



Activity Report 2018

Team SIROCCO

Analysis Representation, Compression and Communication of Visual Data

Joint team with Inria Rennes – Bretagne Atlantique

D5 – Digital Signals and Images, Robotics



Table of contents

1. Team, Visitors, External Collaborators	1
2. Overall Objectives	2
2.1. Introduction	2
2.2. Visual Data Analysis	2
2.3. Signal processing and learning methods for visual data representation and compression	2
2.4. Algorithms for inverse problems in visual data processing	3
2.5. Distributed coding for interactive communication	3
3. Research Program	3
3.1. Introduction	3
3.2. Data Dimensionality Reduction	3
3.3. Deep neural networks	4
3.4. Coding theory	5
4. Application Domains	6
4.1. Overview	6
4.2. Compression of emerging imaging modalities	6
4.3. Networked visual applications	6
4.4. Editing, post-production and computational photography	7
5. Highlights of the Year	7
6. New Software and Platforms	7
6.1. QuantizationAE	7
6.2. LF-Inpainting	7
6.3. LF-HLRA	8
6.4. GBR-MVimages	8
6.5. FastLFInpainting	8
6.6. M360CT	8
6.7. Platforms	9
6.7.1. Light field editor	9
6.7.2. Acquisition of multi-view sequences for Free viewpoint Television	9
6.7.3. Light fields datasets	9
7. New Results	10
7.1. Visual Data Analysis	10
7.1.1. Super-rays for efficient light fields processing	10
7.1.2. Scene depth estimation from light fields	10
7.1.3. Scene flow estimation from light fields	11
7.2. Signal processing and learning methods for visual data representation and compression	11
7.2.1. Multi-shot single sensor light field camera using a color coded mask	11
7.2.2. Compressive 4D light field reconstruction	12
7.2.3. Light fields dimensionality reduction with low-rank models	12
7.2.4. Graph-based transforms for light fields and omni-directional image compression	13
7.2.5. Neural networks for learning image transforms and predictors	13
7.2.6. Cloud-based predictors and neural network temporal predictors video compression	14
7.3. Algorithms for inverse problems in visual data processing	14
7.3.1. View synthesis in light fields and stereo set-ups	14
7.3.2. Light field inpainting and restoration	14
7.3.3. High dynamic range light fields capture	15
7.4. Distributed processing and robust communication	15
7.4.1. Information theoretic bounds for sequential massive random access to large database of correlated data	15
7.4.2. Correlation model selection for interactive video communication	16

7.4.3.	Compression of spatio-temporally correlated and massive georeferenced data	16
7.4.4.	ICON 3D - Interactive CODing for Navigation in 3D scenes	16
8.	Bilateral Contracts and Grants with Industry	17
8.1.1.	CIFRE contract with Technicolor on light fields editing	17
8.1.2.	CIFRE contract with Technicolor on light fields compressed representation	17
8.1.3.	CIFRE contract with Technicolor on cloud-based image compression	17
8.1.4.	DGA contract on deep learning for image compression	18
9.	Partnerships and Cooperations	18
9.1.	Regional Initiatives	18
9.1.1.	ICON 3D - Interactive CODing for Navigation in 3D scenes	18
9.1.2.	CominLabs InterCom project	18
9.2.	European Initiatives	19
9.3.	International Initiatives	19
9.3.1.	Inria International Labs	19
9.3.2.	Inria International Partners	20
9.4.	International Research Visitors	20
10.	Dissemination	20
10.1.	Promoting Scientific Activities	20
10.1.1.	Scientific Events Organisation	20
10.1.2.	Scientific Events Selection	20
10.1.3.	Journal	20
10.1.4.	Invited Talks	20
10.1.5.	Leadership within the Scientific Community	21
10.1.6.	Research Administration	21
10.2.	Teaching - Supervision - Juries	21
10.2.1.	Teaching	21
10.2.2.	Juries	22
10.3.	Popularization	22
11.	Bibliography	22

Project-Team SIROCCO

Creation of the Project-Team: 2012 January 01

Keywords:

Computer Science and Digital Science:

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

Other Research Topics and Application Domains:

- B6. - IT and telecom

1. Team, Visitors, External Collaborators

Research Scientists

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

Post-Doctoral Fellows

Christian Galea [Inria, from Feb 2018]
Mai Quyen Pham [Inria, from Apr 2018]
Lara Younes [Inria, until Aug 2018]

PhD Students

Jean Begaint [Technicolor, until Oct 2018]
Pierre David [Inria]
Elian Dib [Inria]
Thierry Dumas [Inria, until Oct 2018]
Simon Evain [Inria]
Fatma Hawary [Technicolor]
Matthieu Hog [Technicolor, until Aug 2018]
Navid Mahmoudian Bidgoli [Inria]
Mira Rizkallah [Univ de Rennes I]
Jinglei Shi [Inria]

Technical staff

Pierre Allain [Inria]
Laurent Guillo [CNRS]
Xiaoran Jiang [Inria]
Cédric Le Cam [Inria, until Sep 2018]
Hoai Nam Nguyen [Inria, from Nov 2018]
Xin Su [Inria, until Apr 2018]

Administrative Assistant

Huguette Bechu [Inria]

Visiting Scientist

Zhaolin Xiao [Inria, from Dec 2018]

2. Overall Objectives

2.1. Introduction

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

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

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

While aiming at generic approaches, some of the solutions developed are applied to practical problems in partnership with industry (Technicolor, Ericsson, Orange) or in the framework of national projects. The application domains addressed by the project are networked visual applications taking into account their various requirements and needs in terms of compression, of network adaptation, of advanced functionalities such as navigation, interactive streaming and high quality rendering.

2.2. Visual Data Analysis

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

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

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

2.4. Algorithms for inverse problems in visual data processing

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

2.5. Distributed coding for interactive communication

The availability of wireless camera sensors has also been spurring interest for a variety of applications ranging from scene interpretation, object tracking and security environment monitoring. In such camera sensor networks, communication energy and bandwidth are scarce resources, motivating the search for new distributed image processing and coding solutions suitable for band and energy limited networking environments. Our goal is to address theoretical issues such as the problem of modeling the correlation channel between sources, and to practical coding solutions for distributed processing and communication and for interactive streaming.

3. Research Program

3.1. Introduction

The research activities on analysis, compression and communication of visual data mostly rely on tools and formalisms from the areas of statistical image modelling, of signal processing, of machine learning, of coding and information theory. Some of the proposed research axes are also based on scientific foundations of computer vision (e.g. multi-view modelling and coding). We have limited this section to some tools which are central to the proposed research axes, but the design of complete compression and communication solutions obviously rely on a large number of other results in the areas of motion analysis, transform design, entropy code design, etc which cannot be all described here.

