IN PARTNERSHIP WITH:
CNRS

Université Rennes 1

Activity Report 2013

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, HdR]
Aline Roumy [Inria, Researcher]

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

Engineers
Laurent Guillo [CNRS, Technical staff, 20%]
Alan Bourasseau [Univ. Rennes I, until Jul 2013]
David Gommelet [Univ. Rennes I, from Nov 2013]
Ronan Le Boulch [Inria, until Nov 2013]

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.]
Nicolas Dhollande [Univ. Rennes I, granted by CIFRE with Thomson Video Networks]
Mounira Ebdelli [Inria, granted by ANR-ARSSO project]
Julio Cesar Ferreira [co-tutelle Univ. Rennes I - Univ. Uberlandia Brazil, since Oct. 2013]
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, from Jan 2013]

Post-Doctoral Fellow
Raul Martinez Noriega [Inria, until May 2013]

Visiting Scientist
Zhi Liu [Univ. Rennes I]

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, the objective is then to adapt the data coding, protection and transmission scheme. Classical protection methods use Forward Error Correction (FEC). The code rate is then adapted to the visual data priority. However, the reliability of the estimated PER, impacts the performance of FEC schemes. This is the first problem we propose to investigate focusing on the problem of constructing codes which, together with a scalable source representation, 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 the *cliff* effect from which suffer classical FEC schemes when the loss rate exceeds the error correction capacity of the code. The followed direction is based on Wyner-Ziv coding, used as a tool for lossy systematic error correction. The other problem addressed concerns 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. Classical approaches are based on spatial and/or spatio-temporal interpolation. We investigate new methods relying on video models (based on sparsity, epitomes, ...).

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.

2.6. Highlights of the Year

- A joint contribution between Inria/SIROCCO, Qualcomm and Mediatek has been adopted to be part of the HEVC backward compatible 3DV standard, in July 2013.
- Nomination of C. Guillemot as IEEE Fellow.

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 the 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 $\min\{\|\mathbf{x}\|_0 \;:\; \|A\mathbf{x}-\mathbf{b}\|_p \le \rho\}$, for some $\rho \ge 0$, 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 principal component analysis (PCA) and independent component analysis (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 unary 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

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

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 form 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:

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

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

[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.

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 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. The continuous increase of access network bandwidth leads to increasing numbers of networked digital content users and consumers which in turn triggers needs for higher core bandwidth and higher compression efficiencies. 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, its robustness to transmission impairments, are thus important features for seamless adaptation to varying network conditions and to terminal capabilities.

In medical imaging, the large increase of medical analysis using various image sources for clinical purposes and the necessity to transmit or store these image data with improved performances related to transmission delay or storage capacities, command to develop new coding algorithms with lossless compression algorithms or *almost* lossless compression characteristics with respect to the medical diagnosis.

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. Nethertheless, 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, 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. 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 an 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. 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).

5.4. loss concealment algorithm using examplar-based video inpainting

Participants: Ronan Le Boulch, Mounira Ebdelli, Christine Guillemot, Olivier Le Meur [contact person].

This software recovers regions of a video sequence which can be lost after transmission over a network with no guarantee of quality of service. Motion information of impaired areas is first interpolated from the motion vectors of known areas. An examplar-based video inpainting method is then used to fill in the corrupted areas. This software is being registered at the APP (Agence de Protection des Programmes).

5.5. Standardization

Participants: Christine Guillemot, Laurent Guillo [contact person].

In the continuity of the ADT Picovin-P, we have in 2013, pursued our activities of standardization in the area of multi-view plus depth video coding. We in particular followed the standardization activities within the Joint Collaborative Team on 3D Video Coding Extension (JCT-3V). JCT-3V aims at developing 3D extensions for video codecs, which are AVC (ATM) or HEVC (HTM) based. We have pursued the developments of our proposal related to inter-view motion vector prediction, leading to a joint proposal with Qualcomm and Mediatek which has been adopted in the standard in July 2013.

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, a new method to detect salient objects has been proposed [32], [18]. The principle relies upon low-level visual features and super-pixel segmentation. First, the original image is simplified by performing super-pixel segmentation and adaptive color quantization. On the basis of super-pixel representation, inter-super-pixel similarity measures are then calculated based on difference of histograms and spatial distance between each pair of super-pixels. For each super-pixel, its global contrast measure and spatial sparsity measure are evaluated, and refined with the integration of inter super-pixel similarity measures to finally generate the super-pixel-level saliency map. Experimental results on a dataset containing 1,000 test images with ground truths demonstrate that the proposed saliency model outperforms state-of-the-art saliency models. Figure 1 illustrates some results.

Figure 1. Illustration of the proposed approach: first row: original image; second row: saliency map; third row: extraction of the salient object.

6.1.2. Image Memorability

Participant: Olivier Le Meur.

This work has been carried out in collaboration with Mattei Mancas (researcher of the University of Mons) during his visit of the team. The image memorability consists in the faculty of an image to be recalled after a period of time. Recently, the memorability of an image database was measured and some factors responsible for this memorability were highlighted. In [34] we proposed to improve an existing method by using attention-based visual features. To determine whether the visual attention plays a role in the memorability mechanism, eye tracking experiment has been performed by using a set of images of different memorability scores. Two important results have been observed. First the fixation duration is longer for the most memorable images (especially for the very first fixations) which shows a higher cognitive activity for memorable images. Second the observers congruency (agreement between observers) is significantly higher for the most memorable images. This shows that when there are areas with high attraction on all viewers, this induces higher memorability.

