2022
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

Contents

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]

- Mikael Le Pendu [INRIA, Starting Research Position, until May 2022]

- Thomas Maugey [INRIA, Researcher]

- Claude Petit [INRIA, Researcher]

- Aline Roumy [INRIA, Senior Researcher, HDR]

Post-Doctoral Fellow

- Anju Jose tom [INRIA, until Aug 2022]

PhD Students

- Ipek Anil Atalay Appak [CEA]

- Denis Bacchus [INRIA]

- Tom Bachard [UNIV RENNES I]

- Tom Bordin [Inria, from Oct 2022]

- Nicolas Charpenay [UNIV RENNES I]

- Davi Rabbouni De Carvalho Freitas [INRIA]

- Rita Fermanian [INRIA]

- Kai Gu [INRIA]

- Reda Kaafarani [INRIA, CIFRE, Mediakind]

- Brandon Le Bon [INRIA]

- Arthur Lecert [INRIA]

- Yiqun Liu [ATEME, CIFRE]

- Rémi Piau [INRIA]

- Soheib Takhtardeshir [UNIV MID SWEDEN]

- Samuel Willingham [INRIA]

Technical Staff

- Sébastien Bellenous [INRIA, Engineer]

- Guillaume Le Guludec [INRIA, Engineer]

- Jinglei SHI [Inria, Engineer, until May 2022]

Administrative Assistant

- Caroline Tanguy [INRIA]

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, Ateme, Orange) or in the framework of national projects. The application domains addressed by the project are networked visual applications taking into account their various requirements and needs in terms of compression, of network adaptation, of advanced functionalities such as navigation, interactive streaming and high quality rendering.

2.2 Visual Data Analysis

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

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

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

2.4   Algorithms for inverse problems in visual data processing

Based on the above models, besides compression, our goal is also to develop algorithms for solving a number of inverse problems in computer vision. Our emphasis is on methods to cope with limitations of sensors (e.g. enhancing spatial, angular or temporal resolution of captured data, or noise removal), to synthesize virtual views or to reconstruct (e.g. in a compressive sensing framework) light fields from a sparse set of input views, to recover HDR visual content from multiple exposures, and to enable content editing (we focus on color transfer, re-colorization, object removal and inpainting). Note that view synthesis is a key component of multiview and light field compression. 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 corner-stones 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 $m << n$ and A is of full row rank, one seeks the solution of $\min\{\|\mathbf{x}\|_0 : A\mathbf{x} = \mathbf{b}\}$, where $\|\mathbf{x}\|_0$ denotes the ℓ_0 norm of $\mathbf{x}$, 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 $\min\{\|\mathbf{x}\|_0 : \|A\mathbf{x} - \mathbf{b}\|_p \leq \rho\}$, 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 θ, that computes a representation Y from the data X, and a decoder g_d, parametrized by ϕ, that gives a reconstruction $\hat{X}$ of X (see Figure below). 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 $\hat{Y} = Q(Y)$ (see Figure below). If an autoencoder has fully-connected layers, the architecture, and the number of parameters to be learned, depends on the image size. Hence one autoencoder has to be trained per image size, which poses problems in terms of genericity.

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 preferrable. The obtained representation is thus a set of so-called feature maps.

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

Deep neural networks have also been shown to be very promising for solving inverse problems (e.g. super-resolution, sparse recovery in a compressive sensing framework, inpainting) in image processing. Variational autoencoders, generative adversarial networks (GAN), learn, from a set of examples, the latent space or the manifold in which the images, that we search to recover, reside. The inverse problems can be re-formulated using a regularization in the latent space learned by the network. For the needs of the regularization, the learned latent space may need to verify certain properties such as preserving distances

or neighborhood of the input space, or in terms of statistical 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 [1] is central to our compression and communication activities, in particular to the design of entropy codes and of error correcting codes. Another field in coding theory which has emerged in the context of sensor networks is Distributed Source Coding (DSC). It refers to the compression of correlated signals captured by different sensors which do not communicate between themselves. All the signals captured are compressed independently and transmitted to a central base station which has the capability to decode them jointly. DSC finds its foundation in the seminal Slepian-Wolf [2] (SW) and Wyner-Ziv [3] (WZ) theorems. Let us consider two binary correlated sources X and Y. If the two coders communicate, it is well known from Shannon's theory that the minimum lossless rate for X and Y is given by the joint entropy $H(X, Y)$. Slepian and Wolf have established in 1973 that this lossless compression rate bound can be approached with a vanishing error probability for long sequences, even if the two sources are coded separately, provided that they are decoded jointly and that their correlation is known to both the encoder and the decoder.

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

4   Application domains

4.1   Overview

The application domains addressed by the project are:

- Compression with advanced functionalities of various imaging modalities

- Networked multimedia applications taking into account needs in terms of user and network adaptation (e.g., interactive streaming, resilience to channel noise)

- Content editing, post-production, and computational photography.

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)

_____

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

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

[3] A. Wyner and J. Ziv, "The rate-distortion function for source coding ith side information at the decoder." IEEE Transactions on Information Theory, pp. 1-10, January 1976.

images and videos (higher bit depth, extended colorimetric space), light fields and omni-directional imaging.

Different video data formats and technologies are envisaged for interactive and immersive 3D video applications using omni-directional videos, stereoscopic or multi-view videos. The "omni-directional video" set-up refers to 360-degree view from one single viewpoint or spherical video. Stereoscopic video is composed of two-view videos, the right and left images of the scene which, when combined, can recreate the depth aspect of the scene. A multi-view video refers to multiple video sequences captured by multiple video cameras and possibly by depth cameras. Associated with a view synthesis method, a multi-view video allows the generation of virtual views of the scene from any viewpoint. This property can be used in a large diversity of applications, including Three-Dimensional TV (3DTV), and Free Viewpoint Video (FVV). In parallel, the advent of a variety of heterogeneous delivery infrastructures has given momentum to extensive work on optimizing the end-to-end delivery QoS (Quality of Service). This encompasses compression capability but also capability for adapting the compressed streams to varying network conditions. The scalability of the video content compressed representation and its robustness to transmission impairments are thus important features for seamless adaptation to varying network conditions and to terminal capabilities.

4.3   Networked visual applications

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

Broadband fixed and mobile access networks with different radio access technologies have enabled not only IPTV and Internet TV but also the emergence of mobile TV and mobile devices with internet capability. A major challenge for next internet TV or internet video remains to be able to deliver the increasing variety of media (including more and more bandwidth demanding media) with a sufficient end-to-end QoS (Quality of Service) and QoE (Quality of Experience).

4.4   Editing, post-production and computational photography

Editing and post-production are critical aspects in the audio-visual production process. Increased ways of "consuming" visual content also highlight the need for content repurposing as well as for higher interaction and editing capabilities. Content repurposing encompasses format conversion (retargeting), content summarization, and content editing. This processing requires powerful methods for extracting condensed video representations as well as powerful inpainting techniques. By providing advanced models, advanced video processing and image analysis tools, more visual effects, with more realism become possible. Our activies around light field imaging also find applications in computational photography which refers to the capability of creating photographic functionalities beyond what is possible with traditional cameras and processing tools.

5   Social and environmental responsibility

No social or environmental responsibility.

6 Highlights of the year

6.1 Awards

- C. Guillemot has received the distinction of "Officier de la Légion d'Honneur" from Michel Cosnard, Prof. Emerite at the Univ. of Nice Sophia-Antipolis, 21 April 2022.