3.2. Data Dimensionality Reduction

Manifolds, graph-based transforms, compressive sensing

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

Sparse representations, compressive sensing, and dictionary learning have been shown to be powerful tools for efficient processing of visual data. The objective of *sparse representations* is to find a sparse approximation of a given input data. In theory, given $A \in \mathbb{R}^{m \times n}$, and $\mathbf{b} \in \mathbb{R}^m$ with $m \ll n$ and A is of full row rank, one seeks the solution of $\min\{\|\mathbf{x}\|_0 : A\mathbf{x} = \mathbf{b}\}$, where $\|\mathbf{x}\|_0$ denotes the L_0 norm of x , i.e. the number of non-zero components in z . A is known as the dictionary, its columns a_j are the atoms, they are assumed to be normalized in Euclidean norm. There exist many solutions x to $Ax = b$. The problem is to find the sparsest solution x , i.e. the one having the fewest non zero components. In practice, one actually seeks an approximate and thus even sparser solution which satisfies $\min\{\|\mathbf{x}\|_0 : \|A\mathbf{x} - \mathbf{b}\|_p \leq \rho\}$, for some $\rho \geq 0$, characterizing an admissible reconstruction error.

The recent theory of *compressive sensing*, in the context of discrete signals, can be seen as an effective dimensionality reduction technique. The idea behind compressive sensing is that a signal can be accurately recovered from a small number of linear measurements, at a rate much smaller than what is commonly prescribed by the Shannon-Nyquist theorem, provided that it is sparse or compressible in a known basis. Compressed sensing has emerged as a powerful framework for signal acquisition and sensor design, with a number of open issues such as learning the basis in which the signal is sparse, with the help of dictionary learning methods, or the design and optimization of the sensing matrix. The problem is in particular investigated in the context of light fields acquisition, aiming at novel camera design with the goal of offering a good trade-off between spatial and angular resolution.

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

3.3. Deep neural networks

Autoencoders, Neural Networks, Recurrent Neural Networks

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

An autoencoder is a neural network with an encoder g_e , parametrized by θ , that computes a representation Y from the data X , and a decoder g_d , parametrized by ϕ , that gives a reconstruction \hat{X} of X (see Figure below). Autoencoders can be used for dimensionality reduction, compression, denoising. When it is used for compression, the representation need to be quantized, leading to a quantized representation $\hat{Y} = Q(Y)$ (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.

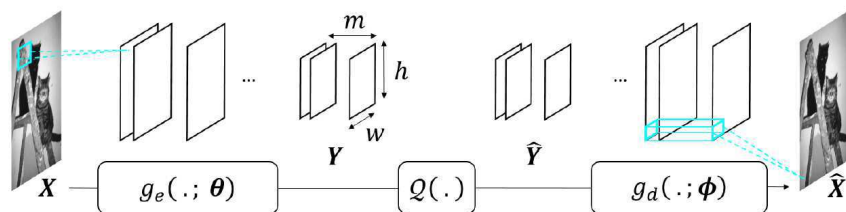


Figure 1. Illustration of an autoencoder.

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

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

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

3.4. Coding theory

OPTA limit (Optimum Performance Theoretically Attainable), Rate allocation, Rate-Distortion optimization, lossy coding, joint source-channel coding multiple description coding, channel modelization, oversampled frame expansions, error correcting codes.

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

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

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

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

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

4. Application Domains

4.1. Overview

The application domains addressed by the project are:

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

4.2. Compression of emerging imaging modalities

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

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

4.3. Networked visual applications

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

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

4.4. Editing, post-production and computational photography

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

5. Highlights of the Year

5.1. Highlights of the Year

5.1.1. Awards

- The paper [20] has received the best student paper award at the European Signal Processing (EUSIPCO) conference, Roma, Sept. 2018.
- The video light field dataset captured by the team has been retained by MPEG-I as test dataset (April 2018) [24].

BEST PAPER AWARD:

[20]

M. RIZKALLAH, F. DE SIMONE, T. MAUGEY, C. GUILLEMOT, P. FROSSARD. *Rate Distortion Optimized Graph Partitioning for Omnidirectional Image Coding*, in "EUSIPCO 2018 - 26th European Signal Processing Conference", Rome, Italy, September 2018, pp. 1-5, <https://hal.inria.fr/hal-01807613>

6. New Software and Platforms

6.1. QuantizationAE

KEYWORDS: Compression - Machine learning

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

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

6.2. LF-Inpainting

Light field inpainting based on a low rank model

KEYWORDS: Light fields - Low rank models - Inpainting

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

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

6.3. LF-HLRA

Light fields homography-based low rank approximation

KEYWORDS: Compression - Light fields - Low rank models - Dimensionality reduction

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

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

6.4. GBR-MVimages

Graph-based representation for multi-view and light field images

KEYWORDS: Light fields - Multi-View reconstruction - Graph

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

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

6.5. FastLFInpainting

Fast light field inpainting

KEYWORDS: Inpainting - Light fields

FUNCTIONAL DESCRIPTION: This software implements a method for fast and efficient inpainting of light fields. It computes disparity based on smoothed structure tensors, that is then used for propagating, by angular warping, the inpainted texture of one view to the entire light field. The code allows inpainting a light field of 80 views in a few seconds (from 4 to 15s, depending on the size of the region to be inpainted). The software has been registered at APP under the number IDDN.FR.001.290017.000.S.P.2018.000.21000.

- Participants: Pierre Allain, Laurent Guillo and Christine Guillemot
- Contact: Laurent Guillo

6.6. M360CT

Multi-360 Calibration Toolkit

KEYWORDS: Omnidirectional camera - Calibration - FTV - 6DoF

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

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

6.7. Platforms

6.7.1. *Light field editor*

Participants: Pierre Allain, Laurent Guillo, Christine Guillemot.

As part of the ERC Clim project, the EPI Sirocco is developing a light field editor, a tool analogous to traditional image editors such as the GNU image manipulation program Gimp or the raster graphic editor Photoshop but dedicated to light fields. As input data, this tool accepts for instance sparse light fields acquired with High Density Camera Arrays (HDCA) or denser light fields captured with microlens array (MLA). Two kinds of features are provided. Traditional features such as changing the angle of view, refocusing or depth map extraction are or will be soon supported. More advanced features are being integrated in our tool as libraries we have developed, such as inpainting to support light field manipulations like object removal, and denoising in the 4D ray space. The next steps are to integrate libraries enabling scene depth estimation and view synthesis. The tool and libraries are developed in C++ and the graphical user interface relies on Qt.

6.7.2. *Acquisition of multi-view sequences for Free viewpoint Television*

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

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

In the project ATeP (funded by InriaHub), we have developed a novel acquisition procedure relying on forty synchronized omnidirectional cameras. The captured content allows an omni-directional visualization of the scene at a set of discrete viewpoints corresponding to the pre-defined camera positions. We also propose a calibration technique to estimate the position and orientation of each camera with respect to a same reference. This solution relies on a calibration of each individual camera, and a graph-based synchronization of all the estimated parameters.

Based on these developed tools, we have built a complete dataset that we share on the following website <https://project.inria.fr/ftv360>. Our dataset is made of two different captures (indoor and outdoor), with, in total 8 different sequences (each of them having 40 synchronized videos of 1 to 4 min long). The calibration parameters are shared with the calibration toolkit that was developed during the project. These data can serve for the development of new tools for FTV, such as: view synthesis, depth estimation, super resolution, inpainting, etc.

