IN PARTNERSHIP WITH:
CNRS

Université Rennes 1

Activity Report 2014

Project-Team SIROCCO

Analysis representation, compression and communication of visual data

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

RESEARCH CENTER
Rennes - Bretagne-Atlantique

THEME
Vision, perception and multimedia interpretation

Table of contents

Project-Team SIROCCO

Keywords: Image Processing, Signal Processing, Video, Sparse Representations

Creation of the Project-Team: 2012 January 01.

1. Members

Research Scientists

Christine Guillemot [Team leader, Inria, Senior Researcher, HdR]
Claude Labit [Inria, Senior Researcher, 20%, HdR]
Thomas Maugey [Inria, Researcher, since Nov 2014]
Aline Roumy [Inria, Researcher, 80% since March 2014]

Faculty Member

Olivier Le Meur [Univ. Rennes I, Associate Professor, HdR]

Engineer

Laurent Guillo [CNRS, Technical staff, 20%]

PhD Students

Jeremy Aghaei Mazaheri [Inria, granted by Astrium SAS]
Martin Alain [Inria, granted by CIFRE with Technicolor]
Marco Bevilacqua [Inria, granted by Alcatel Bell Labs., until April 2014]
Nicolas Dhollande [Univ. Rennes I, granted by CIFRE with Thomson Video Networks]
Mounira Ebdelli [Inria, granted by ANR-ARSSO project, until April 2014]
Julio Cesar Ferreira [co-tutelle Univ. Rennes I - Univ. Uberlandia Brazil, since Oct. 2013]
David Gommelet [Inria, granted by CIFFE with Envivio, from Oct. 2014]
Bihong Huang [Inria, granted by CIFRE with Orange Labs.]
Darya Khaustova [Inria, granted by CIFRE with Orange Labs.]
Mikael Le Pendu [Inria, granted by CIFRE with Technicolor]

Post-Doctoral Fellow

Elif Vural [Inria, since February 2014]

Visiting Scientist

Zhi Liu [Univ. Rennes I, until Aug. 2014]

Administrative Assistant

Huguette Béchu [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]

[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.

- 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 via 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.

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 in order 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 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.

3.4. 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

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

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:

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

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] (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:

[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.

- Compression with advanced functionalities of various image modalities (including multi-view, medical images such as MRI, CT, WSI, or satellite images);
- Networked multimedia applications via 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 with advanced functionalities

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. Medical Imaging (CT, MRI, Virtual Microscopy)

The use of medical imaging has greatly increased in recent years, especially with *magnetic resonance images (MRI) and computed tomography (CT)*. In the medical sector, lossless compression schemes are in general used to avoid any signal degradation which could mask a pathology and hence disturb the medical diagnosis. Nevertheless, some discussions are on-going to use near-lossless coding of medical images, coupled with a detection and segmentation of region-of interest (ROIs) guided by a modeling stage of the image sensor, by a precise knowledge of the medical imaging modalities and by the diagnosis and expertise of practitioners. New application domains using these new approaches of telemedecine will surely increase in the future. The second aspect deals with the legal need of biomedical images storage. The legacy rules of such archives are changing and it could be interesting to propose adaptive compression strategies, i.e to explore reversible lossy-to-lossless coding algorithms and new storage modalities which use, in a first stage, the lossless representation and continuously introduce controlled lossy degradations for the next stages of archives. Finally, it seems promising to explore new representation and coding approaches for 3D biological tissue imaging captured by *3D virtual microscopy*. These fields of interest and scientific application domains commonly generate terabytes of data. Lossless schemes but also lossy approaches have to be explored and optimized, and interactive tools supporting scalable and interactive access to large-sized images such as these virtual microscopy slides need to be developed.

4.5. 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. New Software and Platforms

5.1. Visual Fixation Analysis

Participant: Olivier Le Meur [contact person].

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. This software has been registered at the APP (Agence de Protection des Programmes).

5.2. Hierarchical super-resolution based inpainting

Participant: Olivier Le Meur [contact person].

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. This software is being registered at the APP (Agence de Protection des Programmes).

5.3. Salient object extraction

Participants: Zhi Liu, Olivier Le Meur [contact person].

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. This software is being registered at the APP (Agence de Protection des Programmes).

6. New Results

6.1. Analysis and modeling for compact representation and navigation

3D modelling, multi-view plus depth videos, Layered depth images (LDI), 2D and 3D meshes, epitomes, image-based rendering, inpainting, view synthesis

6.1.1. Salient object detection

Participants: Olivier Le Meur, Zhi Liu.

Salient object detection consists in extracting in an automatic manner the most interesting object in an image or video sequence. From an input image, an object, with well-defined boundaries, is detected based on its saliency. This subject knows an renewed interest these last years. A number of datasets serving as ground truth has been released and can be used to benchmark methods.

In 2013, we proposed a new method for detecting salient objects in still color images. In 2014, this method has been extended to video sequences [21]. Based on the superpixel representation of video frames, motion histograms and color histograms are computed at local and global levels. From these histograms, a superpixel-level temporal saliency measure as well as a spatial saliency measure are obtained. Finally, a pixel-level saliency derivation method is proposed to generate pixel-level temporal saliency map and spatial saliency map. An adaptive fusion method allows to integrate them into an unique spatiotemporal saliency map. Experimental results on two public datasets demonstrate that the proposed model outperforms state-of-the-art spatiotemporal saliency model in terms of both saliency detection and human fixation prediction.

6.1.2. Saliency aggregation

Participants: Olivier Le Meur, Zhi Liu.

In this study [32], we investigate whether the aggregation of saliency maps allows to outperform the best saliency models. Today there exist a number of saliency models for predicting the most visually salient locations within a scene. Although all existing models follow the same objective, they provide results which could be, to some extent, different. The discrepancies are related to the quality of the prediction but also to the saliency map representation. Indeed some models output very focused saliency maps whereas the distribution of saliency values is much more uniform in other models. Others tend to emphasize more on the image edges, the color or luminance contrast. This saliency map manifold contains a rich resource that should be used and from which new saliency maps could be inferred. Combining saliency maps generated using different models might enhance the prediction quality and the robustness of the prediction. Our goal is then to take saliency maps from this manifold and to produce the final saliency map.

This study discussed various aggregation methods; six unsupervised and four supervised learning methods are tested on two existing eye fixation datasets. Results show that a simple average of the TOP 2 saliency maps significantly outperforms the best saliency models. Considering more saliency models tends to decrease the performance, even when robust aggregation methods are used. Concerning the supervised learning methods, we provide evidence that it is possible to further increase the performance, under the condition that an image similar to the input image can be found in the training dataset. Our results might have an impact for critical applications which require robust and relevant saliency maps.

6.1.3. Models for 3D video quality assessment

Participants: Darya Khaustova, Olivier Le Meur.

