Activity Report 2018

Team PERCEPT

Computational Visual Perception and Applications

D6 – Media and Interactions

1 Team composition

Researchers and faculty

Kadi Bouatouch, Emeritus Professor, Univ. Rennes 1
Rémi Cozot, Associate Professor, Univ. Rennes 1, head of the team
Olivier Le Meur, Associate Professor, Univ. Rennes 1, head of the team

Associate members

Christian Bouville

Research engineers, technical staff

Kévin Bannier

PhD students

Maxime Rousselot, CIFRE Harmonic (2016-2019)
Dmitry Kuzovkin, CIFRE Technicolor (2016-2019)
Alexandre Bruckert, Doctoral Contract (2018-2021)
Ific Goudé, ARN REVERY (2018-2021)
Léo Maczyta, DGA (in collaboration with SERPICO INRIA team) (2017-2020)

Administrative assistant

Maryse Fouché, INRIA

Visitors

Abdessettar Ghemougui, assistant professor, University at M'Sila, Algeria. Visiting researcher (12/2018-12/2019)
Francesco Banterle, Visual Computing Lab at CNR-ISTI, Pisa, Italy
Alan Chalmers, University of Warwick, UK

Home page

2 Overall objectives

2.1 Overview

The main topic of the PERCEPT research team deals with applied-visual perception. This is a cross-disciplinary project, spanning computer science, cognitive science and visual perception. In the PERCEPT scope, we consider that visual phenomena are induced by the **stimulus itself** as well as by the intent of the **creative director**.

Besides these two points, visual perception also strongly depends on the **observer**. Indeed, his age, his gender, and his socio-cultural background (to name a few) may significantly influence the way the observer look within the scene.

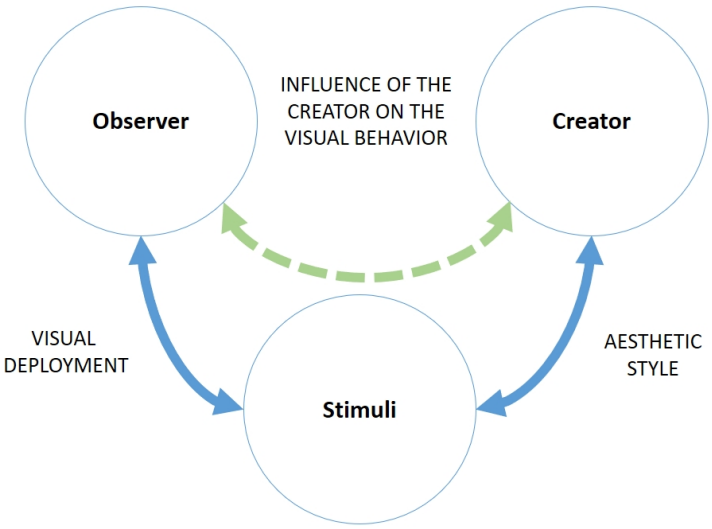

Figure 1: The three research pillars of PERCEPT team

Figure 1 illustrates the dependencies between the observer, the creator and the stimuli. Our research objectives aim to study these three components independently, but, more importantly, jointly. This very last point is the most important scientific challenge for the PERCEPT team. It would require to understand the visual behavior of observers in many viewing conditions, encompassing a variety of content (e.g. natural images, webpages, comics, paintings...), a variety of style and aesthetic.

The visual behavior mainly refers to the **overt attention**, that, per definition, involves eye movement. The observation of eye movements is fundamental for the team objectives. Indeed, eye movements is an exogenous manifestation of where we look at and reflect, to some extent, the visual processing that are involved in the perception of our visual environment. It is common to say that eye movements can be compared to a window on mind and soul [VG07]. The way we look within a visual scene, the way we jump from one location to another may indeed indicate, covertly and subtly, the task we are trying to perform, our gender, our age, and whether we suffer from a visual disease or not. We believe that by monitoring, understanding and modelling the factors

[VG07] R. P. Van Gompel, *Eye movements: A window on mind and brain*, Elsevier, 2007.

that influence our visual behavior, this will open up new avenues for the development of cutting-edge computed-based applications and contents tailored for different groups of people.

Generally speaking and without delving into the details, overt attention is influenced by both bottom-up and top-down processes. The former is driven by properties of the visual stimulus (e.g. data-driven processing), and independent of task or semantics whereas the latter is generally associated with the task at hand. We would like to broaden the definition of top-down processes by considering not only the task at hand but also the specifies of people such as age, gender, cultural heritage and visual disabilities, to name a few. It is also fundamental to consider in the top-down processes the artistic intent, which is connected to aesthetic feeling and emotional arousal, evoked by the onscreen stimulus.

