en restauration d’images à faible luminosité’. In: GRETSI 2023 - XXIXème Colloque Francophone de Traitement du Signal et des Images. Grenoble, France, 2023, pp. 1–4. URL: <https://hal.science/hal-04383530>.
- [41] R. Piau, T. Maugey and A. Roumy. ‘Prédiction de la précision d’apprentissage des réseaux de neurones convolutifs par mesure du chaos’. In: GRETSI 2023. Grenoble, France, 28th Aug. 2023. URL: <https://hal.science/hal-04384164>.

#### Doctoral dissertations and habilitation theses

- [42] P. Bacchus. ‘Deep learning for satellite image compression’. Université Rennes, 12th Dec. 2023. URL: <https://hal.science/tel-04355362>.
- [43] N. Charpenay. ‘Zero-error network information theory: graphs, coding for computing and source-channel duality’. Université de Rennes, 28th Nov. 2023. URL: <https://theses.hal.science/tel-04384481>.

- [44] R. Fermanian. ‘Deep learning for inverse problems and application to omnidirectional imaging’. INRIA Rennes - Bretagne Atlantique and University of Rennes 1, France; SIROCCO, 6th Dec. 2023. URL: <https://theses.hal.science/tel-04363191>.
- [45] B. Le Bon. ‘Deep learning for light field acquisition and restoration’. Université de Rennes 1, 29th Nov. 2023. URL: <https://hal.science/tel-04360264>.
- [46] A. Lecert. ‘Low-light image restoration with deep learning techniques’. Rennes 1; INRIA, 20th Dec. 2023. URL: <https://hal.science/tel-04342921>.
- [47] Y. Liu. ‘Learning for new generation video coders: Complexity Reduction of Inter coding for VVC Codec’. Inria Rennes - Bretagne Atlantique & IRISA, 11th Dec. 2023. URL: <https://hal.science/tel-04364408>.

### Reports & preprints

- [48] Y. Liu, M. Riviere, T. Guionnet, A. Roumy and C. Guillemot. *CNN-based Prediction of Partition Path for VVC Fast Inter Partitioning Using Motion Fields*. 21st Oct. 2023. DOI: [10.48550/arXiv.2310.13838](https://doi.org/10.48550/arXiv.2310.13838). URL: <https://hal.science/hal-04252664>.

### Other scientific publications

- [49] R. Fermanian, T. Maugey and C. Guillemot. ‘SphereDRUNet: A Spherical Denoiser for Omnidirectional Images’. In: ISMAR 2023 - 22nd IEEE International Symposium on Mixed and Augmented Reality. Sydney, Australia: IEEE, 16th Oct. 2023, pp. 1–6. URL: <https://hal.science/hal-04197479>.