This work is carried out in collaboration with Orange labs. The goal is to design objective metrics for quality assessment of 3D video content, by establishing links between human visual perception (visual comfort) and video parameters such as quality and depth quantity, and between visual comfort and visual attention. In 2013 we investigated the differences in 2D visual attention in comparison with 3D visual attention [31]. In 2014, we have focused on the design of an objective stereoscopic quality metric. In stereoscopic video quality, the assessment of spatial and temporal distortions by conventional quality metrics became incomplete because of the added depth dimension. Improperly captured or rendered, depth information can induce visual discomfort, impacting the overall video 3D QoE quality independently of image quality. The model is based on perceptual thresholds, namely visual annoyance, and acceptability. The visual annoyance threshold defines the boundary between annoying and not annoying sensation: 50% of subjects consider a stimulus annoying and 50% as not annoying. Acceptability determines the viewer's expectation level for the perceived video quality in a certain context and situation (inspired by the acceptability for the customer defined as an adequate service.

In order to compute the quality score, the proposed metric requires in input the distortion level of a technical and particular parameter, annoyance threshold and acceptability threshold of the targeted parameter. The performance of proposed objective mode is evaluated by considering five view asymmetries with five degradation levels. Generated contents were assessed by 30 subjects for each asymmetry (focal length mismatch, vertical shift, and rotation, green and white level reduction). The results of the subjective test have demonstrated that it is possible to classify detected problem to one of the objective categories using corresponding acceptability and visual annoyance thresholds.

6.1.4. Epitome-based video representation

Participants: Martin Alain, Christine Guillemot.

In 2014, we have developed fast methods for constructing epitomes from images. An epitome is a factorized texture representation of the input image, and its construction exploits self-similarities within the image. Known construction methods are memory and time consuming. The proposed methods, using dedicated list construction on one hand and clustering techniques on the other hand, aim at reducing the complexity of the search for self-similarities. Experiments show that interesting complexity results can be obtained without degrading the epitome quality for both proposed methods. By limiting the number of exhaustive searches we limit the memory occupation and the processing time, while keeping a good epitome quality (down to 18.08 % of the original memory occupation and 41.39 % of the original processing time) [25]. As an example, images reconstructed using the different techniques are visible in Fig. 1. The epitome construction method is currently being extended from still images to groups of images in video sequences. Denoising and super-resolution algorithms based on the constructed epitomes are also under study.

6.1.5. Light field tomographic reconstruction from a fixed camera focal stack

Participants: Christine Guillemot, Elif Vural.

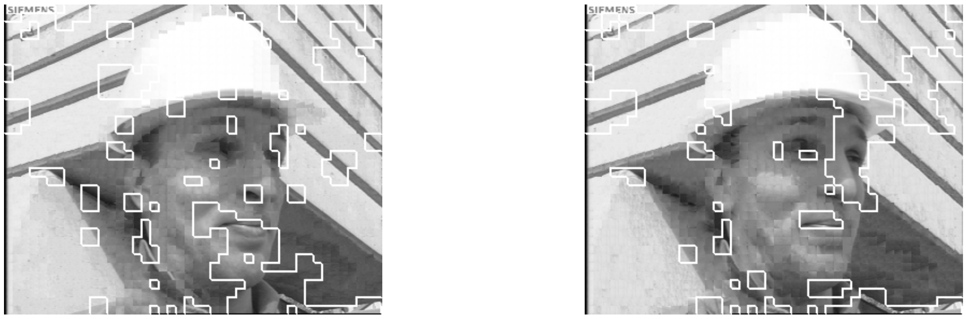

Figure 1. Reconstructed images using the list-based (left) and clustering-based methods. Epitome patches are highlighted in white.

Thanks to the internship of Antoine Mousnier (student at Ecole Centrale Lyon), we have developed a novel approach to partially reconstruct high-resolution 4D light fields from a stack of differently focused photographs taken with a fixed camera. First, a focus map is calculated from this stack using a simple approach combining gradient detection and region expansion with graph cut. Then, this focus map is converted into a depth map thanks to the calibration of the camera. We proceed after this with the tomographic reconstruction of the epipolar images by back-projecting the focused regions of the scene only. We call it masked back-projection. The angles of back-projection are calculated from the depth map. Thanks to the high angular resolution we achieve, we are able to render puzzling perspective shifts although the original photographs were taken from a single fixed camera at a fixed position and render images with extended focus (see Fig. 2). To the best of our knowledge, our method is the first one to reconstruct a light field by using a focal stack captured with an ordinary camera at a fixed viewpoint.

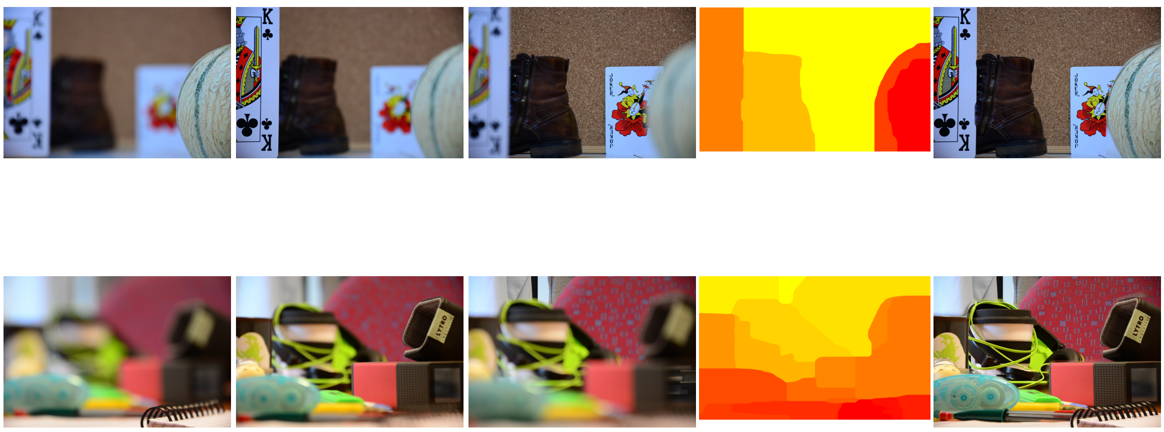

Figure 2. Three images of the focal stack (left); estimated depth map and image with extended focus (right). The focal stack images of the first and second rows have been captured with a Nikon 5200 camera.

6.2. Rendering, inpainting and super-resolution

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

6.2.1. Video inpainting

Participants: Mounira Ebdelli, Christine Guillemot, Olivier Le Meur.

Image (and video) inpainting refers to the process of restoring missing or damaged areas in an image (or a video). This field of research has been very active over the past years, boosted by numerous applications: restoring images from scratches or text overlays, loss concealment in a context of impaired image transmission, object removal in a context of editing, disocclusion in image-based rendering of viewpoints different from those captured by the cameras. Inpainting is an ill-posed inverse problem: given observations, or known samples in a spatial (or spatio-tempoal) neighborhood, the goal is to estimate unknown samples of the region to be filled in. Many methods already exist for image inpainting, either based on PDE (Partial Derivative Equation)-based diffusion schemes, either using sparse or low rank priors or following texture synthesis principles exploiting statistical or self-similarity priors.