6.7.3. *Light fields datasets*

Participants: Pierre Allain, Christine Guillemot, Laurent Guillo.

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

Natural content has been created with Lytro cameras (the original first generation Lytro and the Lytro Illum). The team also owns a R8 Raytrix plenoptic cameras with which still and video contents have been captured. Applications taking advantage of the Raytrix API have been developed to extract views from the Raytrix lightfield. The number of views per frame is configurable and can be set for instance to 3x3 or 9x9 according to the desired sparsity. A dataset of video light fields captured by our raytrix R8 camera has been proposed to the MPEG-I standardization group and retained for test purposes [24].

Synthetic content exists for dense light fields with small baselines. To address issues of scene depth estimation and of view synthesis in more difficult configurations like in the case of large baselines, we have produced two datasets that we use for training neural networks for scene depth estimation from light fields with small and large baselines. Most of our rendered light field scenes are indoor scenes, with light reflection and diffusion on the object surfaces to make them more realistic. Both dense and sparse light fields of 9×9 views of 512×512 pixels have been rendered from the input 3D models, with a disparity range of $[-20, +20]$ for sparse light fields and $[-4, +4]$ for dense light fields. The dense and sparse light fields datasets contains 43 and 53 scene respectively. They are provided together with the ground truth depth maps.

Similarly, as no publicly available dataset exist for video light fields, we have produced our own data set from the Sintel film (<https://durian.blender.org/download/>), which is a short computer animated film by the Blender institute, part of the Blender Foundation. A specific Blender add-on is used to extract views from a frame. As previously, the number of views is configurable. Synthetic contents present the advantage to provide a ground truth useful to evaluate how accurate our algorithms are to compute, for instance, the depth maps and the scene flows. At the moment, the dataset contains two synthetic video light fields of 50 frames.

All these contents are made available via the project web site: <http://clim.inria.fr/DataSoftware.html>

7. New Results

7.1. Visual Data Analysis

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

7.1.1. *Super-rays for efficient light fields processing*

Participants: Matthieu Hog, Christine Guillemot.

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

7.1.2. *Scene depth estimation from light fields*

Participants: Christian Galea, Christine Guillemot, Xiaoran Jiang, Jinglei Shi.

While there exist scene depth and scene flow estimation methods, these methods, mostly designed for stereo content or for pairs of rectified views, do not effectively apply to new imaging modalities such as light fields. We have focused on the problem of *scene depth estimation* for every viewpoint of a dense light field, exploiting information from only a sparse set of views [17]. This problem is particularly relevant for applications such as light field reconstruction from a subset of views, for view synthesis, for 3D modeling and for compression. Unlike most existing methods, the proposed algorithm computes disparity (or equivalently depth) for every viewpoint taking into account occlusions. In addition, it preserves the continuity of the depth space and does not require prior knowledge on the depth range. Experiments show that, both for synthetic and real light fields, our algorithm achieves competitive performance compared to state-of-the-art algorithms which exploit the entire light field and usually generate the depth map for the center viewpoint only. Figure 2 shows the estimated depth map for a synthetic light field in comparison with the ground truth. The estimated depth maps allow us to construct accurate 3D point clouds of the captured scene [16]. This work is now pursued considering deep learning solutions.



Figure 2. Estimated depth map (middle) for the light field 'Buddha' in comparison with the ground truth (right).

7.1.3. Scene flow estimation from light fields

Participants: Pierre David, Christine Guillemot.

Temporal processing of dynamic 3D scenes requires estimating the displacement of the objects in the 3D space, i.e., so-called scene flows. Scene flows can be seen as 3D extensions of optical flows by also giving the variation in depth along time in addition to the optical flow. Estimating dense scene flows in light fields pose obvious problems of complexity due to the very large number of rays or pixels. This is even more difficult when the light field is sparse, i.e., with large disparities, due to the problem of occlusions. We have addressed the complexity problem by designing a sparse estimation method followed by a densification step that avoids the difficulty of computing matches in occluded areas. The developments in this area are also made difficult due to the lack of test data, i.e., there is no publicly available synthetic video light fields with the corresponding ground truth scene flows. In order to be able to assess the performance of the proposed method, we have therefore created synthetic video light fields from the MPI Sintel dataset. This video light field data set has been produced with the Blender software by creating new production files placing multiple cameras in the scene, controlling the disparity between the set of views.

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

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

7.2.1. Multi-shot single sensor light field camera using a color coded mask

Participant: Christine Guillemot.

In collaboration with the University of Linköping (Prof. J. unger, Dr. E. Miandji), we have proposed a compressive sensing framework for reconstructing a light field from a single-sensor consumer camera capture with color coded masks [19]. The proposed *camera architecture* captures incoherent measurements of the light field via a controllable color mask placed in front of the sensor. To enhance the incoherence, hence the reconstruction quality, we propose to utilize multiple shots where, for each shot, the mask configuration is changed to create a new random pattern. To reduce computations and increase the incoherence, we also perform a random sampling of the spatial domain. The compressive sensing framework relies on a dictionary trained over a light field data set. Numerical simulations show significant improvements compared with a similar coded aperture system for light field capture.

7.2.2. Compressive 4D light field reconstruction

Participants: Christine Guillemot, Fatma Hawary.

Exploiting the assumption that light field data is sparse in the Fourier domain, we have also developed a new method for reconstructing a 4D light field from a random set of measurements [14]. The reconstruction algorithm searches for these bases (i.e., their frequencies) which best represent the 4D Fourier spectrum of the sampled light field. The method has been further improved by introducing an orthogonality constraint on the residue, in the same vein as orthogonal matching pursuit but in the Fourier transform domain, as well as a refinement for non integer frequencies. The method achieves a very high reconstruction quality, in terms of PSNR (more than 1dB gain compared to state-of-the-art algorithms).

7.2.3. Light fields dimensionality reduction with low-rank models

Participants: Elian Dib, Christine Guillemot, Xiaoran Jiang.

We have further investigated low-rank approximation methods exploiting data geometry for dimensionality reduction of light fields. While our first solution was considering global low-rank models based on homographies, we have recently developed local low-rank models exploiting disparity. The local support of the approximation is given by super-rays (see section 7.1.1). The super-rays group super-pixels which are consistent across the views while being constrained to be of same shape and size. The corresponding super-pixels in all views are found thanks to disparity compensation. In order to do so, a novel method has been proposed to estimate the disparity for each super-ray using a low rank prior, so that the super-rays are constructed to yield the lowest approximation error for a given rank. More precisely, the disparity for each super-ray is found in order to align linearly correlated sub-aperture images in such a way that they can be approximated by the considered low rank model. The rank constraint is expressed as a product of two matrices, where one matrix contains basis vectors (or eigen images) and where the other one contains weighting coefficients. The eigen images are actually splitted into two sets, one corresponding to light rays visible in all views and a second one, very sparse, corresponding to occluded rays (see Fig. 3). A light field compression algorithm has been designed encoding the different components of the resulting low rank approximation.