7 New software and platforms

7.1 New software

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

Keywords: Image compression, Random access

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

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

Authors: Navid Mahmoudian Bidgoli, Thomas Maugey, Aline Roumy

Contact: Aline Roumy

7.1.2 GRU_DDLF

Name: Gated Recurrent Unit and Deep Decoder based generative model for Light Field compression

Keywords: Light fields, Neural networks, Deep learning, Compression

Functional Description: This code implements a neural network-based light field representation and compression scheme The code is developed in python, based on the pytorch deep learning framework..The proposed network is based on both a generative model that aims at modeling the spatial information that is static, i.e., found in all light field views, and on a convolutional Gated Recurrent Unit (ConvGRU) that is used to model variations between blocks of angular views. The network is untrained in the sense that it is learned only on the light field to be processed, without any additional training data. The network weights can be considered as a representation of the input light field. The compactness of this representation depends on the number of weights or network parameters, but not only. It also depends on the number of bits needed to accurately quantize each weight. Our network is thus learned using a strategy that takes into account weight quantization, in order to minimize the effect of weight quantization noise on the light field reconstruction quality. The weights of the different filters are optimized end-to-end in order to optimize the image reconstruction quality for a given number of quantization bits per weight.

Contact: Christine Guillemot

Participants: Xiaoran Jiang, Jinglei Shi, Christine Guillemot

7.1.3 FPFR+

Name: Fused Pixel and Feature based Reconstruction for view synthesis and temporal interpolation

Keywords: Light fields, View synthesis, Temporal interpolation, Neural networks

Functional Description: This code implements a deep residual architecture that can be used both for synthesizing high quality angular views in light fields and temporal frames in classical videos. The code is developped in python, based on the tensorflow deep learning framework. The proposed framework consists of an optical flow estimator optimized for view synthesis, a trainable feature extractor and a residual convolutional network for pixel and feature-based view reconstruction. Among these modules, the fine-tuning of the optical flow estimator specifically for the view synthesis task yields scene depth or motion information that is well optimized for the targeted problem. In cooperation with the end-to-end trainable encoder, the synthesis block employs both pixel-based and feature-based synthesis with residual connection blocks, and the two synthesized views are fused with the help of a learned soft mask to obtain the final reconstructed view.

Contact: Christine Guillemot

Participants: Jinglei Shi, Xiaoran Jiang, Christine Guillemot

7.1.4 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), diifferent preconditioning methods, and different denoising neural networks.

Contact: Christine Guillemot

Participants: Mikael Le Pendu, Christine Guillemot

7.1.5 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

7.1.6 DeepLFCam

Name: Deep Light Field Acquisition Using Learned Coded Mask Distributions for Color Filter Array Sensors

Keywords: Light fields, Deep learning, Compressive sensing

Functional Description: This code implements a deep light field acquisition method using learned coded mask distributions for color filter array sensors. The code describes and allows the execution of 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: - 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, - a color filtering array performing color multiplexing, - a monochromatic sensor. This composition of filtering, sampling, projection, is modeled as linear projection operator. Those measurements, along with the 'modulation field', or 'shield field' corresponding to the acquisition device (i.e. the light field corresponding to the the transformation of a uniformly white light field, known to completelly characterize the projection operator), are then feed to a deep convolutional residual network to reconstruct the original light field. The provided implementation is meant to be used with Tensorflow 2.1.

Publication: leguludec:hal-03203347

Contact: Guillaume Le Guludec

Participants: Guillaume Le Guludec, Christine Guillemot

7.1.7 DeepULFCam

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

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

Functional Description: This code describes and allows the execution of 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: - 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, - a color filtering array performing color multiplexing, - a monochromatic sensor. The composition of these filterings/projects is modeled as linear projection operator. The light field is then reconstructed by performing the 'unrolling' of an interative reconstruction algorithm (namely the HQS 'half-quadratic splitting' algorithm, a variant of ADMM, an optimization algorithm, applied to the solving a regularized least-squares problem) into a deep convolutional neural network. The algorithm makes use of the structure of the projection operator to efficiently solve the quadratic data-fidelty 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

7.1.8 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

7.1.9 JNR-MLF

Name: Joint Neural Representation for Multiple Light Fields

Keywords: Neural networks, Light fields

Functional Description: The code is designed to learn 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. However, unlike SIGNET which uses the Legendre polynomials as the positional encoding, the model used in the code is using learned Fourier Features.

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. More precisely, the code uses matrix factorization inspired from the Singluar Value Decomposition (SVD).

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: - The representation provider, which takes the index i of a scene, and outputs a weight matrix and bias vector for each layer. - 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, a flavour of the SGD algorithm. 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 backprogagation from a MSE minimization objective. The code uses TensorFlow 2.7.

Contact: Guillaume Le Guludec

Participant: Guillaume Le Guludec

7.2 New platforms

7.2.1 CLIM processing toolbox

Participants: Allain Christine Guillemot, Laurent Guillo.

As part of the ERC Clim project, the EPI Sirocco is developing a light field processing toolbox. The toolbox and libraries are developed in C++ and the graphical user interface relies on Qt. As input data, this tool accepts both sparse light fields acquired with High Density Camera Arrays (HDCA) and denser light fields captured with plenoptic cameras using microlens arrays (MLA). At the time of writing, in addition to some simple functionalities, such as re-focusing, change of viewpoints, with different forms of visualization, the toolbox integrates more advanced tools for scene depth estimation from sparse and

dense light fields, for super-ray segmentation and scene flow estimation, and for light field denoising and angular interpolation using anisotropic diffusion in the 4D ray space. The toolbox is now being interfaced with the C/C++ API of the tensorflow platform, in order to execute deep models developed in the team for scene depth and scene flow estimation, view synthesis, and axial super-resolution.

7.2.2   ADT: Interactive Coder of Omnidirectional Videos

Participants: Sébastien Bellenous, Navid Mahmoudian Bidgoli, Thomas Maugey, Aline Roumy.
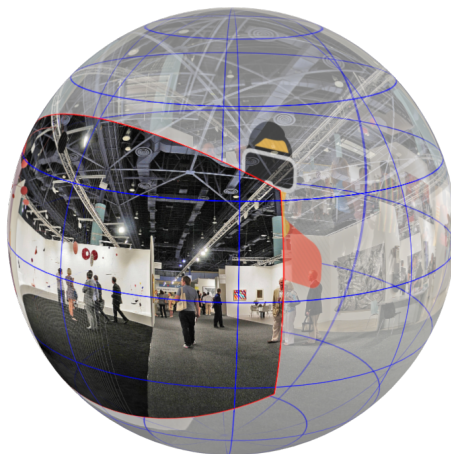
Figure 1: By nature, a video 360° cannot be seen entirely, because it covers all the direction. The developed coding scheme will aim at 1) transmitting only what is necessary and 2) bringing no transmission rate overhead.

In the Intercom project, we have studied the impact of interactivity on the coding performance. We have, for example, tackled the following problem: is it possible to compress a 360 video (as shown in Fig.1) once for all, and then partly extract and decode what is needed for a user navigation, while, keeping good compression performance? After having derived the achievable theoretical bounds, we have built a new omnidirectional video coder. This original architecture enables to reduce significantly the cost of interactivity compared to the conventional video coders. The project ICOV proposes to start from this promising proof of concept and to develop a complete and well specified coder that is aimed to be shared with the community.

8   New results

8.1   Visual Data Analysis

Keywords: Scene depth, Scene flows, 3D modeling, Light-fields, camera design, 3D point clouds.

8.1.1   Light Field Acquisition Using Colour Coded Masks and Meta-Surfaces

Participants: Anil Ipek Atalay Appak, Christine Guillemot, Guillaume Le Guludec.

Compressed sensing using color-coded masks has been recently considered for capturing light fields using a small number of measurements. Such an acquisition scheme is very practical, since any

consumer-level camera can be turned into a light field acquisition camera by simply adding a coded mask in front of the sensor. We developed an efficient and mathematically grounded deep learning model to reconstruct a light field from a set of measurements obtained using a color-coded mask and a color filter array (CFA). Following the promising trend of unrolling optimization algorithms with learned priors, we formulate our task of light field reconstruction as an inverse problem and derive a principled deep network architecture from this formulation [11]. We also introduce a closed-form extraction of information from the acquisition, while similar methods found in the recent literature systematically use an approximation. Compared to similar deep learning methods, we show that our approach allows for a better reconstruction quality. We further show that our approach is robust to noise using realistic simulations of the sensing acquisition process.