In 2014, the problem of video inpainting has been further addressed with free-moving cameras. The algorithm developed first compensates the camera motion between the current frame and its neighboring frames in a sliding window, using a new region-based homography computation which better respects the geometry of the scene compared to state-of-the-art methods. The source frame is first segmented into regions in order to find homogeneous regions. Then, the homography for mapping each region into the target frame is estimated. The overlapping of all aligned regions forms the registration of the source frame into the target one. Once the neighboring frames have been aligned, they form a stack of images from which the best candidate pixels are searched in order to replace the missing ones. The best candidate pixel is found by minimizing a cost function which combines two energy terms. One energy term, called the data term, captures how stationary is the background information after registration, hence enforcing temporal coherency. The second term aims at favoring spatial consistency and preventing incoherent seams, by computing the energy of the difference between each candidate pixel and its 4-neighboring pixels in the missing region. The minimization of the energy term is performed globally using Markov Random Fields and graph cuts. A method of Poisson blending has been implemented in order to further enhance the visual quality of the inpainted videos. The proposed approach, although less complex than state-of-the-art methods, provides more natural results.

6.2.2. Image and video super-resolution in the example-based framework

Participants: Marco Bevilacqua, Christine Guillemot, Aline Roumy.

Super-resolution (SR) refers to the problem of creating a high-resolution (HR) image, given one or multiple low-resolution (LR) images as input. The SR process aims at adding to the LR input(s) new plausible high frequency details, to a greater extent than traditional interpolation methods. We mostly focused on the single-image problem, where only a single LR image is available. We have adopted the example-based framework on one hand and the sparse approximation framework on the other hand.

In the example-based framework, the relation between the LR and HR image spaces is modeled with the help of pairs of small "examples", i.e. texture patches. Each example pair consists of a LR patch and its HR version that also includes high-frequency details; the pairs of patches form a dictionary of patches. For each patch of the LR input image, one or several similar patches are found in the dictionary, by performing a nearest neighbor search. The corresponding HR patches in the dictionary are then combined to form a HR output patch; and finally all the reconstructed HR patches are re-assembled to build the superresolved image. In this procedure, one important aspect is how the dictionary of patches is built. At this regard, two choices are possible: an external dictionary, formed by sampling HR and LR patches from external training images; and an internal dictionary, where the LR/HR patch correspondences are learned by putting in relation directly the input image and scaled versions of it. The advantage of having an external dictionary is that it is built in advance: this leads to a reduction of the computational time, whereas in the internal case the dictionary is generated online at each run of the algorithm. However, external dictionaries have a considerable drawback: they are fixed and so non-adapted to the input image. To be able to satisfactorily process any input image, we need then to include in the dictionary a large variety of patch correspondences, leading to a high computational time. In 2013, external dictionaries have been designed to bridge the gap between external and internal dictionary based methods.

In 2014 instead, we proposed a novel SR method for internal dictionaries [16]. The internal dictionary contains pair of LR/HR patches taken from the image to be processed and is by construction well adapted to the data. However, its size is limited since it results from the sampling of a single image. This leads to an undersampling of the LR space and even more of the HR space. To overcome this problem, state of the art methods select, for each input LR patch, a local neighborhood, learn the local geometry of this neighborhood, and apply it in the HR domain. Therefore, an underlying hypothesis is that the local neighborhoods in the LR and HR domain are similar. To avoid this hypothesis, we employ a regression-based method to directly map LR input patches into their related HR output patches. To make this regression more robust, first the LR patches have been first oversampled (by a bicubic interpolation) such that LR and HR spaces have the same dimension, and second a Tikhonov regularization has been added. When compared to other state-of-the-art algorithms, our proposed algorithm shows the best performance, both in terms of objective metrics and subjective visual results. As for the former, it presents considerable gains in PSNR and SSIM values. When observing the super-resolved images, also, it turns out to be the most capable in producing fine artifact-free HR details.

6.2.3. Image super-resolution in a sparse and manifold learning framework

Participants: Julio Cesar Ferreira, Christine Guillemot, Olivier Le Meur, Elif Vural.

The problem of image super-resolution has also been addressed in a sparse approximation framework. This led to a novel algorithm based on sparse representations in which a structure tensor-based regularization has been introduced [29]. The relative discrepancy between the two eigenvalues of the structure tensor is an indicator of the degree of anisotropy of the gradient in a region of the image. The eigenvalues and eigenvectors of the structure tensor are used to compute, for each pixel belonging to a salient edge, a stream line in the direction perpendicular to the edge (given by the eigenvector corresponding to the highest eigenvalue of the structure tensor). The saliency of an edge is given by the S-norm of the highest eigenvalue. An energy term dealing with the sharpness of edges is then computed and used as a regularization constraint to modify the current estimated high resolution image inside the Iterative Shrinkage Thresholding algorithm. This extra constraint forces the value of the current pixel along the stream line to be as close as possible to pixel values having lowest saliency. The resulting single-image algorithm, called Sharper Edges based Adaptive Sparse Domain Selection (SE-ASDS) allows sharpening edges and reducing the ringing artefacts compared to existing methods. This is illustrated in Fig.3

In the previous method, the dictionaries used for the sparse approximation method are defined as a union of PCA basis learned on clusters of patches of the input image. The clusters are constructed using the classical k-means algorithm with patch distances computed with the Euclidean distance. This study is being pursued by assuming manifold models for the patches of the input images. A method using graph-based clustering has then been used for clustering patches on the manifold, and this method has been extended to cope with the out-of-sample problem. Dedicated dictionary learning methods are currently under development to have dictionaries best adapted to the manifold structure.

6.3. Representation and compression of large volumes of visual data

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

6.3.1. Manifold learning and low dimensional embedding for classification

Participants: Christine Guillemot, Elif Vural.

Typical supervised classifiers such as SVM are designed for generic data types and do not make any particular assumption about the geometric structure of data, while data samples have an intrinsically low-dimensional structure in many data analysis applications. Recently, many supervised manifold learning methods have been proposed in order to take the low-dimensional structure of data into account when learning a classifier. Unlike unsupervised manifold learning methods which only take the geometric structure of data samples into account when learning a low-dimensional representation, supervised manifold learning methods learn an embedding that not only preserves the manifold structure in each class, but also enhances the separation between different classes.

Figure 3. Comparison of SR results (×3). (a) LR image;(b) Nearest-neighbor; (c) Sparse method without structure-based regularization; (d) SE-ASDS results. (e) Comparison between (c) and (d) on patches: edges of (d) are more contrasted than (c).

An important factor that influences the performance of classification is the separability of different classes in the computed embedding. We thus do a theoretical analysis of separability of data representations given by supervised manifold learning. In particular, we focus on the nonlinear supervised extensions of the Laplacian eigenmaps algorithm and examine the linear separation between different classes in the learned embedding. We first consider a setting with two classes and show that the two classes become linearly separable even with a one-dimensional embedding. We characterize the linear separation in terms of the data graph properties such as edge weights, diameter, and volume and some algorithm parameters. We then extend these results to a setting with multiple classes, where the classes are assumed to be categorizable into a few groups with high intra-group affinities. We show that, if the graph is such that the inter-group graph weights are sufficiently small, the learned embedding becomes linearly separable at a dimension that is proportional to the number of groups. These theoretical findings are also confirmed by experimentation on synthetic data sets and image data.

