Activity Report 2016

Project-Team SIROCCO

Analysis representation, compression and communication of visual data

Table of contents

Project-Team SIROCCO

Creation of the Project-Team: 2012 January 01

Keywords:

Computer Science and Digital Science:
  5. - Interaction, multimedia and robotics
  5.3. - Image processing and analysis
  5.4. - Computer vision
  5.9. - Signal processing

Other Research Topics and Application Domains:
  6. - IT and telecom

1. Members

Research Scientists
  Christine Guillemot [Team leader, Inria, Senior Researcher, HDR]
  Thomas Maugey [Inria, Researcher]
  Aline Roumy [Inria, 80%, Researcher]

Faculty Member
  Olivier Le Meur [Univ. Rennes I, Associate Professor, HDR]

Engineers
  Cedric Le Cam [Inria]
  Mikael Le Pendu [Inria]
  Xin Su [Inria]

PhD Students
  Martin Alain [Technicolor, granted by CIFRE, until Mar 2016]
  Jean Begaint [Technicolor, granted by CIFRE]
  Pierre David [Inria, from Oct 2016]
  Elian Dib [Inria, from Oct 2016]
  Thierry Dumas [Inria]
  Julio Cesar Ferreira [co-tutelle between the Univ. of Rennes 1 and the Univ. of Uberlandia, until July 2016]
  David Gommelet [Cifre Ericsson, granted by CIFRE]
  Fatma Hawary [Technicolor, granted by CIFRE]
  Matthieu Hog [Technicolor, granted by CIFRE]
  Hristina Hristova [Univ. Rennes I, also with FRVsense]
  Dmitry Kuzovkin [Technicolor, granted by CIFRE, also with FRVSense]
  Navid Mahmoudian Bidgoli [Inria, from Oct 2016]
  Mira Rizkallah [Univ. Rennes I]
  Maxime Rousselot [Harmonics, granted by CIFRE, also with FRVSense]

Post-Doctoral Fellows
  Oriel Frigo [Inria, from Dec 2016]
  Xiaoran Jiang [Inria]

Visiting Scientists
  Reuben Farrugia [University of Malta, until Aug 2016]
  Sheila S. Hemami [Univ. Rennes I, from May 2016 until Aug 2016]

Administrative Assistant

Huguette Bechu [Inria]

2. Overall Objectives

2.1. Introduction

The goal of the SIROCCO project-team is the design and development of algorithms and practical solutions in the areas of analysis, modelling, coding, and communication of images and video signals. The objective is to cover several inter-dependent algorithmic problems of the end-to-end transmission chain from the capturing, compression, transmission to the rendering of the visual data. The project-team activities are structured and organized around the following inter-dependent research axes:

- Analysis and modeling for compact representation and navigation [1] in large volumes of visual data [2]

- Rendering, inpainting and super-resolution of visual data

- Representation and compression of visual data

- Distributed processing and robust communication of visual data

Given the strong impact of standardization in the sector of networked multimedia, SIROCCO, in partnership with industrial companies, seeks to promote its results in standardization (MPEG). While aiming at generic approaches, some of the solutions developed are applied to practical problems in partnership with industry (Alcatel Lucent, Astrium, Orange labs., Technicolor, Thomson Video Networks) or in the framework of national projects (ANR-ARSSO, ANR-PERSEE). The application domains addressed by the project are networked visual applications taking into account their various requirements and needs in terms of compression, of resilience to channel noise and network adaptation, of advanced functionalities such as navigation, and of high quality rendering.

2.2. Analysis and modeling for compact representation

Analysis and modeling of the visual data are crucial steps for a number of video processing problems: navigation in 3D scenes, compression, loss concealment, denoising, inpainting, editing, content summarization and navigation. The focus is on the extraction of different cues such as scene geometry, edge, texture and motion, on the extraction of high-level features (GIST-like or epitomes), and on the study of computational models of visual attention, useful for different visual processing tasks. In relation to the above problems, the project-team considers various types of image modalities (medical and satellite images, natural 2D still and moving images, multi-view and multi-view plus depth video content).

2.3. Rendering, inpainting and super-resolution

This research axis addresses the problem of high quality reconstruction of various types of visual data after decoding. Depending on the application and the corresponding type of content (2D, 3D), various issues are being addressed. For example, to be able to render 3D scenes, depth information is associated with each view as a depth map, and transmitted in order to perform virtual view generation. Given one view with its depth information, depth image-based rendering techniques have the ability to render views in any other spatial positions. However, the issue of intermediate view reconstruction remains a difficult ill-posed problem. Most errors in the view synthesis are caused by incorrect geometry information, inaccurate camera parameters, and occlusions/disocclusions. Efficient inpainting techniques are necessary to restore disocclusions areas. Inpainting techniques are also required in transmission scenarios, where packet losses result in missing data in the video after decoding. The design of efficient mono-view and multi-view super-resolution methods is also part of the project-team objectives to improve the rendering quality, as well as to trade-off quality against transmission rate.

---

[1] By navigation we refer here to scene navigation by virtual view rendering, and to navigation across slices in volumic medical images.
[2] By visual data we refer to natural and medical images, videos, multi-view sequences as well as to visual cues or features extracted from video content.

2.4. Representation and compression of visual data

The objective is to develop algorithmic tools for constructing low-dimensional representations of multi-view video plus depth data, of 2D image and video data, of visual features and of their descriptors. Our approach goes from the design of specific algorithmic tools to the development of complete compression algorithms. The algorithmic problems that we address include data dimensionality reduction, the design of compact representations for multi-view plus depth video content which allow high quality 3D rendering, the design of sparse representation methods and of dictionary learning techniques. The sparsity of the representation indeed depends on how well the dictionary is adapted to the data at hand. The problem of dictionary learning for data-adaptive representations, that goes beyond the concatenation of a few traditional bases, has thus become a key issue which we address for further progress in the area.

Developing complete compression algorithms necessarily requires tackling visual processing topics beyond the issues of sparse data representation and dimensionality reduction. For example, problems of scalable, perceptual, and metadata-aided coding of 2D and 3D visual data, as well as of near lossless compression of medical image modalities (CT, MRI, virtual microscopy imaging) are tackled. Finally, methods for constructing rate-efficient feature digests allowing processing in lower-dimensional spaces, e.g. under stringent bandwidth constraints, also falls within the scope of this research axis.

2.5. Distributed processing and robust communication

The goal is to develop theoretical and practical solutions for robust image and video transmission over heterogeneous and time-varying networks. The first objective is to construct coding tools that can adapt to heterogeneous networks. This includes the design of (i) sensing modules to measure network characteristics, of (ii) robust coding techniques and of (iii) error concealment methods for compensating for missing data at the decoder when erasures occur during the transmission. The first objective is thus to develop sensing and modeling methods which can recognize, model and predict the packets loss/delay end-to-end behaviour. Given the estimated and predicted network conditions (e.g. Packet Error Rate (PER)), the objective is then to adapt the data coding, protection and transmission scheme. However, the reliability of the estimated PER impacts the performance of FEC schemes. We investigate the problem of constructing codes which would be robust to channel uncertainty, i.e. which would perform well not only on a specific channel but also "universally", hence reducing the need for a feedback channel. This would be a significant advantage compared with rateless codes such as fountain codes which require a feedback channel. Another problem which we address is error concealment. This refers to the problem of estimating lost symbols from the received ones by exploiting spatial and/or temporal correlation within the video signal.

The availability of wireless camera sensors has also been spurring interest for a variety of applications ranging from scene interpretation, object tracking and security environment monitoring. In such camera sensor networks, communication energy and bandwidth are scarce resources, motivating the search for new distributed image processing and coding (Distributed Source Coding) solutions suitable for band and energy limited networking environments. In the past years, the team has developed a recognized expertise in the area of distributed source coding, which in theory allows for each sensor node to communicate losslessly at its conditional entropy rate without information exchange between the sensor nodes. However, distributed source coding (DSC) is still at the level of the proof of concept and many issues remain unresolved. The goal is thus to further address theoretical issues as the problem of modeling the correlation channel between sources, to further study the practicality of DSC in image coding and communication problems.

3. Research Program

3.1. Introduction

The research activities on analysis, compression and communication of visual data mostly rely on tools and formalisms from the areas of statistical image modelling, of signal processing, of coding and information

theory. However, the objective of better exploiting the Human Visual System (HVS) properties in the above goals also pertains to the areas of perceptual modelling and cognitive science. Some of the proposed research axes are also based on scientific foundations of computer vision (e.g. multi-view modelling and coding). We have limited this section to some tools which are central to the proposed research axes, but the design of complete compression and communication solutions obviously rely on a large number of other results in the areas of motion analysis, transform design, entropy code design, etc which cannot be all described here.