We pursue this study by focusing on microscopy with extended depth of field. Indeed, in conventional microscopy imaging, it is not possible to clearly see all objects at different depths due to optics limitations. Extended Depth of Field (EDOF) lenses can solve this problem, however there is a trade-off between the resolution and the depth of field. To overcome this trade-off, phase masks have been used combined them with reconstruction networks.

In collaboration with the university of Tampere (Prof. Humeyra Caglayan, Erdem Sahin) in the context of the Plenoptima Marie Curie project, we adress 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 deep learning based image reconstruction methods. In our design, instead of phase masks, we consider instead meta-surface materials. Meta-surfaces are periodic sub-wavelength structures that can change the amplitude, phase, and polarization of the incident wave. Its parameters are end-to-end optimized to have the highest depth of field with good spatial resolution and high reconstruction quality.

While, in a first step we focused on 2D microscopy imaging, the next step will be to consider light field microscopy. This co-design should indeed lead to a new plenoptic camera design which would take full advantage of the sensor resolution and allow for extended depth of fields, a property that is very important in microscopy.

8.1.2 Sparse Non Parametric BRDF Model and Plenopic Point Clouds

Participants: Davi Freitas, Christine Guillemot.

In collaboration with the Univ. of Linkoping, we have derived a novel sparse non-parametric BRDF model using a machine learning approach to represent the space of possible BRDFs using a set of multidimensional sub-spaces, or dictionaries [16]. By training the dictionaries under a sparsity constraint, the model guarantees high quality representations with minimal storage requirements and an inherent clus- tering of the BDRF-space. The model can be trained once and then reused to represent a wide variety of measured BRDFs. Moreover, the proposed method is flexible to incorporate new unobserved data sets, parameterizations, and transformations. In addition, we show that any two, or more, BRDFs can be smoothlyinterpolated in the coefficient space of the model rather than the significantly higher-dimensional BRDF space. Experimental results show that the proposed approach results in about 9.75dB higher SNR on average for rendered images as compared to current state-of-the-art models.

We pursue this study in the context of the Plenoptima project by designing a novel scene representation, a richer model, in which color would be untangled from lighting, using BRDF models. In explicit representations like plenoptic point clouds (PPC), the RGB colors in the model indeed result from interactions between the incoming light during capture and the material and shape properties of the object in the scene. Hence, if the scene modelled by the PPC is put into an environment with different lighting conditions, it would look out of place since these conditions at the time of capture are "baked in" the colors of the PPC. To address this problem, we leveraged the plenoxel representation and developed a plenoxel point cloud representation in which view-dependent spatial information, represented by BRDF and compressed with spherical harmonics, is associated to the voxel grid.

8.1.3 A Light Field FDL-HCGH Feature in Scale-Disparity Space

Participants: Christine Guillemot.

Many computer vision applications rely on feature detection and description, hence the need for computationally efficient and robust 4D light field (LF) feature detectors and descriptors. In collaboration with Xi'an University (Prof. Zhaolin Xiao), we have proposed a novel light field feature descriptor based on the Fourier disparity layer representation, for light field imaging applications [18]. After the Harris feature detection in a scale-disparity space, the proposed feature descriptor is then extracted using a circular neighborhood rather than a square neighborhood. It is shown to yield more accurate feature matching, compared with the LiFF LF feature, with a lower computational complexity. In order to evaluate the feature matching performance with the proposed descriptor, we generated a synthetic stereo LF dataset with ground truth matching points. Experimental results with synthetic and real-world datasets show that our solution outperforms existing methods in terms of both feature detection robustness and feature matching accuracy. It is in particular shown to yield more accurate feature matching, compared with the reference LIght Field Feature (LiFF) descriptor, with a lower computational complexity. Experimental results with synthetic and real-world datasets show that our solution outperforms existing methods in terms of both feature detection robustness and feature matching accuracy.

8.2   Signal processing and learning methods for visual data representation and compression

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

8.2.1   Neural Network Priors for Light Field Representation

Participants: Christine Guillemot, Xiaoran Jiang, Jinglei Shi.

Deep generative models have proven to be effective priors for solving a variety of image processing problems. However, the learning of realistic image priors, based on a large number of parameters, requires a large amount of training data. It has been shown recently, with the so-called deep image prior (DIP), that randomly initialized neural networks can act as good image priors without learning.

We have proposed a deep generative model for light fields, which is compact and which does not require any training data other than the light field itself. The proposed network is based on both a generative model that aims at modeling the spatial information that is static, i.e., found in all light field views, and on a convolutional Gated Recurrent Unit (ConvGRU) that is used to model variations between angular views [10]. The spatial view generative model is inspired from the deep decoder, itself built upon the deep image prior, but that we enhance with spatial and channel attention modules, and with quantization-aware learning. The attention modules modulate the feature maps at the output of the different layers of the generator. In addition, we offer an option which expressively encodes the upscaling operations in learned weights in order to better fit the light field to process. The deep decoder is also adapted in order to model several light field views, with layers (i.e. features) that are common to all views and others that are specific to each view. The weights of both the convGRU and the deep decoder are learned end-to-end in order to minimize the reconstruction error of the target light field.

To show the potential of the proposed generative model, we have developed a complete light field compression scheme with quantization-aware learning and entropy coding of the quantized weights. Experimental results show that the proposed method outperforms state-of-the-art light field compression methods as well as recent deep video compression methods in terms of both PSNR and MS-SSIM metrics.

8.2.2   Neural Radiance Fields and Implicit Scene Representations

Participants: Christine Guillemot, Guillaume Le Guludec, Jinglei Shi.

The concept of Neural Radiance Fields (NeRF) has been recently introduced for scene representation and light field view synthesis. NeRF is an implicit Multi-Layer Perceptron-based model that maps 5D vectors—3D coordinates plus 2D viewing directions—to opacity and color values, computed by fitting the model to a set of training views. The resulting neural model can then be used to generate any view of the light field, in a continuous manner, using volume rendering techniques.

NeRF can thus be seen as a neural representation of a scene which is learned from a set of input views. While existing light field compression solutions usually consider encoding and transmitting a subset of light field views which are then used at the decoder to synthesize and render the entire light field, we consider an alternative approach which optimizes and compresses the NeRF on the sender side, so that the network itself is transmitted to the receiver.

For high efficiency, we have developed a Distilled Low Rank optimization method for Neural Radiance Fields. The resulting compact neural radiance fields can then be efficiently used for compressing the light field of a scenes. While existing compression methods encode the set of light field sub-aperture images, our proposed method instead learns an implicit scene representation in the form of a Neural Radiance Field (NeRF), which also enables view synthesis. For reducing its size, the model is first learned under a Low Rank (LR) constraint using a Tensor Train (TT) decomposition in an Alternating Direction Method of Multipliers (ADMM) optimization framework. To further reduce the model size, the components of the tensor train decomposition need to be quantized. However, performing the optimization of the NeRF model by simultaneously taking the low rank constraint and the rate-constrained weight quantization into consideration is challenging. To deal with this difficulty, we introduce a network distillation operation that separates the low rank approximation and the weight quantization in the network training. The information from the initial LR constrained NeRF (LR-NeRF) is distilled to a model of a much smaller dimension (DLR-NeRF) based on the TT decomposition of the LR-NeRF. An optimized codebook is then learned to quantize all TT components, producing the final quantized representation. Experimental results show that our proposed method yields better compression efficiency compared with state-of-the-art methods, and it additionally has the advantage of allowing the synthesis of any light field view with a high quality.