Following these first two observations, attention-based visual features were used to predict image memorability scores. A new set of features was then defined and used to train a model. Compared to an existing approach, we improve on the quality of the prediction of 2% while reducing the number of parameters by 14%. More specifically we replace the 512 features related to the GIST by 17 features which are directly related to visual attention.

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. The goal is also to study the differences in 2D visual attention in comparison with 3D visual attention.

Several subjective experiments have been carried out in order to study visual attention in different viewing conditions. The goal of the first experiment, involving 135 observers, was to study visual attention in three different conditions (2D, 3D comfortable and 3D uncomfortable), to eventually establish whether depth influences visual attention and whether there is a link between comfort and visual attention. The use of an eye-tracker allowed to record and to track observer's gaze. By analyzing the results, we found out that visual strategy to observe 2D images and 3D images with uncrossed disparity is very similar; there was no significant influence of discomfort on visual attention.

The second question which has then been addressed is how visual attention is influenced by objects with crossed disparity. A second test has been designed to answer this question, involving 51 observers. Considering scenes with crossed disparity it was revealed that objects located in front of the display plane are the most salient, even if observers experience discomfort. In the third experiment, we extended the study using scenes with crossed and uncrossed disparities. We verified the hypothesis that texture and contrast are more influential in guiding our gaze than the amount of depth. The features influencing the saliency of the objects in stereoscopic conditions were also evaluated with low-level visual stimuli. It was discovered that texture is the most salient feature in comparison to depth. Crossed disparity significantly influences the process of selecting the objects, while uncrossed disparity is less important, the process of selection being in this latter case similar to 2D conditions.

6.1.4. Epitome-based video representation

Participants: Martin Alain, Christine Guillemot.

This work is carried out in collaboration with Technicolor (D. Thoreau, Ph. Guillotel) and aims at studying novel spatio-temporal representations for videos based on epitomes. An epitome is a condensed representation of an image (or a video) signal containing the essence of the textural properties of this image. Different forms of epitomes have been proposed in the literature, such as a patch-based probability model learned either from still image patches or from space-time texture cubes taken from the input video. These probability models together with appropriate inference algorithms, are useful for content analysis inpainting or super-resolution. Another family of approaches makes use of computer vision techniques, like the KLT tracking algorithm, in order to recover self similarities within and across images. In parallel, another type of approach consists in extracting epitome-like signatures from images using sparse coding and dictionary learning.

We have in the past (in the context of the PhD thesis of S. Cherigui) developed a method for constructing epitomes for representing still images. The algorithm tracks self-similarities within the image using a block matching (BM) algorithm. The epitome is constructed from disjoint pieces of texture ("epitome charts") taken from the original image and a transform map which contains translational parameters (see Fig.2. Those parameters keep track of the correspondences between each block of the input image and a block of the epitome. An Intra image compression scheme based on the epitome has been developed showing significant rate savings on some images, including the rate cost of the epitome texture and of the transform map. The entire image can be reconstructed from the epitome texture with the help of the transform map. The method is currently being extended to construct epitome representations of video segments rather than simple images. Such spatio-temporal epitome should pave the way for novel video coding architectures and open perspectives for other video processing problems which we have started to address such as denoising and super-resolution.

Figure 2. Original image and corresponding epitome.

6.2. Rendering, inpainting and super-resolution

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

6.2.1. Image and 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.

Novel methods have been developed investigating two complementary directions first for image inpainting. The first direction which has been explored is the estimation of the unknown pixel with different neighbor embedding methods, i.e. Locally Linear embedding (LLE), LLE with a low-dimensional neigborhood representation (LLE-LDNR), Non-Negative Matrix Factorization (NMF) with various solvers [16]. The second method developed uses a two-steps hierarchical (or coarse to fine) approach to reduce the execution time [17]. In this hierarchical approach, a low resolution version of the input image is first inpainted, this first step being followed by a second one which recovers the high frequency details of the inpainted regions, using a single-image super-resolution method. To be less sensitive to the parameters setting of the inpainting, the low-resolution input picture is inpainted several times with different settings. Results are then efficiently combined with a loopy belief propagation. Experimental results in a context of image editing, texture synthesis and 3D view synthesis demonstrate the effectiveness of the proposed method.

The problem of video inpainting has also been considered. A first video inpainting algorithm has been developed in 2012, using a spatio-temporal examplar-based method. The algorithm proceeds in three steps. The first one inpaints missing pixels in moving objects using motion information. Then the static background is inpainted exploiting similarity between neighboring frames. The last step fills in the remaining holes in the current frame using spatial inpainting. This approach works well with static cameras but not so well when the video has been captured by free-moving cameras.

In 2013, we have therefore addressed the problem of video inpainting 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. The proposed approach, although less complex than state-of-the-art methods, provides more natural results (see Fig.3).