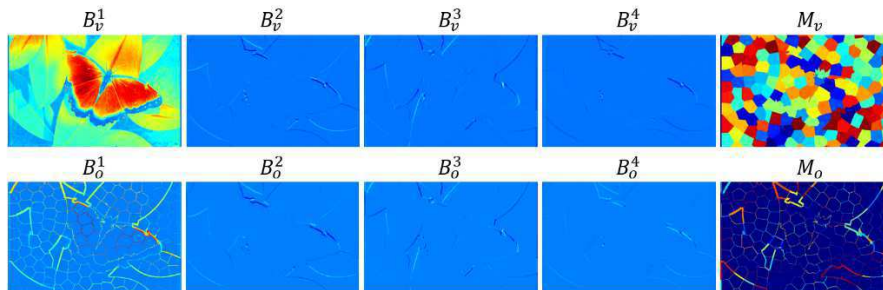


Figure 3. Eigen-images and segmentation maps for visible and occluded sets of pixels.

7.2.4. Graph-based transforms for light fields and omni-directional image compression

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

Graph-based transforms are interesting tools for low-dimensional embedding of light field data. This embedding can be learned with a few eigenvectors of the graph Laplacian. However, the dimension of the data (e.g., light fields) has obvious implications on the storage footprint of the Laplacian matrix and on the eigenvectors computation complexity, making graph-based *non separable* transforms impractical for such data. To cope with this difficulty, in [21], we have first developed *local super-rays based separable (spatial followed by angular)* weighted and unweighted transforms to jointly capture light fields correlation spatially and across views. While separable transforms on super-rays allow us to significantly decrease the eigenvector computation complexity, the basis functions of the spatial graph transforms to be applied on the super-ray pixels of each view are often not compatible, resulting in decreased correlation of the coefficients across views, hence in a loss of performance of the angular transform, compared to the non-separable case. We have therefore developed a graph construction optimization procedure which seeks to find the eigen-vectors having the best alignment with those computed on a reference frame while still approximately diagonalizing their respective Laplacians. Fig.4 shows the second eigenvector of different super-pixels belonging to the same super-ray before and after optimization. A rate-distortion optimized graph partitioning algorithm has also been developed [20] for coding 360° videos signals, to achieve a good trade-off between distortion, smoothness of the signal on each subgraph, and the coding cost of the graph partition.

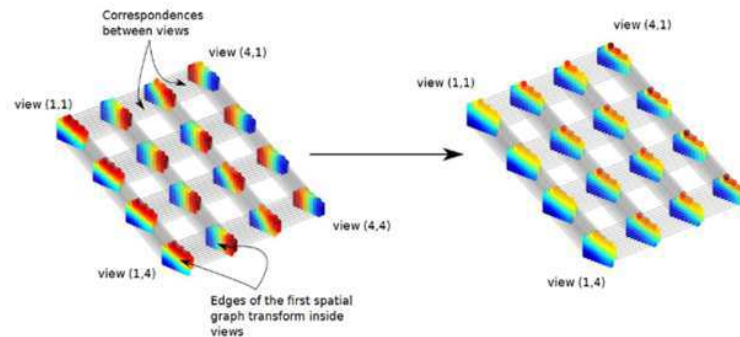


Figure 4. Second eigenvector of super-pixels forming a super-ray before and after optimization.

7.2.5. Neural networks for learning image transforms and predictors

Participants: Thierry Dumas, Christine Guillemot, Aline Roumy.

We have explored the problem of learning transforms for image compression via autoencoders. Learning a transform is equivalent to learning an autoencoder, which is of its essence unsupervised and therefore more difficult than classical supervised learning. In compression, the learning has in addition to be performed under a rate-distortion criterion, and not only a distortion criterion. Usually, the rate-distortion performances of image compression are tuned by varying the quantization step size. In the case of autoencoders, this in principle would require learning one transform per rate-distortion point at a given quantization step size. We have shown in [12] that comparable performances can be obtained with a unique learned transform. The different rate-distortion points are then reached by varying the quantization step size at test time. This approach saves a lot of training time.

Another important operator in compression algorithm is the predictor that aims at capturing spatial correlation. We have developed a set of neural network architectures, called Prediction Neural Networks Set (PNNS), based on both fully-connected and convolutional neural networks, for intra image prediction. It is shown that, while

fully-connected neural networks give good performances for small block sizes, convolutional neural networks provide better predictions in large blocks with complex textures. Thanks to the use of masks of random sizes during training, the neural networks of PNNS well adapt to the available context that may vary, depending on the position of the image block to be predicted. Unlike the H.265 intra prediction modes, which are each specialized in predicting a specific texture, the proposed PNNS can model a large set of complex textures.

7.2.6. *Cloud-based predictors and neural network temporal predictors video compression*

Participants: Jean Begaint, Christine Guillemot.

Video codecs are primarily designed assuming that rigid, block-based, two-dimensional displacements are suitable models to describe the motion taking place in a scene. However, translational models are not sufficient to handle real world motion such as camera zoom, shake, pan, shearing or changes in aspect ratio. Building upon the region-based geometric and photometric model proposed in [5] to exploit correlation between images in the cloud, we have developed a region-based inter-prediction scheme for video compression. The proposed predictor is able to estimate multiple homography models in order to predict complex scene motion. We also introduce an affine photometric correction to each geometric model. Experiments on targeted sequences with complex motion demonstrate the efficiency of the proposed approach compared to the state-of-the-art HEVC video codec [11]. To further improve the accuracy of the temporal predictor, we have explored the use of deep neural networks for frame prediction and interpolation, and preliminary results have shown gains going up to 5% compared with the latest HEVC video codec.

7.3. Algorithms for inverse problems in visual data processing

Inpainting, view synthesis, super-resolution

7.3.1. *View synthesis in light fields and stereo set-ups*

Participants: Simon Evain, Christine Guillemot, Matthieu Hog, Xiaoran Jiang.

We have developed a lightweight convolutional neural network architecture able to perform view synthesis with occlusion handling in a stereo context, from one single, unlabelled and unannotated image, beyond state-of-the-art performance and with only a small amount of data required for training. In particular, it is able, at training and at test time, to estimate the disparity map corresponding to the problem at hand, and to evaluate a confidence in its prediction when using said disparity map for the synthesis. Knowing this confidence measure, it is then able to refine the value of the pixels wrongly estimated, with a refinement network component. The end result is a prediction built from a geometrical analysis of the scene, and completed in wrongly predicted areas by occlusion handling. Since 3D scene information is extracted in the course of the analysis, multiple new views can then be generated by interpolation.

Finally, in collaboration with Technicolor (N. Sabater and M. Hog), we have explored a novel way using recurrent neural networks to solve the problem of view synthesis in light fields. In particular, we proposed a novel solution using Long Short Term Memory Networks on a plane sweep volume. The approach has the advantage of having very few parameters and can be run on arbitrary sequence length. We have shown that the approach yields results that are competitive with the state of the art for dense light fields. Experimental results also show promising results when run on wider baselines.