3.2. Parameter Estimation and Inference

Bayesian estimation, Expectation-Maximization, stochastic modelling

Parameter estimation is at the core of the processing tools studied and developed in the team. Applications range from the prediction of missing data or future data, to extracting some information about the data in order to perform efficient compression. More precisely, the data are assumed to be generated by a given stochastic data model, which is partially known. The set of possible models translates the a priori knowledge we have on the data and the best model has to be selected in this set. When the set of models or equivalently the set of probability laws is indexed by a parameter (scalar or vectorial), the model is said parametric and the model selection resorts to estimating the parameter. Estimation algorithms are therefore widely used at the encoder to analyze the data. In order to achieve high compression rates, the parameters are usually not sent and the decoder has to jointly select the model (i.e. estimate the model parameters) and extract the information of interest.

3.3. Data Dimensionality Reduction

Manifolds, locally linear embedding, non-negative matrix factorization, principal component analysis

A fundamental problem in many data processing tasks (compression, classification, indexing) is to find a suitable representation of the data. It often aims at reducing the dimensionality of the input data so that tractable processing methods can then be applied. Well-known methods for data dimensionality reduction include principal component analysis (PCA) and independent component analysis (ICA). The methodologies which will be central to several proposed research problems will instead be based on sparse representations, on locally linear embedding (LLE) and on the "non negative matrix factorization" (NMF) framework.

The objective of *sparse representations* is to find a sparse approximation of a given input data. In theory, given $A \in \mathbb{R}^{m \times n}$, $m < n$, and $\mathbf{b} \in \mathbb{R}^m$ with $m << n$ and A is of full rank, one seeks the solution of $\min\{\|\mathbf{x}\|_0 \ : \ A\mathbf{x} = \mathbf{b}\}$, where $\|\mathbf{x}\|_0$ denotes the L_0 norm of x, i.e. the number of non-zero components in z. There exist many solutions x to $Ax = b$. The problem is to find the sparsest, the one for which x has the fewest non zero components. In practice, one actually seeks an approximate and thus even sparser solution which satisfies $\min\{\|\mathbf{x}\|_0 \ : \ \|A\mathbf{x} - \mathbf{b}\|_p \leq \rho\}$, for some $\rho \geq 0$, characterizing an admissible reconstruction error. The norm p is usually 2, but could be 1 or ∞ as well. Except for the exhaustive combinatorial approach, there is no known method to find the exact solution under general conditions on the dictionary A. Searching for this sparsest representation is hence unfeasible and both problems are computationally intractable. Pursuit algorithms have been introduced as heuristic methods which aim at finding approximate solutions to the above problem with tractable complexity.

Non negative matrix factorization (NMF) is a non-negative approximate data representation [3]. NMF aims at finding an approximate factorization of a non-negative input data matrix V into non-negative matrices W and H, where the columns of W can be seen as *basis vectors* and those of H as coefficients of the linear approximation of the input data. Unlike other linear representations like PCA and ICA, the non-negativity constraint makes the representation purely additive. Classical data representation methods like PCA or Vector Quantization (VQ) can be placed in an NMF framework, the differences arising from different constraints being placed on the W and H matrices. In VQ, each column of H is constrained to be unitary with only one non-zero coefficient which is equal to 1. In PCA, the columns of W are constrained to be orthonormal and the rows of H to be orthogonal to each other. These methods of data-dependent dimensionality reduction will be at the core of our visual data analysis and compression activities.

---

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

3.4. Perceptual Modelling

Saliency, visual attention, cognition

The human visual system (HVS) is not able to process all visual information of our visual field at once. To cope with this problem, our visual system must filter out irrelevant information and reduce redundant information. This feature of our visual system is driven by a selective sensing and analysis process. For instance, it is well known that the greatest visual acuity is provided by the fovea (center of the retina). Beyond this area, the acuity drops down with the eccentricity. Another example concerns the light that impinges on our retina. Only the visible light spectrum lying between 380 nm (violet) and 760 nm (red) is processed. To conclude on the selective sensing, it is important to mention that our sensitivity depends on a number of factors such as the spatial frequency, the orientation or the depth. These properties are modeled by a sensitivity function such as the Contrast Sensitivity Function (CSF).

Our capacity of analysis is also related to our visual attention. Visual attention which is closely linked to eye movement (note that this attention is called *overt* while the covert attention does not involve eye movement) allows us to focus our biological resources on a particular area. It can be controlled by both top-down (i.e. goal-directed, intention) and bottom-up (stimulus-driven, data-dependent) sources of information [4]. This detection is also influenced by prior knowledge about the environment of the scene [5]. Implicit assumptions related to prior knowledge or beliefs play an important role in our perception (see the example concerning the assumption that light comes from above-left). Our perception results from the combination of prior beliefs with data we gather from the environment. A Bayesian framework is an elegant solution to model these interactions [6]. We define a vector $\overrightarrow{v}_l$ of local measurements (contrast of color, orientation, etc.) and vector $\overrightarrow{v}_c$ of global and contextual features (global features, prior locations, type of the scene, etc.). The salient locations S for a spatial position $\overrightarrow{x}$ are then given by:

$$S(\overrightarrow{x}) = \frac{1}{p(\overrightarrow{v}_l \,|\, \overrightarrow{v}_c)} \times p(s, \overrightarrow{x} \,|\, \overrightarrow{v}_c) \qquad (1)$$

The first term represents the bottom-up salience. It is based on a kind of contrast detection, following the assumption that rare image features are more salient than frequent ones. Most of existing computational models of visual attention rely on this term. However, different approaches exist to extract the local visual features as well as the global ones. The second term is the contextual priors. For instance, given a scene, it indicates which parts of the scene are likely the most salient.

3.5. Coding theory

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

Source coding and channel coding theory [7] 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 [8] (SW) and Wyner-Ziv [9]

---

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

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

[6]L. Zhang, M. Tong, T. Marks, H. Shan, H. and G.W. Cottrell, "SUN: a Bayesian framework for saliency using natural statistics", Journal of Vision, vol. 8, pp. 1-20, 2008.

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

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

[9]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.

(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. Introduction

The application domains addressed by the project are:

- Compression with advanced functionalities of various image modalities (including multi-view, medical images such as MRI, CT, WSI, or satellite images);
- Networked multimedia applications taking into account their various needs in terms of image and 2D and 3D video compression, or in terms of network adaptation (e.g., resilience to channel noise);
- Content editing and post-production.

4.2. Compression of emerging imaging modalities

Compression of images and of 2D video (including High Definition and Ultra High Definition) 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), super-resolution, formats for immersive displays allowing panoramic viewing and 3DTV.

Different video data formats and technologies are envisaged for interactive and immersive 3D video applications using omni-directional videos, stereoscopic or multi-view videos. The "omni-directional video" set-up refers to 360-degree view from one single viewpoint or spherical video. Stereoscopic video is composed of two-view videos, the right and left images of the scene which, when combined, can recreate the depth aspect of the scene. A multi-view video refers to multiple video sequences captured by multiple video cameras and possibly by depth cameras. Associated with a view synthesis method, a multi-view video allows the generation of virtual views of the scene from any viewpoint. This property can be used in a large diversity of applications, including Three-Dimensional TV (3DTV), and Free Viewpoint Video (FTV). The notion of "free viewpoint video" refers to the possibility for the user to choose an arbitrary viewpoint and/or view direction within a visual scene, creating an immersive environment. Multi-view video generates a huge amount of redundant data which need to be compressed for storage and transmission. 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

3D and Free Viewpoint TV: The emergence of multi-view auto-stereoscopic displays has spurred a recent interest for broadcast or Internet delivery of 3D video to the home. Multiview video, with the help of depth information on the scene, allows scene rendering on immersive stereo or auto-stereoscopic displays for 3DTV applications. It also allows visualizing the scene from any viewpoint, for scene navigation and free-viewpoint TV (FTV) applications. However, the large volumes of data associated to multi-view video plus depth content raise new challenges in terms of compression and communication.