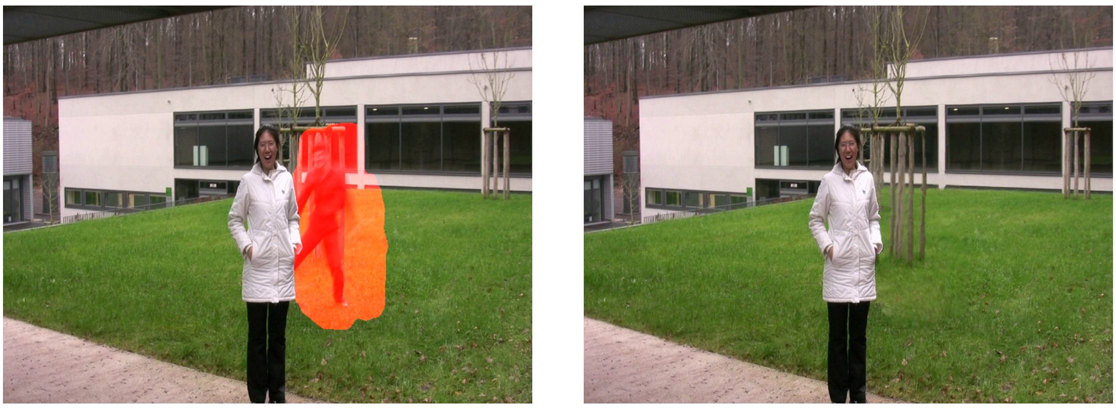

Figure 3. Mask of the image to be inpainted; Results with the proposed video inpainting algorithm.

6.2.2. Image priors for inpainting

Participants: Raul Martinez Noriega, Aline Roumy.

Image inpainting is an ill-posed inverse problem which has no well-defined unique solution. To make this problem more "well-defined" it is necessay to introduce image priors. We consider here the problem of extracting such priors to help restoring the connection of long edges across the missing region. The prior is defined as a binary image that contains the locations of salient edge points located at the boundary of the missing region as well as the linear edges that join these points across the missing region. A method has been developed to extract such priors. It first detect edges which are then successively pruned in order to keep only informative edges, i.e., which have coherent gradients and are either part of a salient structure, or at the border between two different textures. Edges which are quasi-perpendicular to the boundary of the missing region are finally retained. Directions of the retained edges are computed and pairs of edges with similar directions are then connected with straight lines. These lines are used to segment the image into different regions and to define the processing order of the patches to be inpainted. Only patches from the known part and belonging to the same region as the input patch are used. This avoids bringing details of one texture into another one, as well as the unconnected edge problem [35].

6.2.3. Image and video super-resolution

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 (see, for example, Fig. 4 for a comparison between bicubic interpolation and SR). We mostly focused on the single-image problem, where only a single LR image is available.

We have adopted the example-based framework, where 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 super-resolved 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 learnt 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.

To overcome this problem, in [23] we proposed a novel method to build a compact external dictionary. The method consists in first jointly clustering LR and HR patches. The aim of this procedure, which we called JKC (Jointly K-means Clustering), is to prune the dictionary of the “bad” pairs of patches, i.e. those ones for which the cluster assignments of the related LR and HR patches do not correspond. Once the dictionary is clustered, it is summarized, by sampling some prototype patches, and applying on them simple geometrical transformations, in order to enrich the dictionary. The so constructed compact dictionary is shown to give equivalent or even better performance than the initial large dictionary with any input image.

The dictionary construction method described in [23] has been used as a basis for designing a full single-image SR algorithm. The new algorithm, presented in [25], follows the traditional scheme of example-based SR with an external dictionary, where a new way to generate the training patches is introduced. Given a HR training image H, the corresponding LR image L is generated; but instead of directly sampling patches from H and L, as usually done, the training images are further processed. An enhanced interpolation of L, using an iterated back projection, is used as a source of LR patches, and a high-frequency residual image, given by the difference between H and the interpolated LR image, is used for extracting HR patches. The JKC procedure is then applied to get the final compact dictionary. A special example-based SR algorithm has been designed, where the final HR output patches are constructed by combining selected HR residual patches from the dictionary with nonnegative weights. In the context of this study, we have also introduced a novel non-negative dictionary learning method [24]. The proposed method consists of two steps which are alternatively iterated: a sparse coding and a dictionary update stage. As for the dictionary update, an original method has been proposed, which we called K-WEB, as it involves the computation of k WEighted Barycenters.

Besides SR for still images, a preliminary work on video sequences has been also conducted [26]. In particular, we have considered the case of a LR video sequence with periodic high-resolution (HR) key frames. Given this scenario, a specific SR procedure has been designed to upscale each intermediate frame, by using the internal dictionary constructed from the two neighbor key frames.

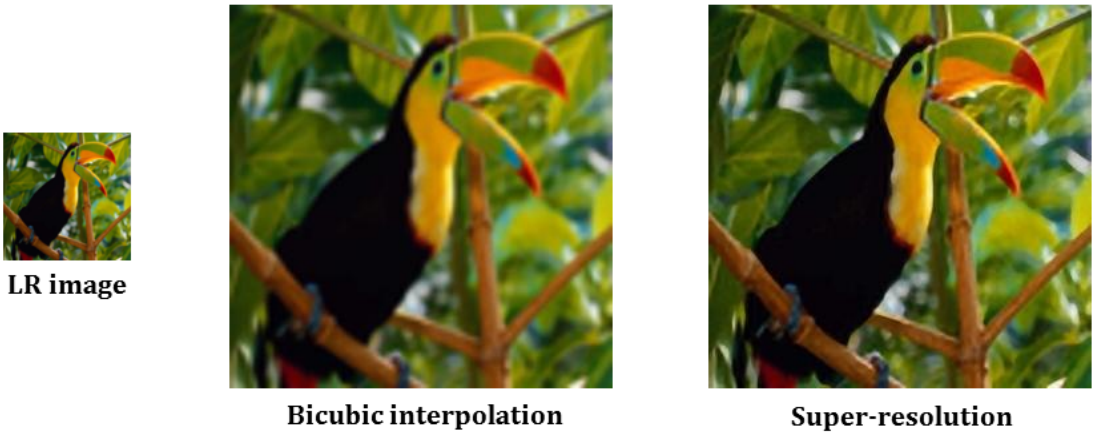

Figure 4. Comparison between bicubic interpolation and SR when upscaling the same LR image (by a factor of 3).