GT view(5,5)	RLVC	HLVC	HEVC	JPL	Ours
boxes	PSNR=35.49dB bpp=0.110	PSNR=34.68dB bpp=0.094	PSNR=35.73dB bpp=0.099	PSNR=36.51 bpp=0.100	PSNR=38.19dB bpp=0.105
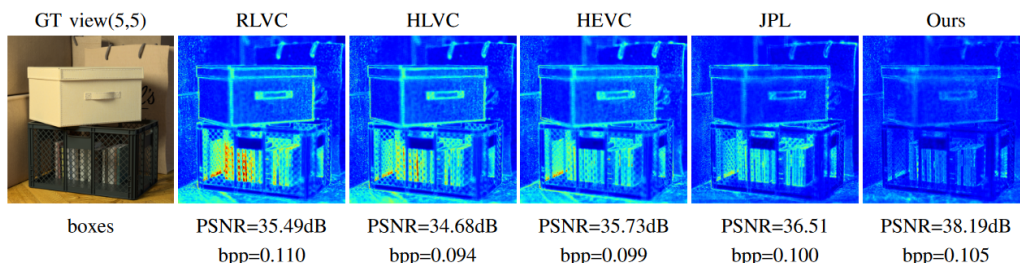
Figure 2: Averaged reconstruction error maps of decompressed light fields, when using the Jpeg-Pleno standard (JPL), the HEVC compression standard, and deep learning-based video compression solutions (RLVC, HLVC). We can see that our low-rank constrained neural radiance field representation gives a better reconstruction quality (error map with lower values, higher PSNR).

Neural implicit representations have thus 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, as described above, we have first focused on finding 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. We show 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.

8.2.3   Omni-directional Neural Radiance Fields

Participants: Kai Gu, Christine Guillemot, Thomas Maugey.

In this study, we explore the use of fish-eye cameras to acquire spherical light fields, the goal being to design novel representations and reconstruction methods from omni-directional imagery. Our motivation is to be able to capture or reconstruct light fields with a very large field of view, i.e. 360°, using fish-eye lenses and developing the corresponding advanced reconstruction and synthesis algorithms. We focus on the question: how do we extract omni-directional information and potentially benefit from it when reconstructing a spherical light field of a large-scale scene with a non- converged camera setup?.

Figure 3: Multiple fish-eye images of a scene.

To address this question, we have developed an omni-directional Neural radiance Field estimation method, called omni-NERF [25]. While NeRF was originally designed for ideal pin-hole camera models with known camera poses, in omni-NERF a new projection model for the distortion of fisheye lenses as a function between the incident and refracted rays has been defined. The model can reconstruct a 360 light field and can fit different fisheye projection models and optimize the camera parameters jointly with the radiance field. We have shown that the spherical sampling improves the performance when training with panoramic fisheye or wide-angle perspective images, and that it improves the performance of view synthesis compared to planar sampling of the scene.

8.2.4   Satellite image compression and restoration

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

In the context of the Lichie project, in collaboration with Airbus, we address two problems for satellite imaging: quasi-lossless compression and restoration, using deep learning 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 consider in this work, 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 filter array. Second, (ii) it should be *adapted to the image statistics*. Indeed, satellite mages 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. We have proposed a compression algorithm, based on an autoencoder, that can efficiently deal with these three constraints.

Indeed, 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 compression and demosaicking of satellite images. 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.

To restore the low light images, we have designed a GAN-based solution to decompose the low light images into its components according to the retinex model (illumination, reflectance). We have introduced different regularization constraints when learning the Retinex model components and we have explored different variants of the retinex model, either assuming a monochomatic or a coloured illumination component. The approach yields state of the art solutions while being very fast for the restoration.

8.2.5   Neural networks for video compression acceleration

Participants: Christine Guillemot, Yiqun Liu, Aline Roumy.

In the context of the Cifre contract with Ateme we investigate deep learning architectures for the inference of coding modes in video compression algorithms with the ultimate goal of reducing the encoder complexity. We focused on the acceleration of the Versatile Video Coding (VVC) standard.

The Versatile Video Coding (VVC) standard has been finalized in 2020. Compared to the High Efficiency Video Coding (HEVC) standard, VVC offers about 50% compression efficiency gain, at the cost of about 10x more encoder complexity. We have developed a Convolutional Neural Network (CNN)-based method to speed up the inter partitioning in VVC [27]. Our method operates at the Coding Tree Unit (CTU) level, by splitting each CTU into a fixed grid of 8×8 blocks. Then each cell in this grid is associated with information about the partitioning depth within that area. A lightweight network for predicting this grid is employed during the rate-distortion optimization to avoid partitions that are unlikely to be selected. Experiments show that the proposed method can achieve acceleration ranging from 17% to 30% in the Random Access Group Of Picture 32 (RAGOP32) mode of VVC Test Model (VTM)10 with a reasonable efficiency drop ranging from 0.37% to 1.18% in terms of BD-rate increase.

8.3   Algorithms for inverse problems in visual data processing

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

8.3.1   Plug-and-play optimization and learned priors

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

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.

In the context of the AI chair DeepCIM, we have first studied Plug-and-Play optimization methods for solving inverse problems by plugging a denoiser into the ADMM (Alternating Direction Multiplier Method) optimization algorithm. We have extended the concept of plug-and-play optimization to use denoisers that can be parameterized for non-constant noise variance. In that aim, we have introduced a preconditioning of the ADMM algorithm, which mathematically justifies the use of such an adjustable denoiser [12]. We additionally proposed a procedure for training a convolutional neural network for high quality non-blind image denoising that also allows for pixel-wise control of the noise standard deviation. We have shown that our pixel-wise adjustable denoiser, along with a suitable preconditioning strategy, can further improve the plug-and-play ADMM approach for several applications, including image completion, interpolation, demosaicing and Poisson denoising. An illustration of Poisson denoising results is given in Fig.(4).
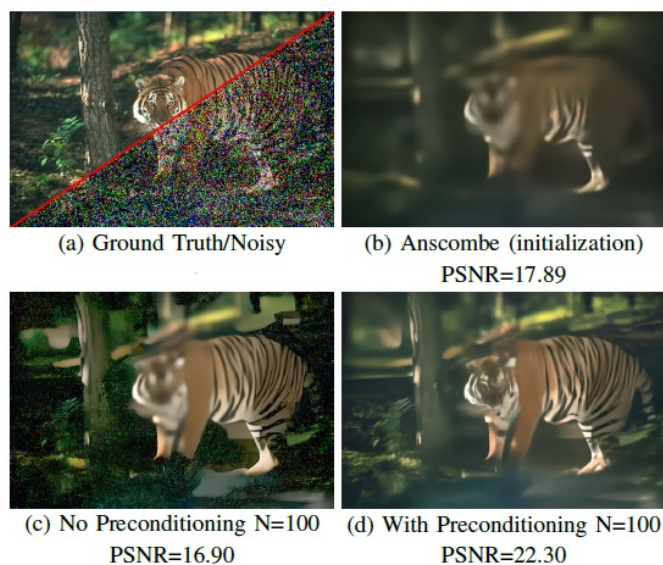

(a) Ground Truth/Noisy

(b) Anscombe (initialization)
PSNR=17.89

(c) No Preconditioning N=100
PSNR=16.90

(d) With Preconditioning N=100
PSNR=22.30

Figure 4: Poisson Denoising for very high noise level. N is the number of iterations of the ADMM iterative algorithms.

We have further shown that it is possible to train a network directly modeling the gradient of a MAP regularizer while jointly training the corresponding MAP denoiser [8]. We use this network in gradient-based optimization methods and obtain better results comparing to other generic Plug-and-Play approaches. We also show that the regularizer can be used as a pre-trained network for unrolled gradient descent. Lastly, we show that the resulting denoiser allows for a better convergence of the Plug-and-Play ADMM.

8.3.2 Unrolled Optimization and deep equilibrium models

Participants: Christine Guillemot, Brandon Le Bon, Mikael Le Pendu, Samuel Willingham.

One advantage of plug-and-play methods with 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. 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.