Internet and mobile video: Broadband fixed (ADSL, ADSL2+) and mobile access networks with different radio access technologies (RAT) (e.g. 3G/4G, GERAN, UTRAN, DVB-H), have enabled not only IPTV and Internet TV but also the emergence of mobile TV and mobile devices with internet capability. A major challenge for next internet TV or internet video remains to be able to deliver the increasing variety of media (including more and more bandwidth demanding media) with a sufficient end-to-end QoS (Quality of Service) and QoE (Quality of Experience).

Mobile video retrieval: The Internet has changed the ways of interacting with content. The user is shifting its media consumption from a passive to a more interactive mode, from linear broadcast (TV) to on demand content (YouTubes, iTunes, VoD), and to user-generated, searching for relevant, personalized content. New mobility and ubiquitous usage has also emerged. The increased power of mobile devices is making content search and retrieval applications using mobile phones possible. Quick access to content in mobile environments with restricted bandwidth resources will benefit from rate-efficient feature extraction and description.

Wireless multi-camera vision systems: Our activities on scene modelling, on rate-efficient feature description, distributed coding and compressed sensing should also lead to algorithmic building blocks relevant for wireless multi-camera vision systems, for applications such as visual surveillance and security.

4.4. Editing and post-production

Video editing and post-production are critical aspects in the audio-visual production process. Increased ways of "consuming" video content also highlight the need for content repurposing as well as for higher interaction and editing capabilities. Content captured at very high resolutions may need to be repurposed in order to be adapted to the requirements of actual users, to the transmission channel or to the terminal. 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. Other applications such as video annotation/retrieval, video restoration/stabilization, augmented reality, can also benefit from the proposed research.

5. Highlights of the Year

5.1. Highlights of the Year

5.1.1. *Awards*

C. Guillemot has been granted an ERC advanced grant for a project on computational light fields imaging.

6. New Software and Platforms

6.1. Salient object extraction

FUNCTIONAL DESCRIPTION This software detects salient object in an input picture in an automatic manner. The detection is based on super-pixel segmentation and contrast of histogram. This software is dedicated to people working in image processing and post production.
- Participants: Zhi Liu and Olivier Le Meur
- Contact: Olivier Le Meur

6.2. VideoInpainting

KEYWORDS: Video Inpainting - Motion informations - Loss concealment - BMFI (Bilinear Motion Field Interpolation)

SCIENTIFIC DESCRIPTION From an input binary mask and a source picture, the software performs an examplar-based inpainting. The method is based on the combination of multiple inpainting applied on a low resolution of the input picture. Once the combination has been done, a single-image super-resolution method is applied to recover the details and the high frequency in the inpainted areas. The developments have been pursued in 2014, in particular by introducing a Poisson blending step in order to improve the visual quality of the inpainted video. This software is dedicated to people working in image processing and post production.

- Participants: Ronan Le Boulch and Olivier Le Meur
- Contact: Olivier Le Meur

6.3. Visual Fixation Analysis

SCIENTIFIC DESCRIPTION From a set of fixation data and a picture, the software called Visual Fixation Analysis extracts from the input data a number of features (fixation duration, saccade length, orientation of saccade...) and computes a human saliency map. The software can also be used to assess the degree of similarity between a ground truth (eye fixation data) and a predicted saliency map. This software is dedicated to people working in cognitive science and computer vision.

- Participants: Olivier Le Meur and Thierry Baccino
- Partner: Université de Rennes 1
- Contact: Olivier Le Meur

6.4. Saccadic model

SCIENTIFIC DESCRIPTION The software called Scanpath Prediction aims at predicting the visual scanpath of an observer. The visual scanpath is a set of fixation points. The computational model is based on bottom-up saliency maps, viewing tendencies (that have been learned from eye tracking datasets) and inhibition-of-return. A presentation of this model is available on the following link: http://fr.slideshare.net/OlivierLeMeur/saccadic-model-of-eye-movements-for-freeviewing-condition. This software is dedicated to people working in computer science, computer vision and cognitive science. This software is being registered at the APP (Agence de Protection des Programmes) under the number IDDN.FR.001.240029.000.S.P.2016.000.10000.

- Participants: Olivier Le Meur
- Partner: Université de Rennes 1
- Contact: Olivier Le Meur

6.5. ADT-ATep

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 the view angle from which he wants to observe the scene. Despite the great interest for FTV, the lack of realistic and ambitious datasets penalizes the research effort. The acquisition of such sequences is very costly in terms of hardware and working effort, which explains why no multi-view videos suitable for FTV has been made available yet.

A project founded by Inriahub has recently started in the SIROCCO team. Called ATeP for "Acquisition, Traitement et Partage" (Acquisition, Processing and Sharing), it targets the acquisition of such dataset. Another interesting aspect of this project is that the acquisition system relies on omnidirectional cameras. The dataset will thus interest all the industries and scientists currently working on the development of efficient processing and coding tools for 360 videos.

- Participants: Cédric Le Cam, Thomas Maugey
- Partner: Inria
- Contact: Thomas Maugey

7. New Results

7.1. Analysis and modeling for compact representation

3D modelling, multi-view plus depth videos, light-fields, 3D meshes, epitomes, image-based rendering, inpainting, view synthesis

7.1.1. *Visual attention*
Participant: Olivier Le Meur.

Visual attention is the mechanism allowing to focus our visual processing resources on behaviorally relevant visual information. Two kinds of visual attention exist: one involves eye movements (overt orienting) whereas the other occurs without eye movements (covert orienting). Our research activities deal with the understanding and modeling of overt attention.

Saccadic model: Previous research showed the existence of systematic tendencies in viewing behavior during scene exploration. For instance, saccades are known to follow a positively skewed, long-tailed distribution, and to be more frequently initiated in the horizontal or vertical directions. In 2016, we investigated the fact that these viewing biases are not universal, but are modulated by the semantic visual category of the stimulus. We showed that the joint distribution of saccade amplitudes and orientations significantly varies from one visual category to another. These joint distributions turn out to be, in addition, spatially variant within the scene frame. We demonstrated that a saliency model based on this better understanding of viewing behavioral biases and blind to any visual information outperforms well-established saliency models. We also proposed an extension of the saccadic model developed in 2015. The improvement consists in accounting for spatially-variant and context-dependent viewing biases. This model outperforms state-of-the-art saliency models, and provides scanpaths in close agreement with human behavior.

Inference of age from eye movements: We have presented evidence that information derived from eye gaze can be used to infer observers' age. From simple features extracted from the sequence of fixations and saccades, we predict the age of an observer. To reach this objective, we used the eye data from 101 observers split in 4 age groups (adults, 6-10 year-old, 4-6 year-old. and 2 year-old) to train a computational model. Participant's eye movements were monitored while participants were instructed to explore color pictures taken from children books for 10 seconds. The analysis of eye gaze provided evidence of age-related differences in viewing patterns. Fixation durations decreased with age while saccades turned out to be shorter when comparing children with adults. We combine several features, such as fixation durations, saccade amplitudes, and learn a direct mapping from those features to age using Gentle AdaBoost classifiers. Experimental results show that the proposed method succeeds in predicting reasonably well the observer's age.

7.1.2. *Graph structure in the rays space for fast light fields segmentation*
Participants: Christine Guillemot, Matthieu Hog.

In collaboration with Technicolor (Neus Sabater), we have introduced a novel graph representation for interactive light field segmentation using Markov Random Field (MRF). The greatest barrier to the adoption of MRF for light field processing is the large volume of input data. The proposed graph structure exploits the redundancy in the ray space in order to reduce the graph size, decreasing the running time of MRF-based optimisation tasks. The concepts of free rays and ray bundles with corresponding neighbourhood relationships are defined to construct the simplified graph-based light field representation. We have then developed a light field interactive segmentation algorithm using graph-cuts based on such ray space graph structure, that guarantees the segmentation consistency across all views. Our experiments with several datasets show results that are very close to the ground truth, competing with state of the art light field segmentation methods in terms of accuracy and with a significantly lower complexity. They also show that our method performs well on both densely and sparsely sampled light fields [18] (see Figure 1).