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. Multi-view plus depth video compression

Participants: Christine Guillemot, Laurent Guillo.

Multi-view plus depth video content represent very large volumes of input data wich need to be compressed for storage and tranmission to the rendering device. The huge amount of data contained in multi-view sequences indeed motivates the design of efficient representation and compression algorithms. The team has worked on motion vector prediction in the context of HEVC-compatible Multi-view plus depth (MVD) video compression. The HEVC compatible MVD compression solution implements a 6 candidate vector list for merge and skip modes. When a merge or a skip mode is selected, a merge index is written in the bitstream. This index is first binarized using a unary code, then encoded with the CABAC. A CABAC context is dedicated to the first bin of the unary coded index while the remaining bins are considered as equiprobable. This strategy is efficient as long as the candidate list is ordered by decreasing index occurrence probability. We have improved the construction of the candidate list by proposing two new candidates derived from disparity motion vectors in order to exploit inter-view correlation. This work has led to a joint proposal with Qualcomm and Mediatek which has been adopted in the HEVC-3DV standard in July 2013.

6.3.2. Spatio-temporal video prediction with neighbor embedding

Participants: Martin Alain, Christine Guillemot.

The problem of texture prediction can be regarded as a problem of texture synthesis. Given observations, or known samples in a spatial neighborhood, the goal is to estimate unknown samples of the block to be predicted. We have in 2012 developed texture prediction methods as well as inpainting algorithms using sparse representation as with learned dictionaries [19], or using neighbor embedding techniques [11], [30]. The methods which we have more particularly considered are Locally Linear Embedding (LLE), LLE with Low-dimensional neighborhood representation (LDNR), and Non-negative Matrix Factorization (NMF) using various solvers. In 2013, we have addressed the problem of temporal prediction for inter frame coding of video sequences using locally linear embedding (LLE). LLE-based prediction computes the predictor as a linear combination of K nearest neighbors (K-NN) searched within one or several reference frames. We have explored different K-NN search strategies in the context of temporal prediction, leading to several temporal predictor variants using or not motion information [22]. A parallel was also drawn between such multi-patch based prediction and the adaptive interpolation filtering (AIF) method. The LLE-based inter prediction techniques, when used as extra modes for inter prediction in an H.264 codec, are shown to bring significant Rate-Distortion (RD) performance gains compared to H.264 (up to 21.76 $\%$ bit-rate saving) and with respect to the use of AIF.

6.3.3. Dictionary learning for sparse coding 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, therefore the rate without compression is very high (about 70 Gbits/sec). The goal is to achieve a rate after compression of 600 Mbits/sec, i.e., 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.

Sparse representation of a signal consists in representing a signal $y \in \Re^n$ as a linear combination of columns, known as atoms, from a dictionary matrix. The dictionary $D \in \Re^{n \times K}$ is generally overcomplete and contains K atoms. The approximation of the signal can thus be written $y \approx Dx$ and is sparse because a small number of atoms of D are used in the representation, meaning that the vector x has only a few non-zero coefficients. Sparsity of the representation depends on how the dictionary is representative of the data at hand, hence the need to learn appropriate dictionaries.

We have developed methods for learning adaptive tree-structured dictionaries, called Tree K-SVD [20]. 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. We recently developped a new sparse coding method adapted to this tree-structure to improve the results [20]. The tree-structured dictionary has been further improved by studying different branch pruning strategies. The use of these dictionaries in an HEVC-based intra coder is under study. The dictionaries are also considered for scene classification and for detecting the MTF (Modulation Transfer Function) of the optical capturing system.

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 are generally used to represent the pixel data. Floating point video formats are widely used in the visual effects industry. Moreover, the development of a new standardized workflow ACES intends to generalize the use of such formats to the whole cinema production pipeline. The increasing use of floating point representations, however, comes with a technical issue concerning the storage space required for those videos with higher precision than the current 8 or 10 bit standards.

In collaboration with Technicolor (D. Thoreau), we worked on floating point video compression. Different approaches exist in the literature. Several methods consists in compressing directly the floating point data using its internal representation (i.e. sign, exponent and mantissa bits). These methods are generally limited to lossless compression schemes. Another type of approach makes use of the existing compression standards such as H264/AVC or HEVC to encode a floating point sequence of images previously converted to lower bit depth integers. In this approach, the conversion is designed to be reversible with minimal loss. However the converted integer images are not intended for being displayed directly. Finally a last family of approach aims at keeping backward compatibility with an existing compression standard. The original image sequence is first tone mapped and encoded to obtain a low dynamic range (LDR) version that can be visualized on a standard LDR display. In parallel, a residual information needed to reconstruct the HDR image from the LDR version is also encoded.

In our study, a floating point to integer conversion method was developed to be applied before HEVC compression. The original floating point RGB values are converted to high bit depth integers with an approximate logarithmic encoding that is reversible without loss. The RGB values are then converted to a YUV color space. The bit depth must also be reduced to be supported by the compression standard. This bit depth reduction is performed adaptively depending on the minimum and maximum values (i.e. darkest and brightest points respectively) which characterize the real dynamic of the data. In the best case, the difference between the extreme values is sufficiently low to perform this operation without loss.