Next, we consider the problem of out-of-sample generalizations for manifold learning. Most manifold learning methods compute an embedding in a pointwise manner, i.e., data coordinates in the learned domain are computed only for the initially available training data. The generalization of the embedding to novel data samples is an important problem, especially in classification problems. Previous works for out-of-sample generalizations are designed for unsupervised methods. We study the problem for the particular application of data classification and propose an algorithm to compute a continuous function from the original data space to the low-dimensional space of embedding. In particular, we construct an interpolation function in the form of a radial basis function that maps input points as close as possible to their projections onto the manifolds of their own class. Experimental results show that the proposed method gives promising results in the classification of low-dimensional image data such as face images.

6.3.2. Dictionary learning for sparse coding and classification of satellite images

Participants: Jeremy Aghaei Mazaheri, Christine Guillemot, Claude Labit.

In the context of the national partnership Inria-Astrium, we explore novel methods to encode images captured by a geostationary satellite. These pictures have to be compressed on-board before being sent to earth. Each picture has a high resolution and so the rate without compression is very high (about 70 Gbits/sec) and the goal is to achieve a rate after compression of 600 Mbits/sec, that is a compression ratio higher than 100. On earth, the pictures are decompressed with a high reconstruction quality and visualized by photo-interpreters. The goal of the study is to design novel transforms based on sparse representations and learned dictionnaries for satellite images.

We have developed methods for learning adaptive tree-structured dictionaries. Each dictionary in the structure is learned on a subset of residuals from the previous level, with the K-SVD algorithm. The tree structure offers better rate-distortion performance than a "flat" dictionary learned with K-SVD, especially when only a few atoms are selected among the first levels of the tree. The tree-structured dictionary allows efficient coding of the indices of the selected atoms. Besides coding, these structured dictionaries turn out to be useful tools for MTF (Modulation Transfer Function) estimation and supervised classification. The MTF estimation consists in estimating the MTF of the instrument used to take this picture. The learned structured dictionaries are currently studied to perform supervised classification in a context of scene recognition in satellite images. In that case, dictionaries should be learned for specific scenes. Then, patches (around each pixel) of a test picture to classify are decomposed over the different dictionaries to determine for each pixel the dictionary giving the best approximation and thus the corresponding class. A graph-cut algorithm can be applied to smooth the classification results. We are currently trying to learn more discriminant dictionaries for this specific application. For that purpose, the objective function to minimize to learn the dictionaries should not only be reconstructive, but also discriminative.

6.3.3. Adaptive clustering with Kohonen self-organizing maps for second-order prediction

Participants: Christine Guillemot, Bihong Huang.

The High Efficiency Video Coding standard (HEVC) supports a total of 35 intra prediction modes which aim at reducing spatial redundancy by exploiting pixel correlation within a local neighborhood. However the correlation remains in the residual signals of intra prediction, leading to some high energy prediction residuals. In 2014, we have studied several methods to exploit remaining correlation in residual domain after intra prediction. The method uses vector quantization with codebooks learned and dedicated to the different prediction modes in order to model the directional characteristics of the residual signals. The best matching code vector is found in a rate-distortion optimization sense. Finally, the index of the best matching code vector is sent to the decoder and the vector quantization error, the difference between the intra residual vector and the best matching code vector, is processed by the conventional operations of transform, scalar quantization and entropy coding. In a first approach, the codebooks are learned using the k-means algorithm. The learning algorithm proceeds in two passes so that the training set of residual vectors corresponds to the case where the vector quantization is the best mode in rate-distortion sense for the second-order prediction. It has been observed that the codebooks learned for different Quantization Parameters (QP) are very similar, leading eventually to QP-independent codebooks. A second method is being developed using clustering with Kohonen self-organizing maps in the codebook learning stage.

6.3.4. HDR video compression

Participants: Christine Guillemot, Mikael Le Pendu.

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, leading to new compression challenges. In collaboration with Technicolor, we have developed a method for converting the floating point RGB values to high bit depth integers with an approximate logarithmic encoding that is reversible without loss. This bit depth reduction is performed adaptively depending on the minimum and maximum values which characterize the dynamic of the data. A 50% rate saving has been obtained at high bitrates compared to the well-known adaptive LogLuv transform [33]. A reversible tone mapping-operator (TMO) has also been designed for efficient compression of High Dynamic Range (HDR) images using a Low Dynamic Range (LDR) encoder. Based on a statistical model of the HDR compression scheme and assumptions on the rate of the encoded LDR image, a closed form solution has been derived for the optimal tone curve in a rate-distortion sense [34].

6.3.5. HEVC-based UHD video coding optimization

Participants: Nicolas Dhollande, Christine Guillemot, Olivier Le Meur.

The HEVC (High Efficiency Video Coding) standard brings the necessary quality versus rate performance for efficient transmission of Ultra High Definition formats (UHD). However, one of the remaining barriers to its adoption for UHD content is the high encoding complexity. We address the problem of HEVC encoding complexity reduction by proposing a strategy to infer UHD coding modes and quadtree from those optimized on the lower (HD) resolution version of the input video. A speed-up by a factor of 3 is achieved compared to directly encoding the UHD format at the expense of a limited PSNR-rate loss [28]. Another method which is still under investigation is to extract from the input video sequence a number of low-level features for adapting the coding decision such as the decomposition of the quadtree. The low-level features are related to gradient-based statistics, structure tensors statistics or entropy etc.

6.4. 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

6.4.1. Universal distributed source coding

Participant: Aline Roumy.

In 2012, we started a new collaboration with Michel Kieffer and Elsa Dupraz (Supelec, L2S) on universal distributed source coding. Distributed source coding (DSC) refers to the problem where several correlated sources need to be compressed without any cooperation at the encoders. Decoding is however performed jointly. This problem arises in sensor networks but also in video compression techniques, where the successive frames are seen as distributed such that the correlation between the frames is not directly used at the encoder. Traditional approaches for DSC (from an information theoretical but also practical point of view) assume that the joint distribution of the sources is perfectly known. Since this assumption is not satisfied in practice, a way to get around this is to use a feedback channel (from the decoder to the encoder), that can trigger the encoder. Instead, we consider universal distributed source coding, where the joint source distribution is unknown.

More precisely, we considered the problem of compressing one source, while a second source, called side information, is available at the decoder. Further, we assumed that the conditional distribution of the side information given the source is unknown at both encoder and decoder. First, we proposed in [18] four uncertainty models for this conditional distribution, and derived the information theoretical bounds. These models differ through the (partial) knowledge on the distribution the user has. This partial knowledge includes the variation speed (slow/fast), the set of possible distributions, and eventually, some a priori distribution on the class of distributions. A complete coding scheme has also been proposed that works well for any distribution in the class. At the encoder, the proposed scheme encompasses the determination of the coding rate and the design of the encoding process. These determinations directly result from the information-theoretical compression bounds. Then a novel decoder is proposed that jointly estimate the source symbols and the conditional distribution. As the proposed decoder is based on the Expectation-Maximization algorithm, which is very sensitive to initialization, we also propose a method to produce first a coarse estimate of the distribution. The proposed scheme avoids the use of a feedback channel or the transmission of a learning sequence, which both result in a rate increase at finite length. Moreover, the proposed algorithm use non-binary LDPC codes, such that the usual binarization of the source, which induce compression inefficiency, can be avoided.

6.4.2. Rate Distortion analysis of Compressed sensing and distributed Compressed sensing

Participant: Aline Roumy.