While iterative methods usually iterate until idempotence, the number of iterations in unrolled optimization methods is set to a small value. This makes it possible to learn a component end-to-end within the optimization algorithm, hence in a way which takes into account the data term, i.e., the degradation operator. But learning networks end-to-end within an unrolled optimization scheme requires high GPU memory usage since the memory used for the backpropagation scales linearly with the number of iterations. This explains why the number of iterations used in an unrolled optimization method is limited. To cope with these limitations, we have developed a stochastic implicit unrolled proximal point algorithm with a learned denoiser, in which sub-problems are defined per iteration. We exploit the fact that the Douglas-Rachford algorithm is an application of the proximal point algorithm to re-define the unrolled step as a proximal mapping. We focused on the unrolled ADMM, which has been demonstrated to be a special case of the Douglas-Rachford algorithm, hence of the proximal point algorithm. This allows us to introduce a novel unrolled proximal gradient method coupling an implicit model and a stochastic learning strategy. We have shown that this stochastic iteration update strategy better controls the learning at each unrolled optimization step, hence leads to a faster convergence than other implicit unrolled methods, while maintaining the advantage of a low GPU memory usage, as well as similar reconstruction quality to the best unrolled methods for all considered image inverse problems.

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. 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.

8.4   User centric compression

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

8.4.1   Zero-error interactive compression

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 no 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. Here, we consider the problem of compression with interaction, where the user can partly extract and decode what is needed for his navigation in the data. This problem can be cast into a source coding where the side information may be absent at the decoder. For this problem, we show that zero-error and vanishing error rate coincide [22] [31].

8.4.2 Coding for computing

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. Therefore we study the problem of coding for computing, when only the class of the tasks, and not the task itself is known upon compression. When the classes of tasks are quite representative (in the database, each class is requested at least twice), we derive a single letter characterization of the optimal compression bounds. More importantly, we show that there exists a coding strategy that outperforms the compression of the whole data [33]. In coding for computing, the optimal coding performance depends on a characteristic graph, that represents the joint distribution of the data and the tasks. We provide new properties of the entropy of AND and OR products of these graphs. This allows to increase the number of graphs for which these graph entropies can be computed.

8.4.3   Semantic Alignment for Multi-item Compression

Participants: Tom Bachard, Thomas Maugey, Anju Jose Tom.

Coding algorithms usually compress independently the images of a collection, in particular when the correlation between them only resides at the semantic level, i.e., information related to the high-level image content. However, in many image sets, even though pixel correlation is absent (i.e., no correlation exists at the spatial level), semantic correlation can be present (i.e., high level content may be similar among the images). This should be exploited to reduce the storage burden. However, no such solution exist. Even worst, the conventional (learning-based or not) image compression of image data set leads to coded representations that are quite spread in the latent space, and thus strongly uncorrelated.

In this work, we have proposed a coding solution able to exploit the semantic redundancy to decrease the storage cost of data collections. First we formally introduced the multi-item compression framework, as the compression of a given image set, in which each image belongs to one class of object (e.g., cat, tree, train, etc.). In other words, the semantic correlation is present as several images may contain the same type of object (even though their pixel similarity is different). Then we derived a loss term to shape the latent space of a variational auto-encoder so that the latent vectors of semantically identical images (i.e., belonging to the same class) can be aligned (see Figure 5). The fact that these latent vectors are aligned further enables an inter-item coder to exploit this redundancy and reduce the bitrate. Finally, we experimentally demonstrated that this alignment leads to a more compact representation of the data collection. This work has been published in [21].
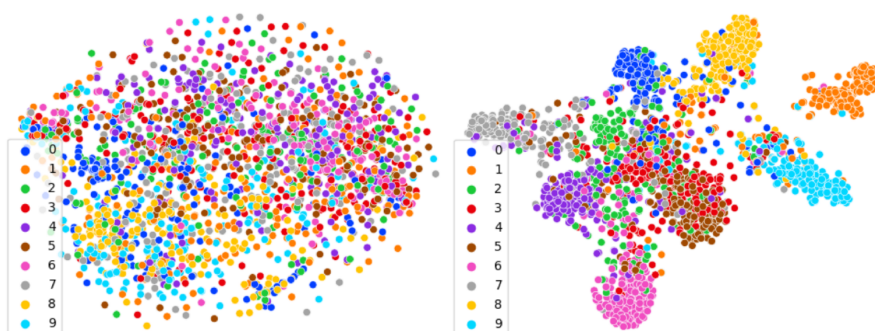

Figure 5: Representation of the multi-image coder latent space before (left) and after (right) semantic alignment. The numbers represent the image class.

8.5 Image Embedding and User Multi-Preference Modeling for Data Collection Sampling

Participants: Thomas Maugey, Anju Jose Tom.

A drastic though efficient way of compressing a data collection consists in deleting a subpart of the data. Contrary to data summarization problem that aims at giving a good overview of an image set, sampling for compression implies that the non-retained images are totally deleted, and thus potentially affects the user experience. This requires to properly model the information contained in an image and thus in an image collection. Modeling the user's taste is also very crucial in order to know what information to preserve and what type of content to delete. Finally one needs to build an efficient sampling strategy to maximize this perceived information.

In this work, we have proposed an end-to-end user-centric sampling method aimed at selecting the images able to maximize the image perceived by a given user. As main contributions, we first introduce a novel metrics that assess the amount of perceived information retained by the user when experiencing a set of images. This metric is set as the volume spanned by the images in a given latent space. Given the actual information present in a set of images, we then show how to take into account 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 show the ability of the proposed approach to accurately integrate user's preference while keeping a reasonable diversity in the sampled image set.

9 Bilateral contracts and grants with industry

9.1 Bilateral contracts with industry

CIFRE contract with Ateme on neural networks for video compression

Participants: Christine Guillemot, Yiqun Liu, Aline Roumy.

- Title : Neural networks for video compression of reduced complexity

- Partners : Ateme (T. Guionnet, M. Abdoli), Inria-Rennes.

- Funding : Ateme, 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.

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 sattelite 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.

10   Partnerships and cooperations

10.1   International research visitors

- Hadi Amirpour, researcher at the Univ. of Klagenfurt, Austria, has visited the team for the period 15-09-2022/09-10-2022.

10.2   European initiatives

10.2.1   H2020 projects

CLIM

> **Participant:**   Christine Guillemot.

CLIM project on cordis.europa.eu

Title: Computational Light fields IMaging

Duration: From September 1, 2016 to February 28, 2022

Partners:

- INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET AUTOMATIQUE (INRIA), France

Inria contact: Christine Guillemot

Coordinator: Christine Guillemot

Summary: Light fields technology holds great promises in computational imaging. Light fields cameras capture light rays as they interact with physical objects in the scene. The recorded flow of rays (the light field) yields a rich description of the scene enabling advanced image creation capabilities from a single capture. This technology is expected to bring disruptive changes in computational imaging. However, the trajectory to a deployment of light fields remains cumbersome. Bottlenecks need to be alleviated before being able to fully exploit its potential. Barriers that CLIM addresses are the huge amount of high-dimensional (4D/5D) data produced by light fields, limitations of capturing devices, editing and image creation capabilities from compressed light fields. These barriers cannot be overcome by a simple application of methods which have made the success of digital imaging in past decades. The 4D/5D sampling of the geometric distribution of light rays striking the camera sensors imply radical changes in the signal processing chain compared to traditional imaging systems.

The ambition of CLIM is to lay new algorithmic foundations for the 4D/5D light fields processing chain, going from representation, compression to rendering. Data processing becomes tougher as dimensionality increases, which is the case of light fields compared to 2D images. This leads to the first challenge of CLIM that is the development of methods for low dimensional embedding and sparse representations of 4D/5D light fields. The second challenge is to develop a coding/decoding architecture for light fields which will exploit their geometrical models while preserving the structures that are critical for advanced image creation capabilities. CLIM targets ground-breaking solutions which should open new horizons for a number of consumer and professional markets (photography, augmented reality, light field microscopy, medical imaging, particle image velocimetry).