Three variants of the method have been compared. The conversion can be performed either by Groups of Pictures (GOP), or independently on each frame of the sequence, or even more locally, by blocks of pixels. The GOP-wise approach combined with spatial and temporal predictions in the encoder gives the best results for low bit rate compression. The block-wise approach can reduce the bit depth with less data loss but breaks the continuity between the blocks, which degrades the Rate Distortion (RD) performance especially at low bit rates. However, we have shown that this approach gives the best results in the context of near lossless compression. The frame-wise version is intermediate between the global (GOP-wise) and local (block-wise) versions. It is adapted to high quality compression. This method was also compared to another frame-wise conversion method in the recent literature called adaptive LogLuv transform, and a 50% rate saving was obtained at high bitrates.

6.3.5. HEVC coding optimization

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

The team has two collaborations in the area of HEVC-based video coding optimization. The first research activity is carried out in collaboration with Orange labs (Felix Henry) and UPC (Philippe Salembier) in Barcelona. The objective is to design novel methods for predicting the residues resulting from spatio-temporal prediction. We have indeed observed that the redundancy in residual signals (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.

The second collaboration is with Thomson Video Networks and aims at designing an innovative architecture for effective real-time broadcast encoders of Ultra High Definition (UHD) contents. Currently, the only way to transmit acceptable UHD contents around $10-20$ Mbits/sec is the new compression standard HEVC (finalized in January 2013). Yet, UHD requires at minimum 8 times more computation than the actual HDTV formats, and HEVC has a computing complexity which is already from 2 to 10 times that of MPEG4-AVC.

To reduce the encoding complexity on UHD content, a pre-analysis with a lower resolution version (HD) of the input content has been considered to infer some decisions and coding parameters on the UHD video. A speed-up of a factor 3 has already been achieved for a small rate loss of $4-5\%$.

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. Loss concealment based on video inpainting

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

We have developed a loss concealment scheme based on a new hierarchical video examplar-based inpainting algorithm. The problem of loss concealment is to estimate unknown pixels after decoding when the corresponding transport packets have been lost on the transmission network. Before proceeding to the video texture inpainting, the motion vectors of the lost blocks must first be estimated from the motion vectors of the received blocks in the spatial neighborhood. The Motion vectors (MV) of damaged blocks are estimated using a Bilinear Motion Field Interpolation (BMFI) technique.

The algorithm follows a coarse to fine approach and first inpaints a low resolution version of the damaged video. Moving objects, detected thanks to the estimated motion vectors, are processed first. The most similar patches (similar to the known pixels of the patch to be completed) is searched within a motion-compensated window in adjacent frames, and used as an estimate of the pixels to be filled in. Then the static background is inpainted using known co-located pixels of neighboring frames. The remaining holes are filled-in using spatial inpainting.

In a second step, the high frequency details of the inpainted areas are recovered using a super-resolution technique, in the same vein as described in Section 6.2.1 for still images. The inpainted low resolution video is first interpolated using a simple lanczos interpolation. The idea is then to search for the nearest neighbor (the best match) of the interpolated version of each inpainted block, within the known part of the current image of the impaired video at the native resolution. The found correspondences form a so-called nearest neighbor field (NNF) which connects inpainted and interpolated patches of the low resolution video to high resolution patches of known parts of the high resolution (HR) video. The found NN patch is then copied to replace the low resolution inpainted patch.The two-step approach allows significantly reducing the execution time of the video inpainting process, while preserving a satisfactory quality.

6.4.2. Universal distributed 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 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 correlation between the successive frames is not directly used at the encoder, and are therefore seen as distributed. Traditional approaches (from an information theoretical but also practical point of view) assume that the correlation channel between 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 correlation channel is unknown and belongs to a class parametrized by some unknown parameter vector. We proposed four uncertainty models that depend on the partial knowledge we have on the correlation channel and derived the information theoretical bounds [28]. A complete coding scheme has also been proposed that works well for any distribution in the class [27]. At the encoder, the proposed scheme encompasses the determination of the coding rate and the design of the encoding process. Both contributions result from the information-theoretical compression bounds of universal lossless source coding with side information. Then a novel decoder is proposed that takes into account the available information regarding the class. The proposed scheme avoids the use of a feedback channel or the transmission of a learning sequence, which both would result in a rate increase at finite length.

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 : § 6.3.3.
- 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.

7.1.2. Collaboration with Alcatel on robust video compression

Participants: Marco Bevilacqua, Christine Guillemot, Ronan Le Boulch, Aline Roumy.

- Title: Self adaptive video codec
- Research axis: 6.2.3
- Funding: Joint research laboratory between Inria and Alcatel
- Period: Oct. 2010 - Dec. 2013.

In the framework of the joint research lab between Alcatel-Lucent and Inria, we participate in the ADR (action de recherche) Selfnets (or Self optimizing wireless networks). The objective is, jointly with the Alcatel Lucent team, to develop video representations and compression tools allowing smooth network adaptation on one hand and loss resilience on the other hand. In that context, the PhD thesis of M. Bevilacqua focuses on the development and study of image and video super-resolution as a tool for constructing scalable representations, hence enabling network adaptation of transmitted video streams. Single-image super-resolution algorithms have been developed, using different methods (neighbor embedding, local learning with regression), and dictionaries learned from external training images or learned on a multi-resolution pyramid constructed from the input low resolution image. These methods have been extended to video super-resolution, the dictionary being constructed from key frames.

7.1.3. 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-Feb.13.

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.

7.1.4. Contract with SHOM (Service Hydrographique et Océanographique de la Marine)

Participants: Alan Bourasseau, Olivier Le Meur.