Input images
and scribbles   GT labels   Results of Wanner et al.   Our results
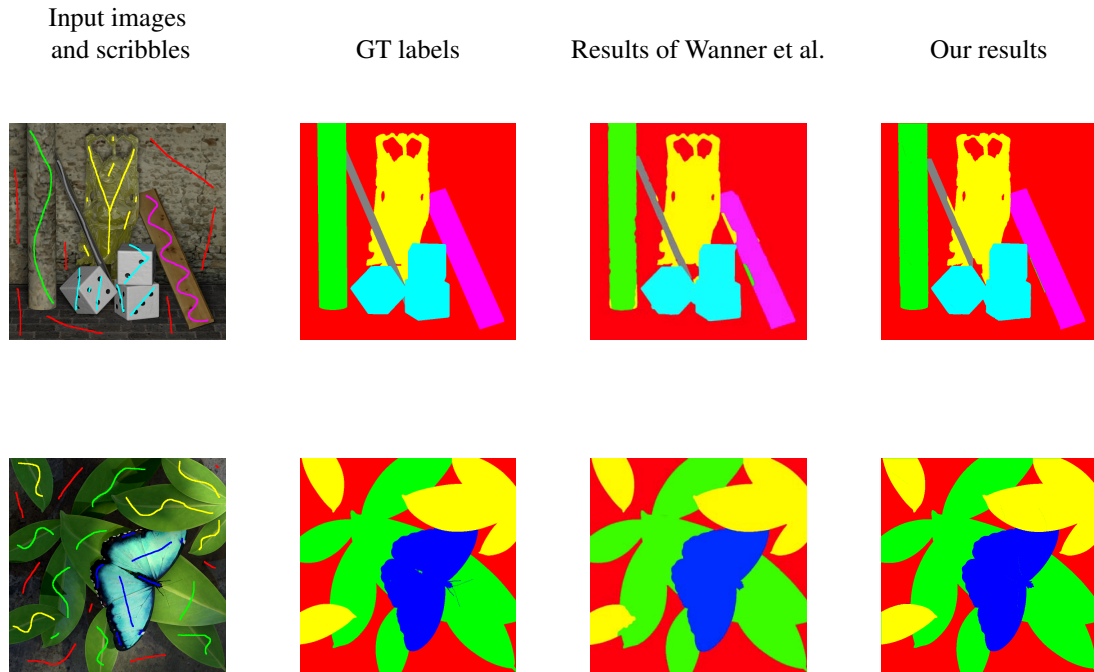

Figure 1. Light-field segmentation results obtained with synthetic light-fields. From left to right, we show, the input central view with scribbles, the ground truth labelling, the results of Wanner et al. and our results.

7.2. Rendering, inpainting and super-resolution

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

7.2.1. *Joint color and gradient transfer through Multivariate Generalized Gaussian Distribution*

Participants: Hristina Hristova, Olivier Le Meur.

Multivariate generalized Gaussian distributions (MGGDs) have aroused a great interest in the image processing community thanks to their ability to describe accurately various image features, such as image gradient fields, wavelet coefficients, etc. However, so far their applicability has been limited by the lack of a transformation between two of these parametric distributions. In collaboration with FRVSense (Rémi Cozot and Kadi Bouatouch), we have proposed a novel transformation between MGGDs, consisting of an optimal transportation of the second-order statistics and a stochastic-based shape parameter transformation. We employ the proposed transformation in both color and gradient transfers between images. We have also proposed a new simultaneous transfer of color and gradient.

7.2.2. *High-Dynamic-Range Image Recovery from Flash and Non-Flash Image Pairs*

Participants: Hristina Hristova, Olivier Le Meur.

In 2016, in collaboration with FRVSense (Rémi Cozot and Kadi Bouatouch), we have proposed a novel method for creating High Dynamic Range (HDR) images from only two images - flash and non-flash images. The proposed method consists of two main steps, namely brightness gamma correction and bi-local chromatic adaptation transform (CAT). First, the brightness gamma correction performs series of increases and decreases of the brightness of the non-flash image and that way yields multiple images with various exposure values. Second, a proposed CAT method, called bi-local CAT enhances the quality of the computed images, by recovering details in the under-/over-exposed regions, using detail information from the flash image. The

final multiple exposure images are then merged together to compute an HDR image. Evaluation shows that our HDR images, obtained by using only two LDR images, are close to HDR images, obtained by combining five manually taken multi-exposure images. The proposed method does not require the usage of a tripod and it is suitable for images of non-still objects, such as people, candle flames, etc. Figure 2 illustrates some results of the proposed method. The HDR-VDP-2 color-coded map (right-most image) shows the main luminance differences (the red areas) between our HDR result and the real HDR image. Snippets (a) and (b) show that the proposed method sharpens fine details, e.g. the net on the lamp. The net on the lamp of the real HDR image is blurry, due to a movement in the real multi-exposure images.

Figure 2. HDR image recovery from two input images, i.e. flash and non-flash images. Our HDR result and the real HDR image are tone-mapped for visualization on an LDR display.

7.2.3. Depth inpainting
Participant: Olivier Le Meur.

To tackle the disocclusion inpainting of RGB-D images appearing when synthesizing new views of a scene by changing its viewpoint, in collaboration with Pierre Buyssens from the Greyc laboratory from the Caen University, we have developed a new examplar-based inpainting method of depth map. The proposed method is based on two main components. First, a novel algorithm to perform the depth-map disocclusion inpainting has been proposed. In particular, this intuitive approach is able to recover the lost structures of the objects and to inpaint the depth-map in a geometrically plausible manner. Then, a depth-guided patch-based inpainting method has been defined in order to fill-in the color image. Depth information coming from the reconstructed depth-map is added to each key step of the classical patch-based algorithm from Criminisi et al. in an intuitive manner. Relevant comparisons to state-of-the-art inpainting methods for the disocclusion inpainting of both depth and color images have illustrated the effectiveness of the proposed algorithms.

7.2.4. Super-resolution and inpainting for face recognition
Participants: Reuben Farrugia, Christine Guillemot.

Most face super-resolution methods assume that low- and high-resolution manifolds have similar local geometrical structure, hence learn local models on the low-resolution manifold (e.g. sparse or locally linear embedding models), which are then applied on the high-resolution manifold. However, the low- resolution manifold is distorted by the one-to-many relationship between low- and high- resolution patches.

We have developed a method which learns linear models based on the local geometrical structure on the high-resolution manifold rather than on the low-resolution manifold. For this, in a first step, the low-resolution patch is used to derive a globally optimal estimate of the high-resolution patch. The approximated solution is shown to be close in Euclidean space to the ground-truth but is generally smooth and lacks the texture details needed by state-of-the-art face recognizers. Unlike existing methods, the sparse support that best estimates the first approximated solution is found on the high-resolution manifold. The derived support is then used to extract the atoms from the coupled dictionaries that are most suitable to learn an upscaling function between the low- and high- resolution patches.

The proposed solution has also been extended to compute face super-resolution of non-frontal images. Experimental results show that the proposed method out- performs six face super-resolution and a state-of-the-art cross- resolution face recognition method. These results also reveal that the recognition and quality are significantly affected by the method used for stitching all super-resolved patches together, where quilting was found to better preserve the texture details which helps to achieve higher recognition rates. The proposed method was shown to be able to super-resolve facial images from the IARPA Janus Benchmark A (IJB-A) dataset which considers a wide range of poses and orientations.

A method has also been developed to inpaint occluded facial regions with unconstrained pose and orientation. This approach first warps the facial region onto a reference model to synthesize a frontal view [15]. A modified Robust Principal Component Analysis (RPCA) approach is then used to suppress warping errors. It then uses a novel local patch-based face inpainting algorithm which hallucinates missing pixels using a dictionary of face images which are pre-aligned to the same reference model. The hallucinated region is then warped back onto the original image to restore missing pixels. Experimental results on synthetic occlusions demonstrate that the proposed face inpainting method has the best performance achieving PSNR gains of up to 0.74dB over the second-best method. Moreover, experiments on the COFW dataset and a number of real-world images show that the proposed method successfully restores occluded facial regions in the wild even for Closed-Circuit Television (CCTV) quality images.

7.2.5. *Light-field inpainting*
Participants: Christine Guillemot, Xiaoran Jiang, Mikael Le Pendu.

Building up on the advances in low rank matrix completion, we have developed a novel method for propagating the inpainting of the central view of a light field to all the other views. After generating a set of warped versions of the inpainted central view with random homographies, both the original light field views and the warped ones are vectorized and concatenated into a matrix. Because of the redundancy between the views, the matrix satisfies a low rank assumption enabling us to fill the region to inpaint with low rank matrix completion. To this end, a new matrix completion algorithm, better suited to the inpainting application than existing methods, has also been developed. Unlike most of the existing light field inpainting algorithms, our method does not require any depth prior. Another interesting feature of the low rank approach is its ability to cope with color and illumination variation between the input views of the light field (see Fig.3. As it can be seen in Figure 3, the proposed method yields inpainting consistency across views.