PLENOPTIMA

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

PLENOPTIMA project on cordis.europa.eu

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 UNIVERSITAT BERLIN (TUB), Germany
- TAMPEREEN KORKEAKOULUSAATIO SR (TAMPERE UNIVERSITY), Finland
- "INSTITUTE OF OPTICAL MATERIALS AND TECHNOLOGIES ""ACADEMICIAN JORDAN MALINOWSKI"" - BULGARIAN ACADEMY OF SCIENCES" (IOMT), Bulgaria

Inria contact: Christine Guillemot

Coordinator: Tampere University (Finland, Atanas Gotchev)

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 joints five of strong 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 will be registered in Joint/Double degree doctoral programmes at academic institutions in Bulgaria, Finland, France, Germany and Sweden. The programmes will be made sustainable through a set of measures in accordance with the Salzburg II Recommendations of the European University Association.

10.3 National initiatives

10.3.1 Project Action Exploratoire "Data Repurposing"

Participants: Anju Jose Tom, Thomas Maugey.

- Funding: Inria.

- Period: Sept. 2020 - Aug. 2023.

Lossy compression algorithms trade bits for quality, aiming at reducing as much as possible the bitrate needed to represent the original source (or set of sources), while preserving the source quality. In the exploratory action "DARE", we propose a novel paradigm of compression algorithms, aimed at minimizing the *information loss perceived by the final user* instead of the actual source quality loss, under compression rate constraints. In particular, we plan to measure the amount of information spanned by a data collection in the semantic domain. First, it enables to identify the high-level information contained in each of the image/video of a data collection. Second, it permits to take into account the redundancies and dissmilarities in the calculation of the global volume of information that is contained in a data collection. Finally, we propose to take into account the user's preferences in this calculation, since two users may have different tastes and priorities. Once the measure of information is set, we plan to build efficient sampling algorithms to reduce the data collection's size. This project also enables to explore new ideas for image generative compression, when part of the content can be "invented" at the decoder side.

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

Participants: Rita Fermanian, Christine Guillemot, Brandon Lebon, Guillaume Le Guludec, Mikael Le Pendu, Jinglei Shi.

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

- Period: Sept. 2020 - Aug. 2024.

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 based techniques designed to learn signal priors, tcan be used as regularization models.

10.3.3 CominLabs MOVE project: Mature Omnidirectional Video Exploration.

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

- Funding: Labex CominLabs.

- Period: Jan. 2021 - Dec. 2022.

This project aims to secure the industrial impact of the InterCom project, ended in Dec. 2020, and funded by the Labex CominLabs. Indeed, the goal of the Cominlabs InterCom project was to design novel compression algorithms to allow interactive communication between users and a server. One instance of this interactive scenario is a visual immersive experience, where a user can navigate freely in a 3D scene. This leads to tremendous amount of visual data, such that the whole scene cannot be sent to the user. Fortunately, the user watches only a part of the scene, such that only the requested part needs to be sent to the user. This however presents a great challenge from the point of view of the compression. Indeed, the data needs to be compressed once online, but decompressed in many manners, one manner being one requested point of view. In the case of a static 360 camera, the navigation has 3 Degrees of Freedom such that there is one decompression per value of a 3D-vector. One of the achievement of the InterCom project is a prototype for an interactive compression scheme for 360-degree images. The MOVE project helps the ADT ICOV project towards the construction of a full demonstrator. It also helps the maturation of the startup project lead by Navid Mahmoudian Bidgoli (ANAX).

10.3.4 CominLabs Colearn project: Coding for Learning

Participants: Thomas Maugey, Rémi Piau, Aline Roumy.

- Partners: Inria-Rennes (Sirocco team); LabSTICC, IMT Atlantique, (team Code and SI); 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.

10.3.5   Inria Start-up Studio: Anax

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

Two former PhD students of the team (Navid Mahmoudian Bidgoli and Simon Evain) plan to launch a startup, named Anax, on the theme of omnidirectional/image processing. They have obtained a one year grant (Sept 2021 - Aug 2022) for both of them funded by the Inria Startup Studio. Here is the description of the Anax project.

Anax aims to provide a deep tech software solution for processing 360-degree visual content with artificial intelligence (AI) specially designed for the preparation of virtual tours. Anax is aimed at various actors who wish to offer an immersive visit experience to improve their visibility, such as real estate agencies, cultural institutions, and interior designers. Anax is developing a technology that allows retrieving a faithful 3D reconstruction of a building from 360-degree images, opening up a wide range of applications based on AI image processing such as automatic recommendation of similar apartments, augmented reality, automatic inventory, etc. The envisaged solution, based on artificial intelligence, works even with consumer 360-degree capturing devices that are readily accessible to the general public.

10.3.6   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.

11   Dissemination

Participants: Christine Guillemot, Thomas Maugey, Aline Roumy.

11.1   Promoting scientific activities

11.1.1   Scientific events: organisation

Member of the organizing committees

- T. Maugey was member as publicity chair of the organizing committee of the IEEE International Conference on Image Processing, 16-19 Oct. 2022 (Bordeaux, France)

- T. Maugey was member as Area Chair of the organizing committee of the ACM International Conference on Multimedia 10-14 Oct. 2022 (Lisbon, Portugal)

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 2022 workshop on New Trends in Image Restoration and Enhancement (NTIRE).

- A. Roumy was a member of the technical program committee of the 2022 National Signal Processing workshop (colloque GRETSI).

11.1.2   Journal

Member of the editorial boards

- T. Maugey is Associate Editor of the EURASIP Journal of Advances in Signal Processing

- T. Maugey is Associate Editor of the IEEE Signal Processing Letters

- C. Guillemot has served as Associate Editor for the IEEE Signal Processing Magazine until Oct. 2022.

- A. Roumy is Associate Editor of the IEEE Transactions on Image Processing.

11.1.3   Invited talks

- T. Maugey gave a seminar at ENS Paris-Saclay on "Compression of visual data: beyond conventional approaches" May 2022.

- C. Guillemot gave a seminar on "distilled low-rank neural network optimization" at MidSweden Univ., Dec. 20, 2022.

- A. Roumy gave a seminar at ENS Rennes on "User-centric data compression and application to 360 videos", March 4, 2022.

- A. Roumy gave a seminar at ENSEEIHT Toulouse on "Neural Networks beyond the Euclidean case: omnidirectional images and application to compression", Oct. 20, 2022.

11.1.4   Leadership within the scientific community

- C. Guillemot has served as Chair of the evaluation panel of the computer vision centre of the Centers de Recerca de Catalunya (CERCA), Barcelona, Feb. 2022.

- A. Roumy is a member of the IEEE Image, Video, and Multidimensional Signal Processing Technical Committee (IVMSP TC).

- A. Roumy is a member of the Executive board of the National Research group in Image and Signal Processing (GRETSI).

- T. Maugey is a member of the European Association For Signal Processing (EURASIP) Technical Area Committee.

11.1.5   Scientific expertise

- C. Guillemot is a member of the Advisory Board of the Strategic Research Programme of the Vrije Universiteit Brussels

- C. Guillemot is member of the commission SEN-3 of the "Fond National de la recherche scientifique (FNRS), en Belgique.

- C. Guillemot has participated in the consolidator grant panel at the Swiss National Science Foundation (Dec. 2022).

- C. Guillemot has participated in the ANR committee selecting Franco-German research proposals in the field of artificial intelligence.

- A. Roumy has participated in the committee selecting research proposals funded by Innoviris (Institut bruxellois pour la Recherche et l'Innovation) Innovation and Research Institute, Bruxelles, Belgium.

- T. Maugey is a referee for the Italian National Agency for the Evaluation of the University and Research Systems (ANVUR)

11.1.6   Research administration

- C. Guillemot is member of the commission Recherche de l'Université de Rennes 1.

- C. Guillemot has been a member of the jury of the concours CRCN/ISFP of Rennes, April 2022.