In collaboration with Enrico Magli and Giulio Coluccia (Polito, Torino, Italy), we studied Compressed sensing as a communication tool. Compressed sensing (CS) is an efficient acquisition scheme, where the data are projected onto a randomly chosen subspace to achieve data dimensionality reduction. The projected data are called measurements. The reconstruction is performed from these measurements, by solving underdetermined linear systems under a sparsity a priori constraint. However, the obtained measurements are reals, and therefore require an infinite precision representation. Therefore, using CS as a compression tool (in the information theoretical sense), requires to determine the trade-off between the rate necessary to encode the measurements and the distortion obtained on the data. In [17], we derive the rate-distortion (RD) function of CS and distributed CS, under the assumption that the sparsity support is perfectly known at the decoder. This provides a lower bound for any practical reconstruction algorithm.

The proof technique developed in [17] has application beyond information theory. It also provides novel analyses of CS reconstruction algorithms [27]. Classical performance analysis of reconstruction algorithms, rely on parameters that are difficult to compute (RIP, coherence of the measurement matrix), for which bounds are used. Instead, we derive exact characterization, by performing either averaged (over the measurement matrix) or asymptotic (in the size of the data) analysis.

7. Bilateral Contracts and Grants with Industry

7.1. Bilateral Contracts with Industry

7.1.1. Contract with Astrium on compression of satellite images

Participants: Jeremy Aghaei Mazaheri, Christine Guillemot, Claude Labit.

- Title : Compression of satellite images.
- Research axis : § 6.3.2.
- Partners : Astrium, Inria-Rennes.
- Funding : Astrium.
- Period : Oct.11-Sept.14.

This contract with Astrium addresses the problem of sparse representation and dictionary learning for efficient sparse coding of video signals captured from a geostationary satellite. The goal is to develop a compact spatio-temporal representation taking advantage of the high redundancy present in the video which is of very high resolution and characterized by low motion. Different methods for learning tree-structured dictionaries have been studied. The tree-structured dictionaries are well-tailored to the characteristics of the signals to be processed at each iteration of the greedy matching pursuit algorithms, while allowing efficient encoding of the produced sparse vectors. Adaptive tree-structures have been developed and the use of such dictionaries in HEVC-based intra coding has been investigated. First tests have also been carried out to known to which extent the learned dictionnaries can allow detecting the modulation transfer function (MTF) used to characterize the quality of electro-optical imaging systems on board remote sensing satellites.

7.2. Bilateral Grants with Industry

7.2.1. Contract with EutelSat on video traffic analysis

Participants: Laurent Guillo, Aline Roumy.

- Title : Bit rate statistical analysis of HEVC encoded video in a broadcast transmission.
- Partners : EutelSat, Inria-Rennes.
- Funding : EutelSat.
- Period : Aug.12-Mar.14.

This contract with EutelSat (starting in August 2012) is a consulting contract and aims at analyzing the variation of the video traffic, when the video is encoded by HEVC. Indeed, the main characteristic of satellite broadcasting, as proposed by Eutelsat, is to provide a nearly constant video quality, which is obtained by variable video traffic (bit rate). Then, to address this variability issue, statistical multiplexing is used to share the resource among the users. However, statistical multiplexing needs a precise analysis of this variability. In this contract, we therefore analyze this variability, when the video is compressed with the upcoming video compression standard HEVC.

7.2.2. CIFRE contract with Orange on Generalized lifting for video compression

Participants: Christine Guillemot, Bihong Huang.

- Title : Generalized lifting for video compression.
- Research axis : § 6.3.3.
- Partners : Orange Labs, Inria-Rennes, UPC-Barcelona.
- Funding : Orange Labs.
- Period : Apr.2012-Mar.2015.

This contract with Orange labs. (started in April. 2012) concerns the PhD of Bihong Huang and aims at modelling the redundancy which remains in spatial and temporal prediction residues. The analysis carried out in the first year of the PhD has shown that this redundancy (hence the potential rate saving) is high. In 2013, different methods have been investigated to remove this redundancy, such as generalized lifting and different types of predictors. The generalized lifting is an extension of the lifting scheme of classical wavelet transforms which permits the creation of nonlinear and signal probability density function (pdf) dependent and adaptive transforms. This study is also carried out in collaboration with UPC (Prof. Philippe Salembier) in Barcelona.

7.2.3. CIFRE contract with Orange on 3D quality assessment

Participants: Darya Khaustova, Olivier Le Meur.

- Title : Objective Evaluation of 3D Video Quality.
- Research axis : § 6.1.3.
- Partners : Orange Labs, Inria-Rennes.
- Funding : Orange Labs.
- Period : Dec.2011-Nov.2014.

This contract with Orange labs. (starting in Dec. 2011) concerns the PhD of Darya Khaustova and aims at developping a video quality metric for 3D content. The usage of 3D video is expected to increase in the next years. In order to ensure a good QoE (Quality of Experience), the 3D video quality must be monitored and accuratly measured. The goal of this thesis is to study objective measures suitable for estimating 3D video quality. A comparison with ground truth as well as with the state-of-the-art 2D metrics should be carried out. To be as effective as possible, the feature of the human visual system should be taken into account.

7.2.4. CIFRE contract with Technicolor on High Dynamic Range (HDR) video compression

Participants: Mikael Le Pendu, Christine Guillemot.

- Title : Floating point high dynamic range (HDR) video compression
- Research axis : § 6.3.4.
- Partners : Technicolor, Inria-Rennes.
- Funding : Technicolor, ANRT.
- Period : Dec.2012-Nov.2015.

High Dynamic Range (HDR) images contain more intensity levels than traditional image formats, leading to higher volumes of data. HDR images can represent more accurately the range of intensity levels found in real scenes, from direct sunlight to faint starlight. The goal of the thesis is to design a visually lossless compression algorithm for HDR floating-point imaging data. The first year of the thesis has been dedicated to the design of a quantization method converting the floating point data into a reduced bit depth representation, with minimal loss. The method leads to a bit rate saving of 50% compared to the existing Adaptive LogLuv transform.

7.2.5. CIFRE contract with Technicolor on sparse modelling of spatio-temporal scenes

Participants: Martin Alain, Christine Guillemot.

- Title : Spatio-temporal analysis and characterization of video scenes
- Research axis : § 6.1.4.
- Partners : Technicolor, Inria-Rennes.
- Funding : Technicolor, ANRT.
- Period : Oct.2012-Sept.2015.

A first CIFRE contract has concerned the Ph.D of Safa Cherigui from Nov.2009 to Oct.2012, in collaboration with Dominique Thoreau (Technicolor). The objective was to investigate texture and video scene characterization using models based on sparse and data dimensionality reduction techniques, as well as based on epitomes. The objective was then to use these models and methods in different image processing problems focusing in particular on video compression. While, the first PhD thesis has focused on spatial analysis, processing, and prediction of image texture, a second CIFRE contract (PhD thesis of Martin Alain) has started in Oct. 2012 to push further the study by addressing issues of spatio-temporal analysis and epitome construction, with applications to temporal prediction, as well as to other video processing problems such as denoising and super-resolution.

7.2.6. CIFRE contract with Thomson Video Networks (TVN) on Video analysis for HEVC based video coding

Participants: Nicolas Dhollande, Christine Guillemot, Olivier Le Meur.