7.3.2. *Light field inpainting and restoration*

Participants: Pierre Allain, Christine Guillemot, Laurent Guillo.

With the increasing popularity of computational photography brought by light field, simple and intuitive editing of light field images is becoming a feature of high interest for users. Light field editing can be combined with the traditional refocusing feature, allowing a user to include or remove objects from the scene, change its color, its contrast or other features. A simple approach for editing a light field image can be obtained with an edit propagation, where first a particular subaperture view is edited (most likely the center one) and then a coherent propagation of this edit is performed through the other views. This problem is particularly challenging for the task of inpainting, as the disparity field is unknown under the occluding mask. We have developed a method that is computationally fast while giving coherent disparity in the masked region, allowing us to inpaint a light field of 81 views in a few seconds [10].

We have also developed a novel light field denoising algorithm using a vector-valued regularization operating in the 4D ray space. More precisely, the method performs a PDE-based anisotropic diffusion along directions defined by local structures in the 4D ray space. It does not require prior estimation of disparity maps. The local structures in the 4D light field are extracted using a 4D tensor structure. We use a diffusivity coefficient derived from the amount of local variations in the 4D space to control the smoothing along directions, surfaces, or volumes in the 4D ray space. The diffusivity coefficient is computed as a function of the 4 eigenvalues of the 4D structure tensor. Experimental results show that the proposed denoising algorithm performs well compared to state of the art methods, while keeping tractable complexity, even with high noise levels (see Fig.5).

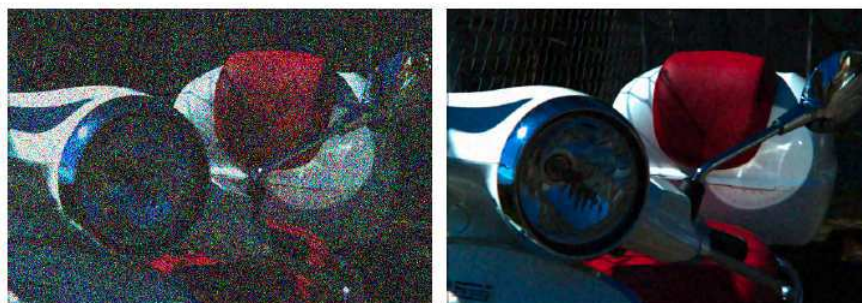


Figure 5. Illustration of denoising results, with additive white Gaussian noise of standard deviation $\sigma = 100$.

7.3.3. High dynamic range light fields capture

Participant: Christine Guillemot.

In collaboration with Trinity College Dublin (Prof. A. Smolic, Dr. M. Le Pendu), we have proposed a method for capturing *High Dynamic Range (HDR) light fields* with dense viewpoint sampling. Analogously to the traditional HDR acquisition process, several light fields are captured at varying exposures with a plenoptic camera. The raw data are de-multiplexed to retrieve all light field viewpoints for each exposure. We then perform a soft detection of saturated pixels. Considering a matrix which concatenates all the vectorized views, we formulate the problem of recovering saturated areas as a Weighted Low Rank Approximation (WLRA) where the weights are defined from the soft saturation detection. The proposed WLRA method [18], extending the matrix completion algorithm of [7] to nonbinary weights, is shown to better handle the transition between the saturated and non-saturated areas. While the Truncated Nuclear Norm (TNN) minimization, traditionally used for single view HDR imaging, does not generalize to light fields, the proposed WLRA method successfully recovers the parallax in the over-exposed areas.

7.4. Distributed processing and robust communication

Information theory, stochastic modelling, robust detection, maximum likelihood estimation, generalized likelihood ratio test, error and erasure resilient coding and decoding, multiple description coding, Slepian-Wolf coding, Wyner-Ziv coding, information theory, MAC channels

7.4.1. Information theoretic bounds for sequential massive random access to large database of correlated data

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

Massive random access is a new source coding paradigm that we proposed. It allows us to extract arbitrary sources from an appropriately compressed database purely by bit extraction. We studied the sequential aspect of this problem where the clients successively access to one source after the other. Theoretical bounds have been derived, and it was shown that the extraction can be done at the same rate as if the database was decoded and the requested sources were re-encoded. As for the storage, a reasonable overhead is required. In [26], we derived the optimal storage and transmission rate regions to the case of more general sources, which occur in practical scenarios. For the lossless source coding problem, we considered non i.i.d. sources (i.e., with memory, but also non necessary ergodic). We also showed that, in the case source statistics are unknown, the rate is increased by a factor that vanishes as the length of the data goes to infinity. Lossy compression is another context of interest, in particular for the application to video. Therefore, we derived achievable storage and transmission rate regions under a distortion constraint for i.i.d. [26] and correlated [13] Gaussian sources. Similarly, the transmission rate-distortion region is the same as if re-encoding of the requested sources was allowed. We are currently extending this work, by studying the constraints of the successive user requests and their influence on the transmission-storage rates performance.

7.4.2. *Correlation model selection for interactive video communication*

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

One application of the sequential massive random access problem is interactive video communication for multi-view videos. In this scheme, the server has to store the views as compactly as possible while allowing interactive navigation. Interactive navigation refers to the possibility for the user to select one view or a subset of views. To achieve this goal, the compression must be done using a model-based coding in which the correlation between the predicted view generated on the user side and the original view has to be modeled by a statistical distribution. A question of interest is therefore how to select a model among a candidate set of models that incurs the lowest extra rate cost to the system. To answer this question, one should evaluate the effect on the transmission rate of using at the decoder a wrong model distribution. This question is related to an open problem in information theory called the mismatch capacity. So, we did not tackle the question for any type of code as in the case of the mismatch capacity. In contrast, we focused on a type of code of practical interest: the linear codes. More precisely, we proposed a criterion to select the model when a linear block code is used for compression. We showed that, experimentally, the proposed bound is an accurate estimate of the effect of using a wrong model.

7.4.3. *Compression of spatio-temporally correlated and massive georeferenced data*

Participants: Thomas Maugey, Aline Roumy.

Another application of the sequential massive random access problem is interactive compression of spatio-temporally correlated sources. For example, highly instrumented smart cities are facing problems of management and storage of a large volume of data coming from an increasing number of sources. In [23] different compression schemes have been proposed that are able to exploit not only the temporal but also the spatial correlation between data sources. A special focus was made on a scheme where some sensors are used as references to predict the remaining sources. Finally, an adaptation of the scheme was proposed to offer interactivity and free selection of some sources by a client. This work was been done in collaboration with the Inria I4S project-team (A. Criniere), IFFSTAR (J. Dumoulin) and the L2S (M. Kieffer).

7.4.4. *ICON 3D - Interactive CODing for Navigation in 3D scenes*

Participants: Navid Mahmoudian Bidgoli, Thomas Maugey.