- Title: Oceanograhic data compression
- Partners: SHOM, Alyotech, Univ. Rennes 1
- Funding: SHOM
- Period: 09/2012-02/2013.

The project consists in developing lossless and lossy compression algorithms for oceanographic data in partnership with ALYOTECH. The SIROCCO team contributes on the design and development of compression algorithms for this specific type of data, based on diffusion methods. The main constraint is the limited bandwidth used by the navy to transmit the data, i.e. an emitted message must be smaller than 4 kilo bytes. In 2013, the obtained quality versus rate performances has been assessed against those given by state of the art solutions (HEVC-Intra and JPEG-2000).

7.2. Grants with Industry

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

Participants: Christine Guillemot, Bihong Huang.

- Title : Generalized lifting for video compression.
- Research axis : § 6.3.5.
- 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.

7.2.2. 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.

7.2.3. 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.

7.2.4. 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.

7.2.5. 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 : § 6.3.5.
- 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.

8. Partnerships and Cooperations

8.1. National Initiatives

8.1.1. ANR-PERSEE

Participants: Christine Guillemot, Laurent Guillo, Olivier Le Meur.

- Title : Perceptual coding for 2D and 3D images.
- Research axis : § 6.2.1, 6.1.1.
- Partners : IRCCYN-Polytech Nantes, INSA-Rennes, Telecom Paris Tech.
- Funding : ANR.
- Period : Oct.2009-Aug.2013

The objective of the project is to develop perceptually driven coding solutions for mono-view and multi-view video. The SIROCCO project-team contributes on different problems for mono-view and multi-view video coding: visual attention modeling (see Section 6.1.1), texture synthesis and inpainting for both 2D and 3D content. Several methods for 2D image inpainting and 2D/3D inpainting to handle disocclusions in virtual view synthesis have been developed (see Sections 6.2.1. A computational model for 3D content has also been studied (see Section 6.1.1).

8.1.2. ANR-ARSSO

Participants: Mounira Ebdelli, Christine Guillemot, Ronan Le Boulch, Olivier Le Meur, Aline Roumy.

- Title : Adaptable, Robust, Streaming SOlutions.
- Research axis: 6.2.1, 6.4.1
- Partners : Inria/Planète, TESA-ISAE, CEA-LETI/LNCA, ALCATEL LUCENT BELL LABS, THALES Communications, EUTELSAT SA.
- Funding : ANR.
- Period : 06/2010-11/2013

The ARSSO project focuses on multimedia content communication systems, characterized by more or less strict real-time communication constraints, within highly heterogeneous networks, and toward terminals potentially heterogeneous too. It follows that the transmission quality can largely differ in time and space. The solutions considered by the ARSSO project must therefore integrate robustness and dynamic adaptation mechanisms to cope with these features. The overall goal is to provide new algorithms, develop new streaming solutions and study their performances. The SIROCCO project-team contributes on the development of loss concealment methods based on video inpainting. The solutions developed in 2012 have been studied in the context of a video compression and transmission chain using the emerging HEVC coding standard and have been integrated in the project demonstrator.

8.2. European Initiatives

8.2.1. 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 proposal SHIVPRO (Saliency-aware High-resolution Video Processing) submitted to the call FP7-PEOPLE-2011-IIF (funding scheme: MC-IIF International Incoming Fellowships (IIF)) has been accepted. Dr. Z. Liu, from Beijing University, has joined the team since August 2012 for two years. The objective of this project is 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.

8.3. International Research Visitors

8.3.1. Visits of International Scientists

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

9. Dissemination

9.1. Scientific Animation

- C. Guillemot is associate editor of the Eurasip International Journal on Image Communication.
- C. Guillemot is senior member of the editorial board of the IEEE Journal on selected topics in signal processing.
- C. Guillemot is member of the award committee of the Eurasip Image communication journal.
- 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 committee in charge of the IEEE Brillouin-Glavieux award.
- C. Guillemot has been a member of technical program committees of international conferences of the field (EUSIPCO 2013, IEEE-MMSP 2013, PV-2013 and PCS-2013).
- 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. Guillemot was a member of the selection committee for recruting a Professor at KTH, Stokholm(2013).

- 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 has been president of the organization committee of 2013 Gretsi-EEA Signal-Image-Vision thesis prize
- C. Labit is member of the GRETSI association board.
- C. Labit is member of the ICT strategic steering committee (CPS-7) of the National Research Agency (ANR).
- O. Le Meur co-organized a special session on "Visual attention and applications" at Wiamis 2013.
- A. Roumy is a member of the GRETSI association board.
- A. Roumy has been a member of the technical program committee of the international conference EUSIPCO 2013.
- A. Roumy is titular member of the National Council of Universities (CNU section 61, 2012-2015).
- A. Roumy is a local liaison officer for Eurasip.
- A. Roumy co-organized a special session on "Compressed Sensing and Sparse Representations for Visual Information Processing" at the Picture Coding Symposium (PCS), 2013.

9.2. Patents and Standardization

The team has contributed to the 3DV standardization initiatives of the ITU/MPEG joint collaborative team on video coding (JCT-VC) [40]. A joint proposal with Qualcomm and Mediatek has been adopted in the 3DV standard in July 2013 [42], [41].
Four patents have been filed jointly by Technicolor and Inria in the area of 2D video coding.

9.3. Teaching - Supervision - Juries

9.3.1. Teaching

Master: C. Guillemot, Image and video compression, 12 hours, M2 computer science, Univ. of Rennes 1, France.

Master: C. Guillemot, Image and video compression, 12 hours, M2 SISEA, Univ. of Rennes 1, France.