(a)　　　　(b)　　　　(c)　　　　(d)
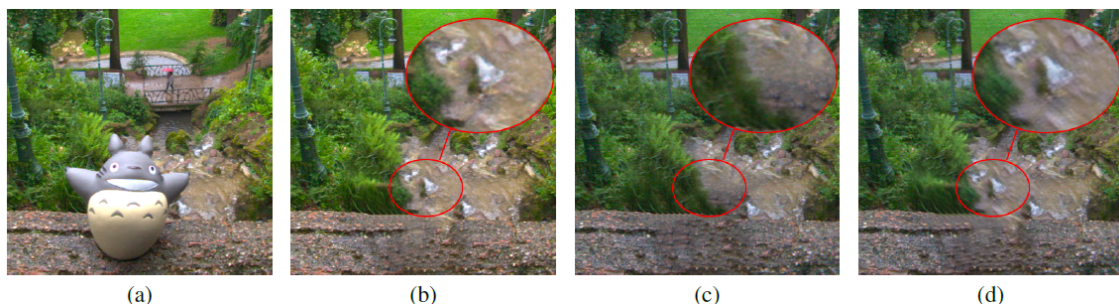
Figure 3. Illustration of our inpainting propagation method : (a) Original central view. (b) Inpainted central view. (c) Another view of the light field inpainted with a state-of-the-art 2D image inpainting method. (d) Propagated inpainting from central view to a different view with the developed low rank method.

7.3. Representation and compression of large volumes of visual data

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

7.3.1. *Graph-based multi-view video representation*

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

One of the main open questions in multiview data processing is the design of representation methods for multiview data, where the challenge is to describe the scene content in a compact form that is robust to lossy data compression. Many approaches have been studied in the literature, such as the multiview and multiview plus depth formats, point clouds or mesh-based techniques. All these representations contain two types of data: i) the color or luminance information, which is classically described by 2D images; ii) the geometry information that describes the scene 3D characteristics, represented by 3D coordinates, depth maps or disparity vectors. Effective representation, coding and processing of multiview data partly rely on a proper representation of the geometry information. The multiview plus depth (MVD) format has become very popular in recent years for 3D data representation. However, this format induces very large volumes of data, hence the need for efficient compression schemes. On the other hand, lossy compression of depth information in general leads to annoying rendering artefacts especially along the contours of objects in the scene. Instead of lossy compression of depth maps, we consider the lossless transmission of a geometry representation that captures only the information needed for the required view reconstructions.

The goal is thus to develop a Graph-Based Representation (GBR) for geometry information, where the geometry of the scene is represented as connections between corresponding pixels in different views. In this representation, two connected pixels are neighboring points in the 3D scene. The graph connections are derived from dense disparity maps and provide just enough geometry information to predict pixels in all the views that have to be synthesized. GBR drastically simplifies the geometry information to the bare minimum required for view prediction. This "task-aware" geometry simplification allows us to control the view prediction accuracy before coding compared to baseline depth compression methods. In 2015, we have first considered multi-view configurations, in which cameras are parallel.

In 2016, we have developed the extension of GBR to complex camera configurations. In [21], Xin Su has implemented a generalized Graph-Based Representation handling two views with complex translations and rotations between them (Fig. 4). The proposed approach uses the epipolar segments to have a row-wise description of the geometry that is as simple as for rectified views. This generalized GBR has been further extended to handle multiple views and scalable description of the geometry, *i.e.*, a geometry data that is coded as a function of the user navigation among the views.
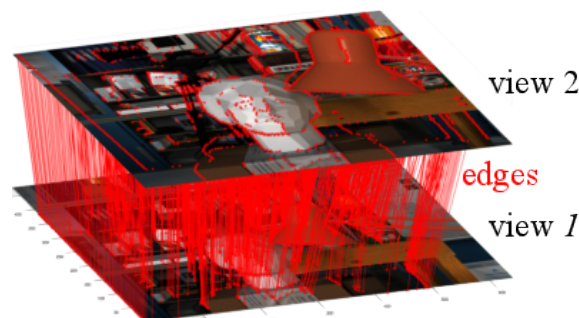

Figure 4. The proposed GBR (i) provides edges describing the geometry information and (ii) link pixels that are neighbors in the 3D scene.

The graph described above links neighboring pixels in the 3D scene as 3D meshes do. This meaningful structure might be used to code the color pixels lying on it. This can be done thanks to the new processing tools developed for signals lying on graphs. These tools rely however on covariance models that are assumed to be suited for the processed data. The PhD work of Mira Rizkallah is currently focussing on the effect of errors in the correlation models on the efficiency of the graph-based transforms.

7.3.2. *Sparse and low rank approximation of light fields*
Participants: Christine Guillemot, Xiaoran Jiang, Mikael Le Pendu.

We have studied the problem of low rank approximation of light fields for compression. A homography-based approximation method has been proposed which jointly searches for homographies to align the different views of the light field together with the low rank approximation matrices. We have first considered a global homography per view and shown that depending on the variance of the disparity across views, the global homography is not sufficient to well-align the entire images. In a second step, we have thus considered multiple homographies, one per region, the region being extracted using depth information. We have first shown the benefit of the joint optimization of the homographies together with the low-rank approximation. The resulting compact representation compressed using HEVC yields compression performance significantly superior to those obtained by directly applying HEVC on the light field views re-structured as a video sequence.

7.3.3. *Deep learning, autoencoders and neural networks for sparse representation and compression*
Participants: Thierry Dumas, Christine Guillemot, Aline Roumy.

Deep learning is a novel research area that attempts to extract high level abstractions from data by using a graph with multiple layers. One could therefore expect that deep learning might allow efficient image compression based on these high level features. However, deep learning, as classical machine learning, consists in two phases: (i) build a graph that can make a good representation of the data (i.e. find an architecture usually made with neural nets), and (ii) learn the parameters of this architecture from large-scale data. As a consequence, neural nets are well suited for a specific task (text or image recognition) and require one training per task. The difficulty to apply machine learning approach to image compression is that it is important to deal with a large variety of patches, and with also various compression rates. To test the ability of neural networks to compress images, we studied shallow sparse autoencoders (AE) for image compression in [14]. A performance analysis in terms of rate-distortion trade-off and complexity is conducted, comparing sparse AEs with LARS-Lasso, Coordinate Descent (CoD) and Orthogonal Matching Pursuit (OMP). A Winner Take All Auto-encoder (WTA AE) is proposed where image patches compete with one another when computing their sparse representation. This allows to spread the sparsity constraint on the whole image. Since the learning is made for this WTA AE, the neural network also learns to deal with various patches, which helps building a general-purpose AE. Finally, we showed that, WTA AE achieves the best rate-distortion trade-off, is robust to quantization noise and it is less complex than LARS-Lasso, CoD and OMP.

7.3.4. *Data geometry aware local basis selection*
Participants: Julio Cesar Ferreira, Christine Guillemot.

Local learning of sparse image models has proven to be very effective to solve a variety of inverse problems in many computer vision applications. To learn such models, the data samples are often clustered using the K-means algorithm with the Euclidean distance as a dissimilarity metric. However, the Euclidean distance may not always be a good dissimilarity measure for comparing data samples lying on a manifold.

In 2015, we have developed, in collaboration with Elif Vural (now Prof. at METU in Ankara, former postdoc in the team), two algorithms for determining a local subset of training samples from which a good local model can be computed for reconstructing a given input test sample, where we take into account the underlying geometry of the data. The first algorithm, called Adaptive Geometry-driven Nearest Neighbor search (AGNN), is an adaptive scheme which can be seen as an out-of-sample extension of the replicator graph clustering method for local model learning. The second method, called Geometry-driven Overlapping Clusters (GOC), is

a less complex nonadaptive alternative for training subset selection. The AGNN and GOC methods have been evaluated in image super-resolution, deblurring and denoising applications and shown to outperform spectral clustering, soft clustering, and geodesic distance based subset selection in most settings. The selected patches are used for learning good local bases using the traditional PCA method. PCA is considered an efficient tool to recover the tangent space of the patch manifold when the manifold is sufficiently regular.

However, when the patch manifold has high curvature, which is observed to be the case for images with high frequencies, PCA may not be suitable. It can be seen in Figure 5 that the PCA basis with respect to a manifold fails to approximate the tangent space as the manifold bends over itself. In other words, PCA basis is not adapted when the curvature is too high. On the other hand, it can be seen in Figure 5 that a union of subspaces with respect to a manifold might generate a local model that yields a more efficient local representation of data.