In the context of the ICON3D project, in collaboration with I3S-Nice (F. Payan), we have proposed a novel prediction tool for improving the compression performance of texture atlases of 3D meshes. This algorithm, called Geometry-Aware (GA) intra coding, takes advantage of the topology of the associated 3D meshes, in order to reduce the redundancies in the texture map. For texture processing, the general concept of the conventional intra prediction, used in video compression, has been adapted to utilize neighboring information on the 3D surface. We have also studied how this prediction tool can be integrated into a complete coding solution. In particular, a new block scanning strategy, as well as a graph-based transform for residual coding

have been proposed. Experimental results show that the knowledge of the mesh topology can significantly improve the compression efficiency of texture atlases.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

8.1.1. CIFRE contract with Technicolor on light fields editing

Participants: Christine Guillemot, Matthieu Hog.

- Title : Light fields editing
- Research axis : 7.1.1
- Partners : Technicolor (N. Sabater), Inria-Rennes.
- Funding : Technicolor, ANRT.
- Period : Oct.2015-Sept.2018.

Editing is quite common with classical imaging. Now, if we want light-field cameras to be in the future as common as traditional cameras, this functionality should also be enabled with light-fields. The goal of the PhD thesis is to develop methods for light-field editing, and in 2018 we have extended our concept of super-rays initially introduced for static light fields to video light fields (see Section 7.1.1). Super-rays group rays within and across views, emitted by the same set of 3D points in the space. A method for dynamic tracking of super-rays with scene flow estimation has been developed. We have further explored a novel way, using recurrent neural networks and in particular long short term memory (LSTM) networks, to solve the problem of view synthesis (see Section 7.3.1).

8.1.2. CIFRE contract with Technicolor on light fields compressed representation

Participants: Christine Guillemot, Fatma Hawary.

- Title : Light fields compressed representation
- Research axis : 7.2.2
- Partners : Technicolor (G. Boisson), Inria-Rennes.
- Funding : Technicolor, ANRT.
- Period : Feb.2016-Jan.2019.

The goal of this PhD thesis is to study reconstruction algorithms from compressed measurements. The goal is to apply these algorithms to scalable compression of light fields. Methods of sparse light field recovery have been developed, based on the assumption of sparsity in the Fourier domain, and using orthogonality constraint in the Fourier transform domain. The method has been further improved by introducing a refinement of the basis functions with non integer frequencies.

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

Participants: Jean Begaint, Christine Guillemot.

- Title : Cloud-based image compression
- Research axis : 7.2.6
- Partners : Technicolor (Ph. Guillotel, F. Galpin), Inria-Rennes.
- Funding : Technicolor, ANRT.
- Period : Nov.2015-Oct.2018.

The goal of this Cifre contract is to develop a novel image compression scheme exploiting similarity between images in a cloud. A region-based geometric and photometric alignment algorithm has been developed and validated for still image compression with an inter-coding set-up using similar images in the cloud as reference frames. This model has been further validated in the context of temporal prediction in a video compression scheme (see Section 7.2.6). Neural network based frame interpolation techniques have also been investigated, showing promising performance gains compared to the state of the art.

8.1.4. DGA contract on deep learning for image compression

Participants: Thierry Dumas, Christine Guillemot, Aline Roumy.

- Title : Deep learning for image compression
- Research axis : 7.2.5
- *Partners:* Inria-Rennes (Sirocco team)
- *Funding:* DGA/Ministry of defense
- Period : Oct.2015-Sept.2018.

This project funded by the DGA/Ministry of Defense concerns the PhD thesis of T. Dumas. The goal was to study deep learning architectures for image compression. Autoencoders have been studied to jointly learn transforms and quantizers with rate-distortion optimization criteria. A set of neural network architectures called Prediction Neural Networks Set (PNNS), based on both fully-connected and convolutional neural networks, has also been developed for intra image prediction (see Section 7.2.5).

9. Partnerships and Cooperations

9.1. Regional Initiatives

9.1.1. ICON 3D - Interactive COding for Navigation in 3D scenes

Participant: Thomas Maugey.

- Title : Interactive COding for Navigation in 3D scenes
- *Partners:* Inria-Rennes (Sirocco) and I3S Sophia-Antipolis (M. Antonini)
- *Funding:* CNRS GDR ISIS
- Period : Sept.2017-Sept.2018.

The project ICON 3D, funded by the GdR-Isis, aims at developing new geometry prediction algorithms for surface meshes. Given a part of a mesh, the prediction algorithm should be able to estimate a neighboring mesh subset corresponding to the one newly visible after user viewpoint angle change.

9.1.2. CominLabs InterCom project

Participants: Aline Roumy, Thomas Maugey.

- Title : Interactive Communication (INTERCOM): Massive random access to subsets of compressed correlated data .
- Research axis : 7.4.1
- *Partners :* Inria-Rennes (Sirocco team and I4S team); LabSTICC, IMT-Atlantique, Signal & Communications Department; External partner: M. Kieffer L2S.
- *Funding :* Labex CominLabs.
- Period : Oct. 2016 - Nov. 2019.

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

9.2. European Initiatives

9.2.1. FP7 & H2020 Projects

9.2.1.1. ERC-CLIM

Participants: Pierre Allain, Pierre David, Elian Dib, Simon Evain, Christian Galea, Christine Guillemot, Laurent Guillo, Xiaoran Jiang, Jinglei Shi.

- Title : Computational Light field Imaging.
- Research axis : [7.1.2](#), [7.1.3](#), [7.2.1](#), [7.2.3](#), [7.2.4](#), [7.3.1](#), [7.3.2](#), [7.3.3](#)
- Partners : Inria-Rennes
- Funding : European Research Council (ERC) advanced grant
- Period : Sept. 2016 - Aug. 2021.

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

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

9.3. International Initiatives

9.3.1. Inria International Labs

9.3.1.1. EPFL-Inria

- Title: Graph-based Omnidirectional video Processing
- International Partner: Ecole Polytechnique Fédérale de Lausanne (Switzerland), LTS4, Pascal Frossard
- Period: 2017-2018

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

9.3.2. Inria International Partners

9.3.2.1. Informal International Partners

We have international collaborations with:

- Reuben Farrugia, Prof. at the University of Malta, with whom we continue collaborating on light field super-resolution. The collaboration started during the sabbatical year (Sept. 2015-Aug. 2016) he spent within the team.
- Ehsan Miandji and Prof. Jonas Unger from Linköping Univ. with whom we collaborate on compressive sampling of light fields.
- Mikael Le Pendu and Prof. Aljosa Smolic from Trinity College Dublin on HDR light field recovery from multiple exposures.

9.4. International Research Visitors

9.4.1. Visits of International Scientists

- Reuben Farrugia, Prof. at the University of Malta, spent 2 weeks in the team (June 2018).
- Alexander Sagel, assistant researcher at the Technical University Munich (Oct.-Nov. 2018).
- Mikael Le Pendu and Martin Alain, postdocs at Trinity College Dublin (Nov. 2018).