Master: L. Guillo, Multimedia Communication, 15 hours, M2 Network Engineering, 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, Acquisition/Image processing/Compression, 16 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, 16 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. Supervision

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

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

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

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

PhD in progress : M. Ebdelli, Video inpainting for editing and loss concealment, Dec. 2010, C. Guillemot and O. Le Meur

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

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

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

PhD in progress : Julio-Cesar Ferreira, multi-view super-resolution, Oct. 2013, C. Guillemot and O. Le Meur (Co-tutelle with University of Uberlandia, Brazil).

9.3.3. Juries

- C. Guillemot has been member (rapporteur) of the jury of the HDR committee of:
  - M. Cagnazzo, Telecom ParisTech, Sept. 2013
- C. Guillemot has been member (rapporteur) of the jury of the PhD committee of:
  - G. Pettrazuoli, Telecom ParisTech,Jan. 2013
- C. Labit has been member (as examiner) of the jury of the HDR committee of:
  - Marc Chaumont, Université de Montpellier 2, Jul. 2013
- C. Labit has been member (as president) of the jury of the PhD committee of:
  - Mikael Carlavan, Université Nice-Sophia Antipolis, June 2013

9.4. Popularization

- L. Guillo is advisor of secondary school pupils following the course "Informatique et Science du Numérique (ISN)" of the Lycée Jean Macé in Rennes, France.

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, http://hal.inria.fr/hal-00457634

[3] J.-J. FUCHS. *A robust matched detector*, in "IEEE Trans. on Signal Processing", Nov. 2007, vol. 55, n° 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° 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° 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° 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° 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° 8

Publications of the year

Articles in International Peer-Reviewed Journals

[11] S. CHERIGUI, C. GUILLEMOT, D. THOREAU, P. GUILLOTEL, P. PEREZ. *Correspondence map-aided neighbor embedding for image intra prediction*, in "IEEE Transaction on Image Processing, 2012", March 2013, vol. 22, n° 3, pp. 1161-1174, http://hal.inria.fr/hal-00755762

[12] H. DU, Z. LIU, J. JIANG, L. SHEN. *Stretchability-aware block scaling for image retargeting*, in "Journal of Visual Communication and Image Representation", May 2013, vol. 24, n° 4, pp. 499-508 [*DOI :* 10.1016/J.JVCIR.2013.03.003], http://hal.inria.fr/hal-00876303

[13] 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° 1, pp. 269 - 279 [*DOI :* 10.1109/TCOMM.2013.120413.130102], http://hal.inria.fr/hal-00935847

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

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

[16] C. GUILLEMOT, M. TURKAN, O. LE MEUR, M. EBDELLI. *Object removal and loss concealment using neighbor embedding*, in "Eurasip Journal on Signal Processing: Image Communication", September 2013, http://hal.inria.fr/hal-00876062

[17] O. LE MEUR, M. EBDELLI, C. GUILLEMOT. *Hierarchical super-resolution-based inpainting*, in "IEEE Transactions on Image Processing", May 2013, vol. 22, n^{o} 10, pp. 3779-3790 [*DOI :* 10.1109/TIP.2013.2261308], http://hal.inria.fr/hal-00876168

[18] Z. LIU, O. LE MEUR, S. LUO, L. SHEN. *Saliency detection using regional histograms*, in "Optics Letters", February 2013, vol. 38, n^{o} 5, pp. 700-702 [*DOI :* 10.1364/OL.38.000700], http://hal.inria.fr/hal-00876272

[19] M. TURKAN, C. GUILLEMOT. *Dictionary learning for image prediction*, in "Journal of Visual Communication and Image Representation", April 2013, vol. 24, n^{o} 3, pp. 426-437, http://hal.inria.fr/hal-00875968

International Conferences with Proceedings

[20] J. AGHAEI MAZAHERI, C. GUILLEMOT, C. LABIT. *Learning A Tree-Structured Dictionary For Efficient Image Representation With Adaptive Sparse Coding*, in "ICASSP - 38th International Conference on Acoustics, Speech, and Signal Processing", Vancouver, Canada, May 2013, http://hal.inria.fr/hal-00876030

[21] J. AGHAEI MAZAHERI, C. GUILLEMOT, C. LABIT. *Learning an Adaptive Dictionary Structure for Efficient Image Sparse Coding*, in "PCS - 30th Picture Coding Symposium - 2013", San Jose, United States, December 2013, http://hal.inria.fr/hal-00876060

[22] M. ALAIN, S. CHÉRIGUI, C. GUILLEMOT, D. THOREAU, P. GUILLOTEL. *Locally Linear Embedding Methods for Inter Image Prediction*, in "ICIP - IEEE International Conference on Image Processing - 2013", Melbourne, Australia, September 2013, http://hal.inria.fr/hal-00875941

[23] M. BEVILACQUA, A. ROUMY, C. GUILLEMOT, M.-L. ALBERI MOREL. *Compact and coherent dictionary construction for example-based super-resolution*, in "IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)", Vancouver, Canada, May 2013, http://hal.inria.fr/hal-00875964

[24] M. BEVILACQUA, A. ROUMY, C. GUILLEMOT, M.-L. ALBERI MOREL. *K-WEB: Nonnegative dictionary learning for sparse image representations*, in "IEEE International Conference on Image Processing (ICIP)", Melbourne, Australia, September 2013, http://hal.inria.fr/hal-00876018