- C. Guillemot has been a member of the jury for recruiting a Professor at the Univ. of Nancy, May 2022.

- A. Roumy has been a member of the Inria Delegation committee of the Rennes Bretagne Atlantique center, Febr. 2022.

- A. Roumy is a member of the Inria evaluation committee.

- A. Roumy has been a member of the jury of the concours CRCN/ISFP of Inria Sophia Antipolis Méditerranée, June 2022.

- T. Maugey served as a jury member of the PhD recruitment committee for Inria-Rennes, June 2022

11.2   Teaching - Supervision - Juries

- T. Maugey has given a Master course on 3D models in a module on advanced video, 8 hours, M2 SISEA, Univ. of Rennes 1, France.

- T. Maugey has given a Master course on Image compression, 10 hours, M2 SISEA, Univ. of Rennes 1, France.

- T. Maugey has given a Master course on Representation, editing and perception of digital images, 20 hours, M2 SIF, Univ. of Rennes 1, France.

- A. Roumy has given a Master course on High dimensional statistical learning, 9 hours, University Rennes 1, SIF Master, France.

- A. Roumy has given a Engineering degree course on Image and Video compression, 10 hours, University Rennes 1, ESIR, France.

11.2.1 Supervision

- Ipek Anil Atalay Appak (PhD in progress): *"Wavefront coding for computational plenoptic cameras"* (Started in August 2021, supervised by C. Guillemot and H. Caglayan(Tampere University)).

- Davi Rabbouni Freitas (PhD in progress): *"Sensing and Reconstruction of Plenoptic Point Clouds"* (started in September 2021, supervised by C.Guillemot and I. Tabus (Tampere University)).

- Denis Bacchus (PhD in progress): *"Deep Learning for selective lossy image compression"* (started in september 2020, supervised by A. Roumy and C. Guillemot).

- Tom Bachard (PhD in progress): *"Extreme compression of image/video databases using GAN-based synthesis"* (started in October 2021, supervised by T. Maugey).

- Tom Bordin (PhD in progress): *"Extreme generative video compression"* (Started in October 2022, supervised by T. Maugey).

- Nicolas Charpenay (PhD in progress): *"Combinatorics and strategic aspects in Information Theory"* (started in October 2020, supervised by M. le Treust and A. Roumy).

- Rita Fermanian (PhD in progress): *"Deep Learning for inverse problems and application to omnidirectional imaging"* (started in October 2020, supervised by C. Guillemot and M. Le Pendu (INTERDIGITAL, Rennes)).

- Kai Gu (PhD in progress): *"Spherical light field representation and reconstruction from omnidirectional imagery"* (Started in June 2021, supervised by C. Guillemot, T. Maugey and S. Knorr (Technische Universität Berlin)).

- Reda Kaafarani (PhD in progress): *"Optimisation of a multi-profile encoding system"* (started in April 2021, supervised by A. Roumy, T. Maugey, J. Le Tanou (MediaKind France) and M. Ropert (MediaKind France)).

- Brandon Le Bon (PhD in progress): *"Deep learning for light field acquisition and restoration"* (started in November 2020, supervised by C. Guillemot and M. Le Pendu (INTERDIGITAL, Rennes)).

- Arthur Lecert (PhD in progress): *"Low-light image enhancement with learning techniques"* (started in October 2020, supervised by C. Guillemot and A. Roumy).

- Yiqun Liu(PhD in progress): *"Learning for new generation video coders"* (started in Aug 2020, supervised by C. Guillemot, A. Roumy, and T. Guionnet (Ateme)).

- Rémi Piau(PhD in progress): *"Coding for learning"* (started in Oct 2021, supervised by A. Roumy, T. Maugey).

- Soheib Takhtardeshir (PhD in progress): *"Representation and compression of multi-sensor video for combined human and computer vision applications"* (Started in Jan 2022, supervised by C. Guillemot, M.Sjöström (Mid Sweden University) and co-supervised by R.Olsson(Mid Sweden University))

- Samuel Willingham (PhD in progress): *"Learning algorithms for inverse problems with modern imaging modalities"* (Started in June 2021, supervised by C. Guillemot and M.Sjöström (Mid Sweden University)).

11.2.2 Juries

- C. Guillemot was member, as chair, of the PhD jury of of F. Nasiri, Univ. Rennes 1, May 2022.

- C. Guillemot was member, as rapporteur, of the PhD jury of B. Chiche, Univ. Paris-Saclay, Oct. 2022

- A. Roumy was member, as rapporteure, of the HDR jury of A. Goupil, Univ. Reims Champagne-Ardenne, April 2022.

- A. Roumy was member, as rapporteure, of the PhD jury of M. Aklouf, Univ. Paris-Saclay, May 2022.

- A. Roumy was member, as chair, of the PhD jury of C. Bonnineau, Univ. Rennes 1, May 2022.

- A. Roumy was member of the PhD jury of L. Chetot, Univ. Lyon, July 2022.

- A. Roumy was member, as rapporteure, of the PhD jury of V. Alves De Oliveira, Univ. Toulouse, Oct. 2022.

- A. Roumy was member, as rapporteure, of the PhD jury of R. Bou Rouphael, Univ. Paris-Cergy, Dec. 2022.

- A. Roumy was member, of the PhD jury of A. Marmoret, Univ. Rennes 1, Dec. 2022.

- A. Roumy was member, as chair, of the PhD jury of B. Hamoum, Univ. Bretagne Sud, Dec. 2022.

- T. Maugey was member of the PhD jury of Sylvia Rossi, University College of London (UCL), London, UK, Jan. 2022

11.3   Popularization

- T. Maugey is scientific mediation officer of the Inria Rennes Scientific mediation team

- T. Maugey is member of a working group on regulation and ethic

- T. Maugey is member of working group on carbon footprint evaluation of research activities

11.3.1   Education

- T. Maugey presented the research profession at Lycée Brequigny in the context of the national "Chiche" program (March, 2022)

- A. Roumy presented her research career at the Inria/Irisa women-only group meeting, June 2022.

11.3.2   Interventions

- T. Maugey has collaborated with the artist Mustapha Azeroual for the project aRza, exhibited at the Jeu de Paume, France, June-Sept. 2022.

12   Scientific production

12.1   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. 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. 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. 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. 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. URL: https://hal-imt-atlantiq ue.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. URL: https://hal.archives-ouvertes.fr/hal-02199839.

12.2 Publications of the year

International journals

[7] H. Amirpour, C. Guillemot, M. Ghanbari and C. Timmerer. 'Advanced Scalability for Light Field Image Coding'. In: *IEEE Transactions on Image Processing* (Nov. 2022), pp. 1–15. URL: https://hal .archives-ouvertes.fr/hal-03848700.

[8] R. Fermanian, M. Le Pendu and C. Guillemot. 'PnP-ReG: Learned Regularizing Gradient for Plug-and-Play Gradient Descent'. In: *SIAM Journal on Imaging Sciences* (Nov. 2022), pp. 1–28. URL: https://hal.archives-ouvertes.fr/hal-03853961.

[9] P. Garus, F. Henry, J. Jung, T. Maugey and C. Guillemot. 'Immersive Video Coding: Should Geometry Information be Transmitted as Depth Maps?' In: *IEEE Transactions on Circuits and Systems for Video Technology* 32.5 (May 2022), pp. 3250–3264. DOI: 10.1109/TCSVT.2021.3100006. URL: https://hal.archives-ouvertes.fr/hal-03303040.

[10] X. Jiang, J. Shi and C. Guillemot. 'An Untrained Neural Network Prior for Light Field Compression'. In: *IEEE Transactions on Image Processing* (Oct. 2022), pp. 1–15. URL: https://hal.archives-ou vertes.fr/hal-03820927.