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. Member of the Organizing Committees

- C. Guillemot was tutorial co-chair of IEEE Int. Conf. on Image Processing (ICIP, Athens, 2018)

10.1.2. Scientific Events Selection

10.1.2.1. Member of the Conference Program Committees

- C. Guillemot has been a member of the international program committee of PCS 2018, ACIVS 2018, VISAP 2018, ICIP 2018
- T. Maugey was publicity chair and member of the technical program committee of the Graph Signal Processing workshop at EPFL (Switzerland) (2018) and area chair of IEEE-VCIP 2018
- A. Roumy has been a member of the technical program committees of the CVPR 2018 workshop on New Trends in Image Restoration and Enhancement (NTIRE), of the ICC 2018 workshop on Promises and Challenges of Machine Learning in Communication Networks (ML4COM), and of the ICT 2018 conference.

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

- C. Guillemot is senior area editor of the IEEE Trans. on Image Processing
- C. Guillemot is associate editor of the International Journal on Mathematical Imaging and Vision
- A. Roumy is associate editor of the Springer Annals of Telecommunications.
- A. Roumy is associate editor of the IEEE Trans. on Image Processing (since Nov. 2018)

10.1.4. Invited Talks

- C. Guillemot gave a keynote talk on light fields image processing at COmpression and REpresentation of Audio-visual Signals (CORESA), Poitiers, 14 Nov. 2018.
- C. Guillemot gave an IEEE Distinguished lecture, Computational light field imaging, invited by the Croatian and Slovenian IEEE signal processing chapter, at Ljubljana (23rd Oct. 2018) and Zagreb (24th. Oct. 2018).
- C. Guillemot gave a lecture at the IEEE signal processing summer school at Midsweden University, Sundvall, Sweden, May 2018.
- C. Guillemot gave a seminar at KTH on light field image processing, 23rd March 2018
- C. Guillemot gave an IEEE Distinguished lecture invited by the Italian IEEE signal processing chapter, Cavalese, 22 Jan. 2018
- Thomas Maugey gave a seminar at ENS-Rennes on Free Viewpoint Television, April 2018.
- A. Roumy gave a talk on Source coding under massive random access: theory and applications, at ENSEA, Cergy-Pontoise, April 2018.
- A. Roumy gave a talk on Source coding under massive random access: theory and applications, at INSA, Lyon, Sept. 2018.
- A. Roumy gave a talk on universal compression of images, at EPFL, Dec. 2018.

10.1.5. Leadership within the Scientific Community

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

10.1.6. Research Administration

- C. Guillemot is a member of Inria evaluation committee
- C. Guillemot is member of the “bureau du Comité des Projets”
- C. Guillemot has served as a member of the selection committee of an Assistant Prof. at Telecom ParisTech (Jan. 2018)
- L. Guillo is a member of the technological development and technology transfer committee at Inria Rennes.
- A. Roumy served as a jury member for the selection of Inria CR (researcher) candidates, Inria, Rennes, France, 2018.
- A. Roumy is a member of the technological development and technology transfer committee at Inria Rennes.
- A. Roumy is a member of the Inria Joint Administrative Committee (CAP commission administrative paritaire).

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

- Master: C. Guillemot, Image and video compression, 10 hours, and advanced video processing, 4 hours, M2 SISEA, Univ. of Rennes 1, France.

- Undergraduate: L. Guillo, course of 35 hours on Functional immutable programming, 1st year of Mathématiques, Informatique, Electronique, mathematique-Economie" (MIEE), Univ. of Rennes 1
- Undergraduate: L. Guillo, talk about agility and software development, 3rd year, ENS Rennes.
- Master: T. Maugey, course on 3D models in a module on advanced video, 4 hours, M2 SISEA, Univ. of Rennes 1, France.
- Master: T. Maugey, course on Representation, editing and perception of digital images, 12 hours, M2 SIF, Univ. of Rennes 1, France.
- Engineering degree: A. Roumy, Sparse methods in image and signal processing, 13 hours, INSA Rennes, 5th year, Mathematical engineering, France.
- Master: A. Roumy, Foundations of smart sensing, 18 hours, ENSAI, Master of Science in Statistics for Smart Data, France.
- Master: A. Roumy, High dimensional statistical learning, 9 hours, University Rennes 1, SIF Master, France.
- Master: A. Roumy, Information theory, 15 hours, University Rennes 1, SIF master, France.

10.2.2. Juries

- C. Guillemot has been member of the PhD committees of:
 - D. Liu, KTH, Stockholm, March 2018
 - K. Reuze, INSA, Rennes, Nov. 2018
 - S. Ferdowsi, Univ. Geneva, Dec. 2018
- Thomas Maugey has been member of the jury of the PhD committee of:
 - Renata Khasanova, EPFL, Dec. 2018

10.3. Popularization

- C. Guillemot and A. Roumy participated in the editorial board of a series of articles, towards a general audience, on Fourier, in the context of his 250th birth anniversary (2018)

11. Bibliography

Publications of the year

Doctoral Dissertations and Habilitation Theses

- [1] J. BÉGAINT. *Towards novel inter-prediction methods for image and video compression*, Rennes 1, November 2018, <https://hal.archives-ouvertes.fr/tel-01960088>
- [2] D. GOMMELET. *Methods for improving the backward compatible High Dynamic Range compression*, Université Rennes 1, September 2018, <https://tel.archives-ouvertes.fr/tel-01947401>
- [3] M. HOG. *Light Field Editing and Rendering*, Rennes 1, November 2018, <https://hal.archives-ouvertes.fr/tel-01966828>
- [4] A. ROUMY. *Universal transmission and compression: send the model?*, Université Rennes 1, November 2018, Habilitation à diriger des recherches, <https://hal.inria.fr/tel-01916058>

Articles in International Peer-Reviewed Journals

- [5] J. BÉGAINT, D. THOREAU, P. GUILLOTTEL, C. GUILLEMOT. *Region-based prediction for image compression in the cloud*, in "IEEE Transactions on Image Processing", 2018, vol. 27, n^o 4, pp. 1835-1846 [DOI : 10.1109/TIP.2017.2788192], <https://hal.archives-ouvertes.fr/hal-01662639>
- [6] R. A. FARRUGIA, C. GUILLEMOT. *Light Field Super-Resolution using a Low-Rank Prior and Deep Convolutional Neural Networks*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", 2019, pp. 1-15 [DOI : 10.1109/TPAMI.2019.2893666], <https://hal.archives-ouvertes.fr/hal-01984843>
- [7] M. LE PENDU, X. JIANG, C. GUILLEMOT. *Light Field Inpainting Propagation via Low Rank Matrix Completion*, in "IEEE Transactions on Image Processing", April 2018, vol. 27, n^o 4, pp. 1981 - 1993 [DOI : 10.1109/TIP.2018.2791864], <https://hal.archives-ouvertes.fr/hal-01720622>
- [8] R. MA, T. MAUGEY, P. FROSSARD. *Optimized Data Representation for Interactive Multiview Navigation*, in "IEEE Transactions on Multimedia", July 2018, vol. 20, n^o 7, pp. 1595-1609, <https://hal.inria.fr/hal-01633660>
- [9] E. VURAL, C. GUILLEMOT. *A Study of the Classification of Low-Dimensional Data with Supervised Manifold Learning*, in "Journal of Machine Learning Research", 2018, vol. 18, n^o 1, pp. 5741-5795, <https://hal.archives-ouvertes.fr/hal-01737768>