In 2016, we have proposed a strategy to choose between these two kinds of bases locally depending on the local data geometry. This function is defined as the variability of the tangent space in each cluster.
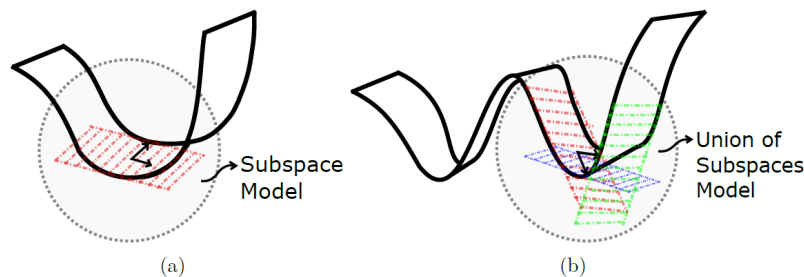

Figure 5. Subspaces computed with data sampled from a neighborhood on a manifold; (a): PCA basis which fails to approximate the subspace as the manifold curvature is too high; (b): union of subspaces generating a local model more coherent with the manifold geometry.

7.3.5. *Rate-distortion optimized tone curves for HDR video compression*

Participants: David Gommelet, Christine Guillemot, Aline Roumy.

High Dynamic Range (HDR) images contain more intensity levels than traditional image formats. Instead of 8 or 10 bit integers, floating point values requiring much higher precision are used to represent the pixel data. These data thus need specific compression algorithms. In collaboration with Ericsson [17], we have developed a novel compression algorithm that allows compatibility with the existing Low Dynamic Range (LDR) broadcast architecture in terms of display, compression algorithm and datarate, while delivering full HDR data to the users equipped with HDR display. The developed algorithm is thus a scalable video compression offering a base layer that corresponds to the LDR data and an enhancement layer, which together with the base layer corresponds to the HDR data. The novelty of the approach relies on the optimization of a mapping called Tone Mapping Operator (TMO) that maps efficiently the HDR data to the LDR data. The optimization has been carried out in a rate-distortion sense: the distortion of the HDR data is minimized under the constraint of minimum sum datarate (for the base and enhancement layer), while offering LDR data that are closed to some "aesthetic" a priori. Taking into account the aesthetic of the scene in video compression is indeed novel, since video compression is traditionally optimized to deliver the smallest distortion with the input data at the minimum datarate.

7.3.6. *Cloud-based image compression*

Participants: Jean Begaint, Christine Guillemot.

The emergence of cloud applications and web services has led to an increasing use of online resources. Image processing applications can benefit from this vast storage and distribution capacity. In collaboration with Technicolor, we investigate the use of this mass of redundant data to enhance image compression schemes. A region-based registration algorithm has been developed to capture complex deformations between two images. The registration method is then used to exploit both global and local correspondences between pairs of images of the same scene. The region-based registration yields a better prediction (hence reduced prediction errors, see Fig.6) which in turn yields a significant rate-distortion performance gain compared to current image coding solutions.
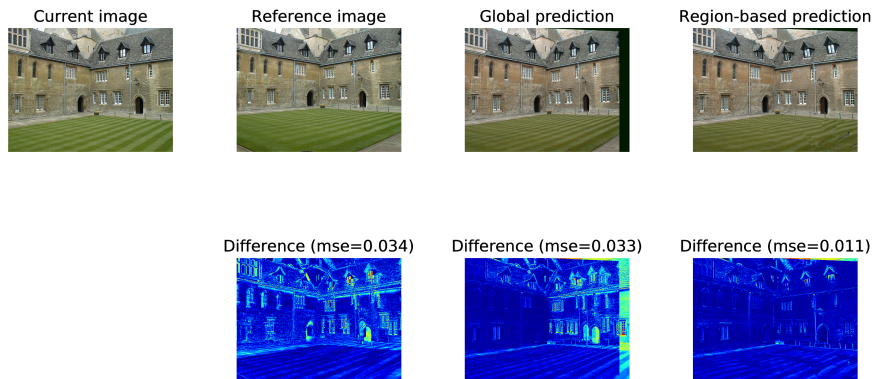

Figure 6. Image registration with global and region-based homographies and corresponding prediction error.

7.4. Distributed processing and robust communication

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

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

Participants: Thomas Maugey, Aline Roumy.

In order to have performing FTV systems, the data transmission has to take into account the interactivity of the user, *i.e.,* the viewpoint that is requested. In other words, a FTV system transmits to the visualisation support only what needs to be updated when a user changes its viewpoint angle (*i.e.*, the new information appearing in its vision field). The Sirocco has recently proposed some promising work using channel coding for interactive data coding. This coding scheme focusses on multi-view plus depth format only. In order to extend this approach to other formats, we have started a collaboration with the I3S laboratory in Nice, expert in 3D mesh compression.

The project ICON 3D funded by the GdR-Isis will be divided into two parts. First, we will study and develop 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 (Fig. 7). The prediction error will be characterized. Then, we will study the channel coding method that should be developed to correct this error.
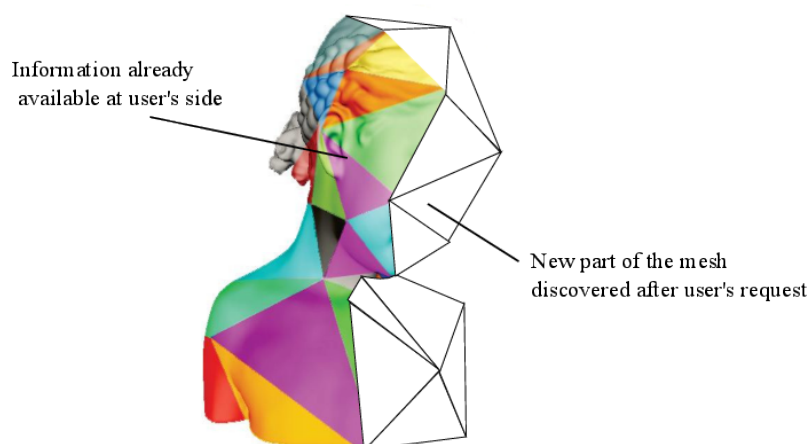
Figure 7. When a user changes his viewpoint angle, he discovers new part of the mesh that has to be transmitted.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Grants with Industry

8.1.1. Consulting contract with Enensys technologies
Participant: Aline Roumy.

- Title : Matrix inversion for video streaming.
- Research axis : REF AT 7.4. Distributed processing and robust communication
- Partners : Enensys, Inria-Rennes.
- Funding : Enensys.
- Period : Apr. 2016 - May 2016.

This contract with Enensys technologies aimed at studying solutions for reducing the complexity of matrix inversion used for encoding data in the context of video streaming. First a bibliographical study has been carried out related to the problem of matrix inversion in a finite field, then a novel solution has been proposed together with some recommendations regarding the algorithmic implementation.

8.1.2. Google faculty research award
Participants: Christine Guillemot, Xiaoran Jiang, Mikael Le Pendu.

- Title : Light fields low rank and sparse approximation
- Research axis : 7.3.2
- Partners : Inria-Rennes.
- Funding : Google.
- Period : Oct.2015-Sept.2016.

The goal of the project was to study low-rank and sparse approximation models for light fields compression. A homography-based low-rank approximation has been developed showing significant PSNR-rate gains compared to a direct encoding of light field views with HEVC-inter coding.

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

Participants: Christine Guillemot, David Gommelet, Aline Roumy.

- Title : LDR-compatible coding of HDR video signals.
- Research axis : § 7.3.5.
- Partners : Envivio.
- Funding : Cifre Envivio.
- Period : Oct.2014-Sept.2017.

The goal of this Cifre contract is to design solutions for LDR-compatible coding of HDR videos. This involves the study of rate-distortion optimized tone mapping operators taking into account constraints of temporal coherency to avoid the temporal flickering which results from a direct frame-by-frame application of classical tone mapping operators. The goal is also to design a coding architecture which will build upon these operators, integrating coding tools tailored to the statistics of the HDR refinement signals.

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

Participants: Maxime Rousselot, Olivier Le Meur.

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

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

8.1.5. CIFRE contract with Technicolor on image collection analysis

Participants: Dmitry Kuzovkin, Olivier Le Meur.

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

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

8.1.6. CIFRE contract with Technicolor on light fields editing

Participants: Christine Guillemot, Matthieu Hog.

- Title : Light fields editing
- Research axis : 7.1.2
- Partners : Technicolor, Inria-Rennes.
- Funding : Technicolor, ANRT.
- Period : Oct.2015-Sept.2018.