- Title : Coding optimization of HEVC by using pre-analysis approaches.
- Research axis : § 6.3.3.
- Partners : Thomson Video Networks, Univ. Rennes 1.
- Funding : Thomson Video Networks (TVN).
- Period : Nov.2012-Sept.2015.

This contract with TVN (started in Oct. 2012) concerns the PhD of Nicolas Dhollande and aims at performing a coding mode analysis and developing a pre-analysis software. HEVC standard is a new standard of compression including new tools such as advanced prediction modes. Compared to the previous standard H.264, HEVC's complexity is three to four times higher. The goal of this thesis is to infer the best coding decisions (prediction modes...) in order to reduce the computational complexity of HEVC thanks to a pre-analysis step. The pre-analysis is expected to provide useful estimates of local video characteristics which will then help selecting the prediction and transform partitions as well as a number of other parameters such as the quantization parameters or the prediction modes.

7.2.7. CIFRE contract with Envivio on LDR compatible HDR video coding

Participants: Christine Guillemot, David Gommelet, Aline Roumy.

- Title : LDR-compatible coding of HDR video signals.
- Research axis : § 6.3.3.
- 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. Partnerships and Cooperations

8.1. European Initiatives

8.1.1. FP7 & H2020 projects

8.1.1.1. FP7-PEOPLE-SHIVPRO

Participants: Olivier Le Meur, Zhi Liu.

- Title : Saliency-aware High-resolution Video Processing.
- Research axis : 6.1.1.
- Partners : Visting professor from Shanghai University.
- Funding : EC-FP7 MC-IIF International Incoming Fellowships (IIF).
- Period : 08/2012-07/2014

The SHIVPRO project has been supporting the visit of Dr. Z. Liu, from Beijing University in the team from August 2012 to August 2014. The objective of this project was to propose an efficient spatio-temporal saliency model to predict salient regions in High-Resolution (HR) videos, and fully exploit it to ease the design and improve the performance of HR video compression and retargeting applications. With the aim to overcome the drawbacks of existing saliency models, based on a multiscale region representation, the proposed model systematically realizes statistical model saliency measuring, intra-scale saliency modification, inter-scale saliency propagation and flexible incorporation of top-down information, to generate a novel saliency representation form with scalability, saliency tree, from which a multiscale saliency fusion scheme is used to derive high-quality saliency maps at various scales.

8.2. International Research Visitors

8.2.1. Visits of International Scientists

Dr. Zhi Liu, from Shanghai University, has been visiting the team from August 2012 until August 2014. His stay has been funded by the FP7-PEOPLE-2011-IIF program. The funding scheme is the MC-IIF International Incoming Fellowships (IIF).

8.2.2. Visits to International Teams

8.2.2.1. Sabbatical programme

C. Guillemot has spent a six months sabbatical stay (Mar. 2014- Aug. 2014) at EPFL (Ecole Polytechnique Federale de Lausanne)

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific events organization

9.1.1.1. Member of the organizing committee

- C. Guillemot has been keynote chair of IEEE-ICIP 2014 and area chair of Eusipco 2014.
- C. Guillemot is member of the IEEE IVMSP technical committee.
- C. Guillemot is member of the award committee of the Eurasip Image communication journal.
- C. Guillemot is member of the committee in charge of the IEEE Brillouin-Glavieux award.
- C. Labit is member of the GRETSI association board.
- C. Labit has been president of the organization committee of 2013 Gretsi-EEA Signal-Image-Vision thesis prize
- A. Roumy is member of the GRETSI association board.
- A. Roumy is local liaison officer for Eurasip.

9.1.2. Scientific events selection

9.1.2.1. Member of the conference programme committee

- C. Guillemot has been a member of technical program committees of international conferences of the field (EUSIPCO 2014, IEEE-ICIP 2014, EUVIP-2014).
- O. Le Meur has been a member of the technical program committees of the international conferences DICTA 2014, QoMex 2014, EUVIP 2014, ICME 2014.
- O. Le Meur co-organized a special session on Visual saliency: emerging models and applications in multimedia processing? at ICME 2014.

9.1.3. Journal

- C. Guillemot is associate editor of the Eurasip International Journal on Image Communication.
- C. Guillemot is associate editor of the IEEE Trans. on Image Processing.
- C. Guillemot is associate editor of the International Journal on Mathematical Imaging and Vision.
- C. Guillemot is senior member of the editorial board of the IEEE Journal on selected topics in signal processing.

9.1.4. Scientific animation

- C. Guillemot is member of the Selection and Evaluation Committee of the “Pôle de Compétitivité” Images and Networks of the Region of Ouest of France.
- C. Guillemot is member as scientific expert of the CCRRDT (Regional Committee of Research and Technological Development) of the Brittany region.
- C. Guillemot is member of the “bureau du Comité des Projets” as well as of the “commission personnels” in charge of the postdoc and delegation recrutments.
- C. Labit is the Vice-president of the Scientific Board, in charge of Research and Innovation, for the University of Rennes1 (since June 1st, 2008)
- C. Labit is president of the Rennes-Atalante Science Park and of the start-up incubator Emergys (since April, 2007).
- C. Labit is member of the ICT strategic steering committee (CPS-7) of the National Research Agency (ANR).

- A. Roumy is a member of the Selection and Evaluation Committee (CDT Commission Developpement Technologique) for the Inria technological development grants.
- A. Roumy is titular member of the National Council of Universities (CNU section 61, 2012-2015).

9.2. Patents and Standardization

9.2.1. Patents

- Four patents have been filed jointly by Technicolor and Inria in the area of video analysis and compression.
- One patent has been filed jointly by Thomson Video Networks, Univ. Rennes 1 and Inria in the area of UHD HEVC-based video coding.

9.3. Teaching - Supervision - Juries

9.3.1. 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, 13 hours, M2, Univ. of Paris 8, France.

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

Master: A. Roumy, Magistère program, Information Theory, Computer science and telecommunications, 18 hours, Ecole Normale Supérieure de Cachan, Ker Lann campus, France.

Engineering degree: C. Guillemot, Video communication, 12 hours, Télécom Lille 1, Villeneuve-d'Ascq, France.

L1: L. Guillo, Programming, Univ. of Rennes 1, France.

Engineering degree: C. Labit, Entrepreneurship and innovation, 3 hours, ESIR, Rennes, 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.

Engineering degree: A. Roumy, Image processing and numerical analysis, 51 hours, ECAM Rennes, France.

9.3.2. Invited Talks

- O. Le Meur gave a talk at UESTC (University of Electronic Science and Technology of China), School of Electronic Engineering, Chengdu, China. Prof. Bing ZENG and Prof. Hongliang LI. Title of the talk: examplar-based inpainting methods, July 2014.
- O. Le Meur gave a talk at Sichuan University, Chengdu, China. Prof. Qionghua WANG. Title of the talk: examplar-based inpainting methods, July 2014.
- O. Le Meur gave a talk at the National Research Group (GdR), Information Signal Image viSion (ISIS). Special session on visual attention. Title of the talk: Modelling the visual scanpath, 19th June 2014.