In the following sections, we present and elaborate on the research topics that the team is pursuing.

2.2 Scientific foundations

As mentioned earlier, we aim to design computational models to predict visual phenomena in a number of viewing conditions and with different visual modalities. For that we follow a rather common foundational methodology in science [Kap04]; this consists of several stages providing a guiding and step-by-step strategy for understanding visual perception and for designing innovative methods. The foundational methodology can be boiled down to four stages that are briefly described below:

- Data collection consists in gathering as much as possible information relevant to the problem domain, namely the visual perception for PERCEPT team. Gathering data is a fundamental, yet challenging problem. Two different cases of data collection could be considered: **generic** data and **specific** data. In the former case, we aim to collect visual-related data as faithfully as possible to represent for instance the general behavior of observers or the general trend of director creators. For the sake of completeness, incorporating rare events or specific conditions is here an important aspect. At the opposite, when we have to investigate a particular behavioral case such as visual deployment on comics or paintings, it is important to limit and to constraint the experimental method in order to get data in line with the target application with as few outliers as possible.

- Data understanding is a mandatory stage to get the initial insights about the data. This stage encompasses data visualization that could appear fundamental to reveal clearly and efficiently important information through graphical means. Beyond this, data understanding consists also in gathering descriptive statistic that provides a simple summary about the observations that have been made.

- Modelling and performance evaluation consist in designing an efficient computational model for simulating the targeted behavior. The design of such models

[Kap04] D. Kaplan, *The Sage handbook of quantitative methodology for the social sciences*, Sage, 2004.

could rely on a number of scientific methods, such as non-parametric or parametric statistical methods, unsupervised or supervised machine learning (e.g. Support Vector Machine, deep learning...). The performance evaluation of the proposed models is also a key point, requiring in many cases the design of new assessment methods able to evaluate objectively the similarity degree between the observation and the prediction.

- Deployment consists here in making the computational models of visual perception available for image processing applications. The use of such models aims to improve existing image processing applications as well as to define new and innovative applications.

According to the foundational methodology presented above, the main scientific challenges targeted by the PERCEPT team are introduced below. As mentioned in section 2.1, these challenges are about the factors that influence the way we look onscreen and the computational models simulating these behaviors.

2.2.1 Computational models of visual attention

Saliency models Most of the computational models of bottom-up overt visual attention, which date back to the 1980s [CF88,TCKW+95,IKN98], have been motivated by the seminal work of Koch and Ullman [KU85]. They proposed a plausible computational architecture to predict human gaze. From a set of feature maps processed in a massively parallel manner, a single topographic saliency map, which encodes the ability of an area to attract our gaze, is computed. Since then, there has been a growing interest in this subject. A number of models, more or less biological and using different mathematical tools, have been proposed. A review of 63 saliency models from the literature showed that 47 of them use bottom-up factors, and 23 use top-down factors [BI13]. Recently, a new generation of models, relying on deep networks, has even brought a new momentum in this field of research [KTB14,KWB16,PCM+17,PSGiN+16,CBSC16].

[CF88] J. J. Clark, N. J. Ferrier, "Modal Control Of An Attentive Vision System", *in: ICCV*, IEEE, p. 514–523, 1988.

[TCKW+95] J. K. Tsotsos, S. M. Culhane, W. Y. Kei W., Y. Lai, N. Davis, F. Nuflo, "Modeling visual attention via selective tuning", *Artificial intelligence 78*, 1, 1995, p. 507–545.

[IKN98] L. Itti, C. Koch, E. Niebur, "A model for saliency-based visual attention for rapid scene analysis", *IEEE Trans. on PAMI 20*, 1998, p. 1254–1259.

[KU85] C. Koch, S. Ullman, "Shifts in selective visual attention: towards the underlying neural circuitry", *Human Neurobiology 4*, 1985, p. 219–227.

[BI13] A. Borji, L. Itti, "State-of-the-art in Visual Attention Modeling", *IEEE Trans. on Pattern Analysis and Machine Intelligence 35*, 1, 2013, p. 185–207.

[KTB14] M. Kümmerer, L. Theis, M. Bethge, "Deep gaze i: Boosting saliency prediction with feature maps trained on imagenet", *arXiv preprint arXiv:1411.1045*, 2014.

[KWB16] M. Kümmerer, T. S. Wallis, M. Bethge, "DeepGaze II: Reading fixations from deep features trained on object recognition", *arXiv preprint arXiv:1610.01563*, 2016.

[PCM+17] J. Pan, C. Canton, K. McGuinness, N. E. O'Connor, J. Torres, E. Sayrol, X. Giro-i Nieto, "Salgan: Visual saliency prediction with generative adversarial networks", *arXiv preprint arXiv:1701.01081*, 2017.