Editing is quite common with classical imaging. Now, if we want light-fields cameras to be in the future as common as traditional cameras, this functionality should also be enabled with light-fields. The goal of the PhD is to develop methods for light-field editing, and the work in 2016 has focused on the design of fast semi-supervised segmentation algorithms with coherence constraints across sub-aperture images (see 7.1.2).

8.1.7. CIFRE contract with Technicolor on light fields compressed representation

Participants: Christine Guillemot, Fatma Hawary.

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

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

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

Participants: Jean Begaint, Christine Guillemot.

- Title : Cloud-based image compression
- Research axis : 7.3.6
- Partners : Technicolor, Inria-Rennes.
- Funding : Technicolor, ANRT.
- Period : Nov.2015-Oct.2018.

The goal of this Cifre contract is to develop a novel image compression scheme exploiting similarity between images in a cloud. The objective will therefore be to develop rate-distortion optimized affine or homographic estimation and compensation methods which will allow us to construct prediction schemes and learn adapted bases from most similar images retrieved by image descriptors. One issue to be addressed is the rate-distortion trade-off induced by the need for transmitting image descriptors.

9. Partnerships and Cooperations

9.1. Regional Initiatives

9.1.1. CominLabs/InterCom project

Participants: Aline Roumy, Thomas Maugey.

- Title : Interactive Communication (INTERCOM): Massive random access to subsets of compressed correlated data .
- Research axis : 7.4.1
- Partners : Inria-Rennes (Sirocco team and i4S team); LabSTICC, Telecom Bretagne, Signal & Communications Department; External partner: Kieffer L2S, CentraleSupelec, Univ. Paris Sud.
- Funding : Labex CominLabs.
- Period : Oct. 2016 - Nov. 2019.

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

9.2. European Initiatives

9.2.1. FP7 & H2020 Projects

9.2.1.1. ERC-CLIM
Participants: Pierre David, Elian Dib, Christine Guillemot, Xin Su.

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. Informal International Partners

- Reuben Farrugia, Prof. at the University of Malta has been one sabbatical year Sept. 2015-Aug. 2016) within the team, working on inverse problems (super-resolution, inpainting) for several applications.
- The study on guided image inpainting is carried out in collaboration with Prof. Pascal Frossard from EPFL (Ecole Polytechique Federal de Lausanne).

9.4. International Research Visitors

9.4.1. Visits of International Scientists

Sheila Hemami, Prof. at Northeastern University, Boston, has visited the team during three months (May 2016-July 2016), working on the problem of demultiplexing and decoding of micro-lenses based light fields.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. Member of Organizing Committees
- C. Guillemot has served as area chair for the European Signal Processing Conference (EUSIPCO), 2016.

- C. Guillemot co-organized (together with Joachim Weickert, Prof. Universität des Saarlandes, and Thomas Pock, Prof. TU Graz, and Gerlind Plonka-Hoch, Prof. Universität Göttingen) a Schloss Dagstuhl seminar on inpainting-based image compression (Nov. 13-18, 2016).

10.1.1.2. Chair of Conference Program Committees

- C. Guillemot was technical co-chair of the IEEE image, video and multi-dimensional signal processing workshop, 11-12 July 2016
- C. Guillemot was general co-chair of the IEEE International workshop on Multimedia Signal Processing (IEEE-MMSP), Montreal, 21-23 Sept. 2016.

10.1.1.3. Member of Conference Program Committees

- O. Le Meur has been a member of technical program committees of international conferences: EUVIP 2016, QoMEX 2016
- A. Roumy has been a member of the technical program committee of the ACCV 2016 workshop on New Trends in Image Restoration and Enhancement.

10.1.2. Journal

10.1.2.1. Member of the Editorial Boards

- C. Guillemot has been associate editor of the Eurasip International Journal on Image Communication (2010-2016).
- C. Guillemot is senior area editor of the IEEE Trans. on Image Processing.
- C. Guillemot is associate editor of the International Journal on Mathematical Imaging and Vision.
- O. Le Meur is member of the editorial board of the IET Image Processing Journal.
- T. Maugey has been Guest editor for the Special Issue on Interactive Multi-view Video Services: from acquisition to Rendering, IEEE Multimedia Communication Technical Committee letters, Vol. 11(2), March 2016 (Guest Editors: Erhan Ekmekcioglu, Thomas Maugey, Laura Toni)
- A. Roumy is associate editor of the Springer Annals of Telecommunications.

10.1.3. Invited Talks

- O. Le Meur has been invited for a talk at Harmonic (Rennes) on "Computational Modelling of visual attention"
- O. Le Meur participated to the TechnoConference at Inria. The presentation dealt with spatio-temporal video inpainting.
- T. Maugey has been invited for a talk at I3S (Nice, France) on "Enabling user to interactively select viewpoint : a challenge for 3D data compression ", May 2016.
- C. Guillemot gave an overview talk on image compression at the Dagstuhl seminar (13-18 Nov. 2016).

10.1.4. Leadership within the Scientific Community

- C. Guillemot is member of the IEEE IVMSP technical committee
- C. Guillemot is senior member of the steering committee of IEEE Trans. on Multimedia (2016-2018).

10.1.5. Scientific Expertise

- C. Guillemot is member as scientific expert of the CCRRDT (Regional Committee of Research and Technological Development) of the Brittany region.

10.1.6. Research Administration

- C. Guillemot has been (Sept. 2015-Dec. 2016) vice-chair of Inria's evaluation committee.
- C. Guillemot is member of the "bureau du Comité des Projets".

- T. Maugey has been member of the selection comitee for the assigment of the ministerial PhD grants at IRISA.
- A. Roumy served as a member of Board of Examiners (Comité de sélection) for an Associate Professor position (Maitre de Conférences) at CREATIS Polytech Lyon (MCU61-46 2016).
- A. Roumy served as a member of Board of Examiners (Comité de sélection) for an Associate Professor position (Maitre de Conférences) at ENSEIRB-MATMECA Bordeaux (MCF61-0127 2016).
- 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, 8 hours, M2 computer science, Univ. of Rennes 1, France.

Master: C. Guillemot, Image and video compression, 8 hours, M2 SISEA, Univ. of Rennes 1, France.

Master: O. Le Meur, Selective visual attention, 6 hours, M2, Univ. of Paris 8, France.

Master: O. Le Meur, Acquisition/Image Processing/Compression, 22 hours, M2 MITIC, Univ. of Rennes 1, France.

Engineer degree: O. Le Meur, Image Processing, video analysis and compression, 54 hours, ESIR2, Univ. of Rennes 1, France.

Engineer degree: O. Le Meur, Visual communication, 65 hours, ESIR3, Univ. of Rennes 1, France.

Professional training: O. Le Meur, Image Processing and OpenCV, 42 hours, Technicolor Rennes.

Engineering degree: A. Roumy, Compressive sensing, 17 hours, INSA Rennes, 5th year, Mathematical engineering, France.

Master: A. Roumy, Compressive sensing, 15 hours, ENSAI Bruz, Master Big Data, France.

10.2.2. Supervision of PhD defended during the year

PhD-VAE: P. Bordes, Adapting video compression to new formats, Univ. of Rennes 1, 18 Jan. 2016, C. Guillemot.

PhD : M. Le Pendu, Backward compatible approaches for the compression of high dynamic range videos, Univ. of Rennes 1, 17 March 2016, C. Guillemot (Cifre contract with Technicolor).

PhD : J. C. Ferreira, Algorithms for Super-resolution of Images based on Sparse Representation and Manifolds, co-tutelle Univ. of Rennes 1/ University of Uberlandia Brazil, 6 July 2016, C. Guillemot.

PhD: C. Chamaret, Color Harmony: experimental and computational modeling, University of Rennes 1, 28th of April, 2016.