[25] M. BEVILACQUA, A. ROUMY, C. GUILLEMOT, M.-L. ALBERI MOREL. *Super-resolution using Neighbor Embedding of Back-projection residuals*, in "18th International Conference on Digital Signal Processing (DSP)", Fira, Santorini, Greece, July 2013, http://hal.inria.fr/hal-00876020

[26] M. BEVILACQUA, A. ROUMY, C. GUILLEMOT, M.-L. ALBERI MOREL. *Video super-resolution via sparse combinations of key-frame patches in a compression context*, in "30th Picture Coding Symposium (PCS)", San Jose, United States, December 2013, http://hal.inria.fr/hal-00876026

[27] E. DUPRAZ, A. ROUMY, M. KIEFFER. *Practical Coding Scheme for Universal Source Coding with Side Information at the Decoder*, in "Data Compression Conference", Snowbird, United States, March 2013, pp. 1-11, http://hal.inria.fr/hal-00819491

[28] E. DUPRAZ, A. ROUMY, M. KIEFFER. *Universal Wyner-Ziv Coding for Gaussian Sources*, in "ICASSP 2013", Vancouver, Canada, May 2013, pp. 1-4, http://hal.inria.fr/hal-00819493

[29] M. EBDELLI, O. LE MEUR, C. GUILLEMOT. *Analysis of patch-based similarity metrics: application to denoising*, in "IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)", Vancouver, Canada, May 2013, http://hal.inria.fr/hal-00875984

[30] C. GUILLEMOT, S. CHÉRIGUI, D. THOREAU. *K-NN search using local learning based on regression for image prediction with neighbor embedding*, in "IEEE Intl. Conf. on Acoustics and Signal Processing (IEEE-ICASSP)", Vancouver, United States, May 2013, http://hal.inria.fr/hal-00876123

[31] C. GUILLEMOT, M. TURKAN, O. LE MEUR, M. EBDELLI. *Image inpainting using LLE-LDNR and linear subspace mappings*, in "IEEE Intl. Conf. on Acoustics and Signal Processing (IEEE-ICASSP)", Vancouver, United States, May 2013, http://hal.inria.fr/hal-00876130

[32] Z. LIU, O. LE MEUR, S. LUO. *Superpixel-based saliency detection*, in "WIAMIS - 14th International Workshop on Image and Audio Analysis for Multimedia Interactive Services", Paris, France, IEEE, July 2013, http://hal.inria.fr/hal-00876184

[33] S. LUO, Z. LIU, L. LINA, X. ZOU, O. LE MEUR. *Efficient saliency detection using regional color and spatial information*, in "EUVIP", Paris, France, June 2013, http://hal.inria.fr/hal-00876192

[34] M. MANCAS, O. LE MEUR. *Memorability of natural scenes: the role of attention*, in "ICIP", Sydney, Australia, September 2013, http://hal.inria.fr/hal-00876173

[35] R. MARTINEZ-NORIEGA, A. ROUMY. *Prior and macro-filling order for image completion*, in "IEEE Intl. Conf. on Image Processing (ICIP)", Melbourne, Australia, September 2013, http://hal.inria.fr/hal-00918346

[36] D. WOLINSKI, O. LE MEUR, J. GAUTIER. *3D view synthesis with inter-view consistency*, in "ACM Multimedia", Barcelone, Spain, October 2013, http://hal.inria.fr/hal-00876171

National Conferences with Proceedings

[37] J. AGHAEI MAZAHERI, C. GUILLEMOT, C. LABIT. *Représentations parcimonieuses par un dictionnaire à structure adaptative : application au codage d'images satellitaires*, in "24ème colloque Gretsi", Brest, France, September 2013, http://hal.inria.fr/hal-00876044

[38] E. DUPRAZ, A. ROUMY, M. KIEFFER. *Codage de Sources avec Information Adjacente et Connaissance Imparfaite de la Corrélation : le problème des cadrans*, in "GRETSI 2013", Brest, France, September 2013, pp. 1-4, http://hal.inria.fr/hal-00935785

[39] E. DUPRAZ, A. ROUMY, M. KIEFFER. *Codage Distribué dans des Réseaux de Capteurs avec Connaissance Incertaine des Corrélations*, in "GRETSI 2013", Brest, France, September 2013, pp. 1-4, http://hal.inria.fr/hal-00935788

Other Publications

[40] L. GUILLO, C. GUILLEMOT. , *3D-CE5.h: Additional merge candidates derived from shifted disparity candidate predictors*, January 2013, Joint Collaborative Team on 3D Video Coding Extension Development of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, JCT3V-C0148, 3rd Meeting: Geneva, CH, 17-23 Jan. 2013, http://hal.inria.fr/hal-00917123

[41] T. VIJAYARAGHAVAN, L. ZHANG, Y. CHEN, L. GUILLO, C. GUILLEMOT, J.-L. LIN, Y. CHEN. , *CE3.h: Merge candidates derivation from vector shifting*, July 2013, Joint Collaborative Team on 3D Video Coding Extensions of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, JCT3V-E0126, 5th Meeting: Vienna, AU, July 27- Aug. 2, 2013, http://hal.inria.fr/hal-00917146

[42] T. VIJAYARAGHAVAN, L. ZHANG, Y. CHEN, T. GUIONNET, L. GUILLO, C. GUILLEMOT. , *CE5.h: Merge candidates derivation from vector shifting*, April 2013, Joint Collaborative Team on 3D Video Coding Extensions of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, JCT3V-D0178, 4th Meeting: Incheon, KR, 20-26 Apr. 2013, http://hal.inria.fr/hal-00917137