[11] G. Le Guludec and C. Guillemot. 'Deep Unrolling for Light Field Compressed Acquisition using Coded Masks'. In: *IEEE Access* (Apr. 2022), pp. 1–17. DOI: 10.1109/access.2022.3168362. URL: https://hal.archives-ouvertes.fr/hal-03643112.

[12] M. Le Pendu and C. Guillemot. 'Preconditioned Plug-and-Play ADMM with Locally Adjustable Denoiser for Image Restoration Mikael'. In: *SIAM Journal on Imaging Sciences* (Nov. 2022), pp. 1–30. URL: https://hal.archives-ouvertes.fr/hal-03857826.

[13] N. Mahmoudian Bidgoli, R. Azevedo, T. Maugey, A. Roumy and P. Frossard. 'OSLO: On-the-Sphere Learning for Omnidirectional images and its application to 360-degree image compression'. In: *IEEE Transactions on Image Processing* (2022), pp. 1–16. URL: https://hal.inria.fr/hal-0383 3562.

[14] J. Shi, X. Jiang and C. Guillemot. 'Deep Residual Architecture Using Pixel and Feature Cues for View Synthesis and Temporal Interpolation'. In: *IEEE Transactions on Computational Imaging* (Mar. 2022), pp. 1–14. DOI: 10.1109/tci.2022.3160671. URL: https://hal.archives-ouvertes.f r/hal-03607417.

[15] N. Thomos, T. Maugey and L. Toni. 'Machine Learning for Multimedia Communications'. In: *Sensors* 22.3 (21st Jan. 2022), pp. 1–31. DOI: 10.3390/s22030819. URL: https://hal.inria.fr/hal-03 541327.

[16] T. Tongbuasirilai, J. Unger, C. Guillemot and E. Miandji. 'A Sparse Non-parametric BRDF Model'. In: *ACM Transactions on Graphics* (May 2022), pp. 1–18. DOI: 10.1145/3533427. URL: https://h al.archives-ouvertes.fr/hal-03654734.

[17] Z. Xiao, J. Shi, X. Jiang and C. Guillemot. 'Axial refocusing precision model with light fields'. In: *Signal Processing: Image Communication* (May 2022), pp. 1–11. DOI: 10.1016/j.image.2022.11 6721. URL: https://hal.archives-ouvertes.fr/hal-03651953.

[18] M. Zhang, H. Jin, Z. Xiao and C. Guillemot. 'A Light Field FDL-HCGH Feature in Scale-Disparity Space'. In: *IEEE Transactions on Image Processing* (Sept. 2022), pp. 1–12. URL: https://hal.arch ives-ouvertes.fr/hal-03763691.

International peer-reviewed conferences

[19]   H. Amirpour, C. Guillemot and C. Timmerer. 'FuRA: Fully Random Access Light Field Image Compression'. In: European Workshop on Video and Image Processing. Proceedings EUVIP Conf. Lisbo, Portugal, Sept. 2022, pp. 1–6. URL: https://hal.archives-ouvertes.fr/hal-0375889 8.

[20]   P. Bacchus, R. Fraisse, A. Roumy and C. Guillemot. 'Quasi Lossless Satellite Image Compression'. In: IGARSS 2022 - International Geoscience and Remote Sensing Symposium. Proceedings. Kuala Lumpur, Indonesia, 17th July 2022, pp. 1–4. URL: https://hal.inria.fr/hal-03807986.

[21]   T. Bachard, A. J. Tom and T. Maugey. 'Semantic alignment for multi-item compression'. In: ICIP 2022 - IEEE International Conference on Image Processing. Bordeaux, France: IEEE, 16th Oct. 2022, pp. 1–5. URL: https://hal.archives-ouvertes.fr/hal-03760650.

[22]   N. Charpenay, M. Le Treust and A. Roumy. 'Zero-error source coding when side information may be present'. In: IZS 2022 - International Zurich Seminar on Information and Communication. Zurich, Switzerland, 2022, pp. 1–5. URL: https://hal.archives-ouvertes.fr/hal-03290860.

[23]   P. Garus, F. Henry, T. Maugey and C. Guillemot. 'Decoder Side Multiplane Images using Geometry Assistance SEI for MPEG Immersive Video'. In: MMSP 2022 - IEEE 24th International Workshop on Multimedia Signal Processing. Shanghai, China: IEEE, 26th Sept. 2022, pp. 1–6. URL: https://hal.inria.fr/hal-03833545.

[24]   P. Garus, F. Henry, T. Maugey and C. Guillemot. 'Motion Compensation-based Low-Complexity Decoder Side Depth Estimation for MPEG Immersive Video'. In: MMSP 2022 - IEEE 24th International Workshop on Multimedia Signal Processing. Shanghai, China: IEEE, 26th Sept. 2022, pp. 1–6. URL: https://hal.inria.fr/hal-03833539.

[25]   K. Gu, T. Maugey, S. Knorr and C. Guillemot. 'Omni-nerf: neural radiance field from 360° image captures'. In: ICME 2022 - IEEE international conference on multimedia and expo. Taipei, Taiwan: IEEE, 18th July 2022, pp. 1–6. URL: https://hal.inria.fr/hal-03646688.

[26]   A. Lecert, R. Fraisse, A. Roumy and C. Guillemot. 'A new regularization for retinex decomposition of low-light images'. In: IEEE International Conference on Image Processing. Proceedings. Bordeaux, France: IEEE, Oct. 2022, pp. 1–4. URL: https://hal.archives-ouvertes.fr/hal-03806390.

[27]   Y. Liu, M. Abdoli, T. Guionnet, C. Guillemot and A. Roumy. 'Light-weight cnn-based vvc inter partitioning acceleration'. In: IVMSP 2022 - IEEE 14th Image, Video, and Multidimensional Signal Processing Workshop. Nafplio, Greece: IEEE, 26th June 2022, pp. 1–5. URL: https://hal.inria.fr/hal-03808305.

[28]   Y. Liu, M. Abdoli, T. Guionnet, C. Guillemot and A. Roumy. 'Statistical analysis of inter coding in vvc test model (VTM)'. In: ICIP 2022 - IEEE International Conference on Image Processing. Bordeaux, France: IEEE, 16th Oct. 2022, pp. 1–4. URL: https://hal.inria.fr/hal-03808292.

[29]   J. Shi, X. Jiang and C. Guillemot. 'Deep Video Frame Rate Up-conversion Network using Feature-based Progressive Residue Refinement a'. In: VISAPP 2022 - 17th International Conference on Computer Vision Theory and Applications. online, France, 6th Feb. 2022, pp. 1–9. URL: https://hal.archives-ouvertes.fr/hal-03432380.

National peer-reviewed Conferences

[30]   N. Mahmoudian Bidgoli, R. Azevedo, T. Maugey, A. Roumy and P. Frossard. 'Méthode d'apprentissage opérant sur la sphère: Application à la compression d'images omnidirectionnelles'. In: GRETSI 2022 - XXVIIIème Colloque Francophone de Traitement du Signal et des Images. Nancy, France, 6th Sept. 2022, pp. 1–4. URL: https://hal.inria.fr/hal-03866225.

Conferences without proceedings

[31]   N. Charpenay, M. Le Treust and A. Roumy. 'Codage de source zéro-erreur pour un réseau de Gray-Wyner lorsque l'information adjacente peut être présente'. In: XXVIIIème Colloque Francophone de Traitement du Signal et des Images (GRETSI'22). Nancy, France, 6th Sept. 2022, pp. 1–4. URL: https://hal.archives-ouvertes.fr/hal-03866341.

Doctoral dissertations and habilitation theses

[32]   T. Maugey. 'Visual data compression: beyond conventional approaches'. Université de Rennes 1, 27th June 2022. URL: https://hal.inria.fr/tel-03754139.

Reports & preprints

[33]   N. Charpenay, M. Le Treust and A. Roumy. *Zero-Error Coding for Computing with Encoder Side-Information.* 15th Dec. 2022. URL: https://hal.archives-ouvertes.fr/hal-03901563.