10.2.3. Juries

- C. Guillemot has been member (rapporteur) of the PhD jury of:
  - T. Biatek, INSA-Rennes, Apr. 2016
  - A. Akl, Univ. of Bordeaux, Feb. 2016
- C. Guillemot has been member of the PhD jury of:
  - A. Gilles, INSA-Rennes, Sept. 2016
  - L. LE Magoarou, Univ. Rennes 1, Nov. 2016
- O. Le Meur has been member (rapporteur) of the jury of the PhD committee of:
  - Romain Cohendet, University Bretagne Loire, IRCCyN, 2016

– Andrea Helo, University Paris Descartes, 2016
– Ala Aboudib, Telecom Bretagne, 2016
– Franck Chi, Telecom Bretagne, 2016

11. Bibliography

Publications of the year

Doctoral Dissertations and Habilitation Theses

[1] M. ALAIN. *A compact video representation format based on spatio-temporal linear embedding and epitome*, Université Rennes 1, January 2016, https://hal.inria.fr/tel-01261590

[2] J. C. FERREIRA. *Algorithms for super-resolution of images based on sparse representation and manifolds*, Université Rennes 1, July 2016, https://tel.archives-ouvertes.fr/tel-01388977

[3] M. LE PENDU. *Backward compatible approaches for the compression of high dynamic range videos*, Université Rennes 1, March 2016, https://tel.archives-ouvertes.fr/tel-01312901

Articles in International Peer-Reviewed Journals

[4] P. BUYSSENS, O. LE MEUR, M. DAISY, D. TSCHUMPERLÉ, O. LÉZORAY. *Depth-guided disocclusion inpainting of synthesized RGB-D images*, in "IEEE Transactions on Image Processing", 2017, vol. 26, n⁰ 2, pp. 525-538 [*DOI : 10.1109/TIP.2016.2619263*], https://hal.archives-ouvertes.fr/hal-01391065

[5] J. C. FERREIRA, E. VURAL, C. GUILLEMOT. *Geometry-Aware Neighborhood Search for Learning Local Models for Image Superresolution*, in "IEEE Transactions on Image Processing", March 2016, vol. 25, n⁰ 3, 14 p. , https://hal.inria.fr/hal-01388955

[6] Y. GAO, G. CHEUNG, T. MAUGEY, P. FROSSARD, J. LIANG. *Encoder-Driven Inpainting Strategy in Multiview Video Compression*, in "IEEE Transactions on Image Processing", January 2016, vol. 25, n⁰ 1, pp. 134-149, https://hal.inria.fr/hal-01217115

[7] S. KHATTAK, T. MAUGEY, R. HAMZAOUI, S. AHMAD, P. FROSSARD. *Temporal and Inter-view Consistent Error Concealment Technique for Multiview plus Depth Video Broadcasting*, in "IEEE Transactions on Circuits and Systems for Video Technology", May 2016, vol. 26, n⁰ 5, pp. 829-840 [*DOI : 10.1109/TCSVT.2015.2418631*], https://hal.inria.fr/hal-01137927

[8] O. LE MEUR, A. COUTROT. *Introducing context-dependent and spatially-variant viewing biases in saccadic models*, in "Vision Research", 2016, vol. 121, pp. 72 - 84 [*DOI : 10.1016/J.VISRES.2016.01.005*], https://hal.inria.fr/hal-01391745

[9] M. LE PENDU, C. GUILLEMOT, D. THOREAU. *Inter-Layer Prediction of Color in High Dynamic Range Image Scalable Compression*, in "IEEE Transactions on Image Processing", August 2016, vol. 25, pp. 3585 - 3596 [*DOI : 10.1109/TIP.2016.2571559*], https://hal.inria.fr/hal-01388961

[10] T. MAUGEY, G. PETRAZZUOLI, P. FROSSARD, M. CAGNAZZO, B. PESQUET-POPESCU. *Reference view selection in DIBR-based multiview coding*, in "IEEE Transactions on Image Processing", April 2016, vol. 25, n⁰ 4, pp. 1808-1819, https://hal.inria.fr/hal-01262258

[11] E. VURAL, C. GUILLEMOT. *Out-of-sample generalizations for supervised manifold learning for classification*, in "IEEE Transactions on Image Processing", March 2016, vol. 25, n[o] 3, 15 p. , https://hal.inria.fr/hal-01388959

Articles in Non Peer-Reviewed Journals

[12] A. ROUMY. *An Information theoretical problem in interactive Multi-View Video services*, in "IEEE Communications Society Multimedia Communications Technical Committee (ComSoc MMTC) E-Letter", March 2016, vol. 11, n[o] 2, 6 p. , https://hal.inria.fr/hal-01394864

International Conferences with Proceedings

[13] M. ALAIN, C. GUILLEMOT, D. THOREAU, P. GUILLOTEL. *Learning Clustering-Based Linear Mappings for Quantization Noise Removal*, in "IEEE International Conference on Image Processing (ICIP) 2016", Phoenix, United States, September 2016 [*DOI :* 10.1109/ICIP.2016.7533151], https://hal.inria.fr/hal-01317625

[14] T. DUMAS, A. ROUMY, C. GUILLEMOT. *Shallow sparse autoencoders versus sparse coding algorithms for image compression*, in "2016 IEEE International Conference on Multimedia and Expo (ICME 2016)", Seattle, WA, United States, July 2016, pp. 1 - 6 [*DOI :* 10.1109/ICMEW.2016.7574708], https://hal.archives-ouvertes.fr/hal-01377907

[15] R. FARRUGIA, C. GUILLEMOT. *Model and Dictionary guided Face Inpainting in the Wild*, in "ACCV workshop on New Trends in Image Restoration and Enhancement", Taipei, Taiwan, November 2016, 17 p. , https://hal.inria.fr/hal-01388971

[16] R. FARRUGIA, C. GUILLEMOT. *Robust Face Hallucination Using Quantization-Adaptive Dictionaries*, in "IEEE International Conference on Image Processing", Phoenix, United States, September 2016, 5 p. , https://hal.inria.fr/hal-01388972

[17] D. GOMMELET, A. ROUMY, C. GUILLEMOT, M. ROPERT, J. LETANOU. *Rate-distortion optimization of a tone mapping with SDR quality constraint for backward-compatible high dynamic range compression*, in "International Conference on Image Processing, ICIP", Phoenix, United States, 2016 IEEE International Conference on Image Processing (ICIP), September 2016, pp. 1384-1388 [*DOI :* 10.1109/ICIP.2016.7532585], https://hal.inria.fr/hal-01377701

[18] M. HOG, N. SABATER, C. GUILLEMOT. *Light Field Segmentation Using a Ray-Based Graph Structure*, in "European Conference on Computer Vision - ECCV", Amsterdam, Netherlands, October 2016, 16 p. , https://hal.inria.fr/hal-01388965

[19] O. LE MEUR, A. COUTROT. *How saccadic models help predict where we look during a visual task? Application to visual quality assessment*, in "SPIE Image Quality And System Performance", San Fransisco, United States, February 2016, https://hal.inria.fr/hal-01391750

[20] M. RIZKALLAH, T. MAUGEY, C. YAACOUB, C. GUILLEMOT. *Impact of Light Field Compression on Refocused and Extended Focus Images*, in "EUSIPCO 2016 - 24th European Signal Processing Conference", Budapest, Hungary, August 2016 [*DOI :* 10.1109/EUSIPCO.2016.7760378], https://hal.inria.fr/hal-01377834

[21] X. SU, T. MAUGEY, C. GUILLEMOT. *Graph-based representation for multiview images with complex camera configurations*, in "ICIP 2016 - IEEE International Conference on Image Processing", Phoenix, United States, September 2016, pp. 1554 - 1558 [*DOI :* 10.1109/ICIP.2016.7532619], https://hal.inria.fr/hal-01378422

National Conferences with Proceedings

[22] P. BUYSSENS, O. LE MEUR, M. DAISY, D. TSCHUMPERLÉ, O. LÉZORAY. *Désoccultation de cartes de profondeurs pour la synthèse de vues virtuelles*, in "RFIA 2016", Clermont-Ferrand, France, June 2016, https://hal.archives-ouvertes.fr/hal-01320967

Scientific Books (or Scientific Book chapters)

[23] M. MANCAS, O. LE MEUR. *Applications of Saliency Models*, in "From Human Attention to Computational Attention. A Multidisciplinary Approach", M. MANCAS, V. P. FERRERA, N. RICHE, J. G. TAYLOR (editors), Springer Series in Cognitive and Neural Systems, Springer, September 2016, vol. 10, pp. 331-377 [*DOI :* 10.1007/978-1-4939-3435-5_18], https://hal.inria.fr/hal-01393254

Other Publications

[24] A. COUTROT, O. LE MEUR. *Visual attention saccadic models: taking into account global scene context and temporal aspects of gaze behaviour*, August 2016, ECVP 2016 - European Conference on Visual Perception, Poster, https://hal.inria.fr/hal-01391751

[25] M. HOG, N. SABATER, C. GUILLEMOT. *Super-rays for Efficient Light Field Processing*, December 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01407852