9.3.3. Supervision

PhD : M. Bevilacqua, Image and video super-resolution using neighbor embedding algorithms, Univ. of Rennes 1, defense on the 4th of June 2014, A. Roumy and C. Guillemot (contract with Alcatel/Lucent)

PhD : S. Cherigui, Techniques de codage d'images basees representations parcimonieuses de scenes et prediction spatiale multi-patches, University of Rennes 1, defense on the 18th of June 2014, C. Guillemot (contract with Technicolor)

PhD : M. Ebdelli, Video inpainting techniques : application to object removal and error concealment, Univ. of Rennes 1, defense on the 20th of June 2014, C. Guillemot and O. Le Meur.

PhD in progress : J. Aghaei Mazaheri, Sparse representations and dictionary learning for satellite image compression, Oct. 2011-Sept. 2014, C. Guillemot and C. Labit (contract with Astrium)

PhD in progress : M. Bevilacqua, Image and video super-resolution using neighbor embedding algorithms, Feb. 2011-Jan. 2014, A. Roumy and C. Guillemot (contract with Alcatel/Lucent)

PhD in progress : B. Huang, Video compression with generalized lifting, Apr. 2012-March 2015, C. Guillemot (Cifre contract with Orange)

PhD in progress : D. Khaustova, Objective evaluation of 3D video quality, Dec. 2011-Nov. 2014, O. Le Meur (Cifre contract with Orange)

PhD in progress : M. Alain, Spatio-temporal linear embedding for epitome-based video compression, Oct. 2012-Sept. 2015, C. Guillemot (Cifre contract with Technicolor)

PhD in progress : N. Dhollande, HEVC codec optimization based on content pre-analysis, Oct. 2012-Sept. 2015, O. Le Meur and C. Guillemot (Cifre contract with Thomson Video Networks)

PhD in progress : M. Le Pendu, HDR video compression, Dec. 2012-Nov. 2015, C. Guillemot (Cifre contract with Technicolor)

PhD in progress : Julio-Cesar Ferreira, multi-view super-resolution, Oct. 2013-March 2015, C. Guillemot, G. Carrijo, E. Da Silva (Co-tutelle with University of Uberlandia, Brazil).

PhD in progress: D. Gommelet, LDR-compatible HDR video compression, Oct. 2014- Sept. 2017, C. Guillemot and A. Roumy (Cifre contract with Envivio).

9.3.4. Juries

- C. Guillemot has been member (rapporteur) of the jury of the PhD committee of:
  - J.J. Micallef, Univ. of Malta, Jan. 2014
  - E. Mora, Telecom Paris Tech, Feb. 2014
  - Tuan Tran Thai, ISAE/Univ. of Toulouse, June 2014
- C. Guillemot has been member (president) of the jury of the PhD committee of:
  - E. Vidal, Univ. of valenciennes, Apr. 2014
  - W. Lu, INSA of Rennes, July 2014
- O. Le Meur has been member (as examiner) of the jury of the PhD committee of:
  - Hamed Rezazadegan Tavakoli, University of Oulu, Finland, 2014.
- A. Roumy has been member (as examiner) of the jury of the PhD committee of:
  - A. UNSAL, Eurecom Telecom ParisTech, Nov. 2014.

9.4. Popularization

- A. Roumy participated in the in-service training for teacher in Computer Science (Informatique et Science du Numérique (ISN), level bac -1 to bac+2) and gave a 3h course on video and image compression.
- As part of a seminar dedicated to teachers of secondary and preparatory schools, Laurent Guillo presented a possible use of Agility methods for pupils' projects and he animated the related workshop, ENS Rennes, October 2014.

10. Bibliography

Major publications by the team in recent years

[1] V. CHAPPELIER, C. GUILLEMOT. *Oriented wavelet transform for image compression and denoising*, in "IEEE Transactions on Image Processing", 2006, vol. 15, n^o 10, pp. 2892-2903, http://hal.inria.fr/inria-00504227

[2] T. COLLEU, S. PATEUX, L. MORIN, C. LABIT. *A polygon soup representation for multiview coding*, in "Journal of Visual Communication and Image Representation", Feb 2010, pp. 1–32

[3] J.-J. FUCHS. *A robust matched detector*, in "IEEE Trans. on Signal Processing", Nov. 2007, vol. 55, n^o 11, pp. 5133-5142

[4] C. GUILLEMOT, A. ROUMY. *Towards constructive Slepian-Wolf coding schemes*, in "Distributed source coding", Elsevier Inc., 2008

[5] T. GUIONNET, C. GUILLEMOT. *Soft decoding and synchronization of arithmetic codes: Application to image transmission over noisy channels*, in "IEEE Trans. on Image Processing", Dec. 2003, vol. 12, pp. 1599-1609

[6] H. JÉGOU, C. GUILLEMOT. *Robust multiplexed codes for compression of heterogeneous data*, in "IEEE Transactions on Information Theory", April 2005, vol. 51, n^o 4, pp. 1393 - 1407, http://hal.inria.fr/inria-00604036

[7] O. LE MEUR, P. LE CALLET, D. BARBA. *Predicting visual fixations on video based on low-level visual features*, in "Vision Research", Sep. 2007, vol. 47, n^o 19, pp. 2493-2498

[8] O. LE MEUR, P. LE CALLET, D. BARBA, D. THOREAU. *A coherent computational approach to model the bottom-up visual attention*, in "IEEE Trans. On PAMI", May 2006, vol. 28, n^o 5, pp. 802-817

[9] L. OISEL, E. MEMIN, L. MORIN, F. GALPIN. *One-dimensional dense disparity estimation for three-dimensional reconstruction*, in "IEEE Trans. on Image Processing", Sept. 2003, vol. 12, n^o 9, pp. 1107–1119

[10] A. ROUMY, S. GUEMGHAR, G. CAIRE, S. VERDU. *Design Methods for Irregular Repeat-Accumulate Codes*, in "IEEE Trans. on Information Theory", August 2004, vol. 50, n^o 8

Publications of the year

Doctoral Dissertations and Habilitation Theses

[11] M. BEVILACQUA. *Algorithms for super-resolution of images and videos based on learning methods*, Université Rennes 1, June 2014, https://tel.archives-ouvertes.fr/tel-01064396

[12] S. CHÉRIGUI. *Sparse representation-based video coding and multi-patches spatial prediction*, Université de Rennes 1, June 2014, https://hal.inria.fr/tel-01093920

[13] M. EBDELLI. *Video inpainting techniques : application to object removal and error concealment* , Université de Rennes 1, June 2014, https://hal.inria.fr/tel-01093202

[14] O. LE MEUR. *Visual attention modelling and applications. Towards perceptual-based editing methods*, University of Rennes 1, November 2014, Habilitation à diriger des recherches, https://hal.inria.fr/tel-01085936

Articles in International Peer-Reviewed Journals

[15] C. BAI, J. ZHANG, Z. LIU, W.-L. ZHAO. *K-means based histogram using multiresolution feature vectors for color texture database retrieval*, in "Multimedia Tools and Applications", May 2014 [*DOI :* 10.1007/s11042-014-2053-8], https://hal.inria.fr/hal-00997873