International Conferences with Proceedings

- [10] P. ALLAIN, L. GUILLO, C. GUILLEMOT. *Fast light field inpainting propagation using angular warping and color-guided disparity interpolation*, in "Acivs 2018 - Advanced Concepts for Intelligent Vision Systems", Poitiers, France, LNCS, volume 11182, Springer, September 2018, pp. 559-570 [DOI : 10.1007/978-3-030-01449-0_47], <https://hal.archives-ouvertes.fr/hal-01846028>
- [11] J. BÉGAINT, F. GALPIN, P. GUILLOTTEL, C. GUILLEMOT. *Region-based models for motion compensation in video compression*, in "PCS 2018 - Picture Coding Symposium", San Francisco, United States, June 2018, pp. 154-158 [DOI : 10.1109/PCS.2018.8456258], <https://hal.archives-ouvertes.fr/hal-01807693>
- [12] T. DUMAS, A. ROUMY, C. GUILLEMOT. *Autoencoder based image compression: can the learning be quantization independent?*, in "ICASSP 2018 - International Conference on Acoustics, Speech and Signal Processing ICASSP", Calgary, Canada, IEEE, April 2018, pp. 1188-1192, <https://arxiv.org/abs/1802.09371> [DOI : 10.1109/ICASSP.2018.8462263], <https://hal.archives-ouvertes.fr/hal-01713644>
- [13] E. DUPRAZ, T. MAUGEY, A. ROUMY, M. KIEFFER. *Rate-Distortion Performance of Sequential Massive Random Access to Gaussian Sources with Memory*, in "DCC 2018 - Data compression conference", Snowbird, United States, March 2018, pp. .1-10 [DOI : 10.1109/DCC.2018.00059], <https://hal.archives-ouvertes.fr/hal-01759076>
- [14] F. HAWARY, G. BOISSON, C. GUILLEMOT, P. GUILLOTTEL. *Compressive 4D Light Field Reconstruction Using Orthogonal Frequency Selection*, in "ICIP 2018 - 25th IEEE International Conference on Image Processing", Athens, Greece, IEEE, October 2018, pp. 3863-3867 [DOI : 10.1109/ICIP.2018.8451110], <https://hal.archives-ouvertes.fr/hal-01810353>

- [15] M. HOG, N. SABATER, C. GUILLEMOT. *Dynamic Super-Rays for Efficient Light Field Video Processing*, in "BMVC 2018 - 29th British Machine Vision Conference", Newcastle upon Tyne, United Kingdom, September 2018, pp. 1-12, <https://hal.archives-ouvertes.fr/hal-01649342>
- [16] X. JIANG, C. GALEA, L. GUILLO, C. GUILLEMOT. *3D Scene Modeling from Dense Video Light Fields*, in "IC3D 2018 - 8th International conference on 3D immersion", Brussels, Belgium, December 2018, pp. 1-7, <https://hal.inria.fr/hal-01925319>
- [17] X. JIANG, M. LE PENDU, C. GUILLEMOT. *Depth estimation with occlusion handling from a sparse set of light field views*, in "ICIP 2018 - IEEE International Conference on Image Processing", Athens, Greece, IEEE, October 2018, pp. 634-638 [DOI : 10.1109/ICIP.2018.8451466], <https://hal.archives-ouvertes.fr/hal-01786049>
- [18] M. LE PENDU, A. SMOLIC, C. GUILLEMOT. *High dynamic range light fields via weighted low rank approximation*, in "ICIP 2018 - IEEE International Conference on Image Processing", Athens, Greece, IEEE, October 2018, pp. 1-5 [DOI : 10.1109/ICIP.2018.8451584], <https://hal.archives-ouvertes.fr/hal-01872277>
- [19] E. MIANDJI, J. UNGER, C. GUILLEMOT. *Multi-Shot Single Sensor Light Field Camera Using a Color Coded Mask*, in "EUSIPCO 2018 - 26th European Signal Processing Conference", Roma, Italy, September 2018, pp. 1-5, <https://hal.archives-ouvertes.fr/hal-01807780>
- [20] *Best Paper*
M. RIZKALLAH, F. DE SIMONE, T. MAUGEY, C. GUILLEMOT, P. FROSSARD. *Rate Distortion Optimized Graph Partitioning for Omnidirectional Image Coding*, in "EUSIPCO 2018 - 26th European Signal Processing Conference", Rome, Italy, September 2018, pp. 1-5, <https://hal.inria.fr/hal-01807613>.
- [21] M. RIZKALLAH, X. SU, T. MAUGEY, C. GUILLEMOT. *Graph-based Transforms for Predictive Light Field Compression based on Super-Pixels*, in "ICASSP 2018 - IEEE International Conference on Acoustics, Speech and Signal Processing", Calgary, Canada, IEEE, April 2018, pp. 1718-1722 [DOI : 10.1109/ICASSP.2018.8462288], <https://hal.inria.fr/hal-01737332>
- [22] X. SU, M. RIZKALLAH, T. MAUGEY, C. GUILLEMOT. *Rate-Distortion Optimized Super-Ray Merging for Light Field Compression*, in "EUSIPCO 2018 - 26th European Signal Processing Conference", Rome, Italy, Proceedings, September 2018, pp. 1-5, <https://hal.archives-ouvertes.fr/hal-01812858>

Conferences without Proceedings

- [23] A. CRINIÈRE, A. ROUMY, T. MAUGEY, M. KIEFFER, J. DUMOULIN. *Compression of spatio-temporally correlated and massive georeferenced Data*, in "EGU 2018 - European Geosciences Union General Assembly", Vienne, Austria, April 2018, <https://hal.inria.fr/hal-01887809>

Research Reports

- [24] L. GUILLO, X. JIANG, G. LAFRUIT, C. GUILLEMOT. *Light field video dataset captured by a R8 Raytrix camera (with disparity maps)*, INTERNATIONAL ORGANISATION FOR STANDARDISATION ISO/IEC JTC1/SC29/WG1 & WG11, April 2018, pp. 1-6, <https://hal.inria.fr/hal-01804578>

Other Publications

-
- [25] T. DUMAS, A. ROUMY, C. GUILLEMOT. *Context-adaptive neural network based prediction for image compression*, July 2018, <https://arxiv.org/abs/1807.06244> - working paper or preprint, <https://hal.archives-ouvertes.fr/hal-01841034>
- [26] E. DUPRAZ, T. MAUGEY, A. ROUMY, M. KIEFFER. *Transmission and Storage Rates for Sequential Massive Random Access*, May 2018, working paper or preprint, <https://hal.inria.fr/hal-01799016>