[16] M. BEVILACQUA, A. ROUMY, C. GUILLEMOT, A. MOREL. *Single-Image Super-Resolution via Linear Mapping of Interpolated Self-Examples*, in "IEEE Transactions on Image Processing", October 2014, vol. 23, n^o 12, pp. 5334-5347 [*DOI :* 10.1109/TIP.2014.2364116], https://hal.archives-ouvertes.fr/hal-01088753

[17] G. COLUCCIA, A. ROUMY, E. MAGLI. *Operational Rate-Distortion Performance of Single-source and Distributed Compressed Sensing*, in "Communications, IEEE Transactions on", June 2014, vol. 62, n^o 6, pp. 2022 - 2033, IEEE Transactions on Communications [*DOI :* 10.1109/TCOMM.2014.2316176], https://hal.inria.fr/hal-00996698

[18] E. DUPRAZ, A. ROUMY, M. KIEFFER. *Source Coding with Side Information at the Decoder and Uncertain Knowledge of the Correlation*, in "IEEE Transactions on Communications", January 2014, vol. 62, n^o 1, pp. 269 - 279 [*DOI :* 10.1109/TCOMM.2013.120413.130102], https://hal-supelec.archives-ouvertes.fr/hal-00935847

[19] R. GUERRAOUI, K. HUGUENIN, A.-M. KERMARREC, M. MONOD, S. PRUSTY, A. ROUMY. *Tracking Freeriders in Gossip-Based Content Dissemination Systems*, in "Computer Networks", May 2014, vol. 64, pp. 322-338 [*DOI :* 10.1016/J.COMNET.2014.02.023], https://hal.inria.fr/hal-00941107

[20] C. GUILLEMOT, O. LE MEUR. *Image inpainting: Overview and recent advances*, in "IEEE Signal Processing Magazine", January 2014, https://hal.inria.fr/hal-00876076

[21] Z. LIU, X. ZHANG, S. LUO, O. LE MEUR. *Superpixel-based spatiotemporal saliency detection*, in "IEEE Transactions on Circuits and Systems for Video Technology", February 2014, 1 p. [*DOI :* 10.1109/TCSVT.2014.2308642], https://hal.archives-ouvertes.fr/hal-00993218

[22] Z. LIU, W. ZOU, O. LE MEUR. *Saliency Tree: A Novel Saliency Detection Framework*, in "IEEE Transactions on Image Processing", May 2014, vol. 23, n^o 5, pp. 1937-1952 [*DOI :* 10.1109/TIP.2014.2307434], https://hal.archives-ouvertes.fr/hal-00993215

[23] Z. LIU, W. ZOU, L. LI, L. SHEN, O. LE MEUR. *Co-saliency detection based on hierarchical segmentation*, in "IEEE Signal Processing Letters", January 2014, vol. 21, n^o 1, pp. 88-92 [*DOI :* 10.1109/LSP.2013.2292873], https://hal.archives-ouvertes.fr/hal-00993210

[24] L. SHEN, Z. ZHANG, Z. LIU. *Adaptive inter-mode decision for HEVC jointly utilizing inter-level and spatio-temporal correlations*, in "IEEE Transactions on Circuits and Systems for Video Technology", March 2014, 1 p. [*DOI :* 10.1109/TCSVT.2014.2313892], https://hal.archives-ouvertes.fr/hal-00996798

International Conferences with Proceedings

[25] M. ALAIN, C. GUILLEMOT, D. THOREAU, P. GUILLOTEL. *Clustering-based Methods for Fast Epitome Generation*, in "European Signal Processing Conference (EUSIPCO 2014)", Lisbonne, Portugal, September 2014, https://hal.inria.fr/hal-01008279

[26] S. CHÉRIGUI, M. ALAIN, C. GUILLEMOT, D. THOREAU, P. GUILLOTEL. *Epitome inpainting with in-loop residue coding for image compression*, in "ICIP - IEEE International Conference on Image Processing 2014", Paris, France, October 2014, https://hal.inria.fr/hal-00994418

[27] G. COLUCCIA, A. ROUMY, E. MAGLI. *Exact Performance Analysis of the Oracle Receiver for Compressed Sensing Reconstruction*, in "IEEE International Conference on Acoustics,

Speech, and Signal Processing (ICASSP)", Florence, Italy, May 2014, ICASSP 2014 proceedings [*DOI :* 10.1109/ICASSP.2014.6853748], https://hal.inria.fr/hal-00996695

[28] N. DHOLLANDE, O. LE MEUR, C. GUILLEMOT. *HEVC intra coding of Ultra HD video with reduced complexity*, in "IEEE International Conference on Image Processing 2014 (ICIP 2014)", Paris, France, France, October 2014, https://hal.inria.fr/hal-00994476

[29] J. C. FERREIRA, O. LE MEUR, C. GUILLEMOT, E. A. B. DA SILVA, G. A. CARRIJO. *Single image super-resolution using sparse representations with structure constraints*, in "IEEE International Conference on Image Processing", Paris, France, October 2014, https://hal.inria.fr/hal-00995052

[30] D. KHAUSTOVA, J. FOURNIER, O. LE MEUR, E. WYCKENS. *How visual attention is modified by disparities and textures changes?*, in "SPIE 8651, Human Vision and Electronic Imaging XVIII", San Francisco, United States, February 2014, 865115 p. , https://hal.inria.fr/hal-00993791

[31] D. KHAUSTOVA, O. LE MEUR, J. FOURNIER, E. WYCKENS. *An investigation of visual selection priority of objects with texture and crossed and uncrossed disparities*, in "SPIE 8651, Human Vision and Electronic Imaging XVIII", San Francisco, United States, February 2014, vol. 9014, pp. 90140D-90140D-13, https://hal.inria.fr/hal-00993786

[32] O. LE MEUR, Z. LIU. *Saliency aggregation: Does unity make strength?*, in "12th Asian Conference on Computer Vision (ACCV 2014)", Singapor, Singapore, November 2014, https://hal.inria.fr/hal-01085898

[33] M. LE PENDU, C. GUILLEMOT, D. THOREAU. *Adaptive Re-Quantization For High Dynamic Range Video Compression*, in "ICASSP - IEEE International Conference on Acoustics Speech and Signal Processing - 2014", Florence, Italy, May 2014, https://hal.inria.fr/hal-00993167

[34] M. LE PENDU, C. GUILLEMOT, D. THOREAU. *Rate Distortion Optimized Tone Curve for High Dynamic Range Compression*, in "EUSIPCO - 22nd European Signal Processing Conference", Lisbon, Portugal, September 2014, https://hal.inria.fr/hal-00998480

[35] L. LI, Z. LIU, W. ZOU, X. ZHANG, O. LE MEUR. *Co-saliency detection based on region-level fusion and pixel-level refinement*, in "International Conference on Multimedia Expo (ICME)", Chengdu, China, July 2014, pp. 1-6, https://hal.archives-ouvertes.fr/hal-00996799

[36] L. ZHANG, Y. CHEN, V. THIRUMALAI, J.-L. LIN, Y. CHEN, J. AN, S. LEI, T. GUIONNET, L. GUILLO, C. GUILLEMOT. *Inter-view motion prediction in 3D-HEVC*, in "ISCAS", Melbourne, Australia, June 2014, https://hal.inria.fr/hal-00992942