[PSGiN+16] J. Pan, E. Sayrol, X. Giro-i Nieto, K. McGuinness, N. E. O'Connor, "Shallow and deep convolutional networks for saliency prediction", *in: Proceedings of the IEEE*

The common denominator between these models is that they all output a 2D static saliency map. Although the saliency map representation is a convenient way to indicate where we look within a scene, these models do not completely account for the complexities of our visual system. One obvious limitation concerns the fact that these models do not make any assumption about eye movements and oculomotor behavioural biases [TV09]. For instance, they implicitly make the hypothesis that eyes are equally likely to move in any direction.

Saccadic models Rather than computing a unique saliency map, saccadic models aim to predict visual scanpaths, *i.e.,* the sequence of fixations and saccades an observer would perform to sample the visual environment. As saliency models, saccadic models have also to predict the salient areas of our visual environment. But the great difference with saliency models is that saccadic models have to output plausible visual scanpaths, *i.e.,* having the same peculiarities as human scanpaths. Ellis and Smith [ES85] pioneered in this field by elaborating a general framework for generating visual scanpaths. They used a stochastic process where the position of a fixation depends on the previous fixation, according to a first-order Markov process.

In the past, we published two important contributions for improving the performance of saccadic model [LML15,LMC16]. The first paper presents a new framework embedding important characteristics of visual scanpaths. These characteristics, also called viewing biases or viewing tendencies [TV08], encompass the fact that saccades of small amplitudes are far more numerous than long saccades, and that there is an asymmetry in saccade orientation. Horizontal saccades (leftwards or rightwards) are more frequent than vertical ones, which are much more frequent than oblique ones [FK10]. Another important bias is the central bias. Fixations are not distributed evenly throughout the scene but are much more frequent at the center of an image. There are a number of reasons for this central bias. Among them we just mention the bias from the photographer, who tends to place the object of interest near the image's center, and the orbital reserve (eye

Conference on Computer Vision and Pattern Recognition, p. 598–606, 2016.

[CBSC16] M. Cornia, L. Baraldi, G. Serra, R. Cucchiara, "A Deep Multi-Level Network for Saliency Prediction", *in : International Conference on Pattern Recognition (ICPR)*, 2016.

[TV09] B. Tatler, B. T. Vincent, "The prominence of behavioural biases in eye guidance", *Visual Cognition, Special Issue: Eye Guidance in Natural Scenes 17*, 6-7, 2009, p. 1029–1059.

[ES85] S. R. Ellis, J. D. Smith, *Patterns of statistical dependency in visual scanning*, Elsevier Science Publishers BV, (eds) Amsterdam, North Holland Press, 1985, ch. Eye Movements and Human Information Processing, p. 221–238.

[LML15] O. Le Meur, Z. Liu, "Saccadic model of eye movements for free-viewing condition", *Vision research 1*, 1, 2015, p. 1–13.

[LMC16] O. Le Meur, A. Coutrot, "Introducing context-dependent and spatially-variant viewing biases in saccadic models", *Vision Research 121*, 2016, p. 72–84.

[TV08] B. Tatler, B. Vincent, "Systematic tendencies in scene viewing", *Journal of Eye Movement Research 2*, 2008, p. 1–18.

[FK10] T. Foulsham, A. Kingstone, "Asymmetries in the direction of saccades during perception of scenes and fractals: Effects of image type and image features", *Vision Research 50*, 8, 2010, p. 779–795.

direction relative to the head). A review of central bias factors is documented by Tseng et al. [TCC+09].

2.2.2 Inferring high-level information from the visual deployment

Through eye tracking experiments, the overt visual attention can be observed and recorded. Eye movements reveal where and how an observer look within a scene displayed onscreen. More specifically, we are interested in analysing the sequence of visual fixations and saccades made during the observation. Fixations aim to bring objects of interest into the fovea where the visual acuity is maximum. Saccades are ballistic changes in eye position, allowing to jump from one position to another. This sequence of fixations and saccades is called visual scanpaths.

Visual information extraction essentially takes place during the fixation period. The way we look within a visual scene, the way we jump from one location to another indicates explicitly, but covertly and subtly, many information about our personality and the cognitive state of our mind. Below we briefly elaborate on the main high-level information that can be extracted from our gaze. The main scientific papers dealing with those aspects are underlined.

Inferring the age of an observer from his gaze. Eye movement patterns during scene perception have been investigated extensively for adults [MCMH04,Ray09]. Concerning early childhood and adolescence, there is only very few eye tracking studies that have been performed with natural scenes. In 2012, Kirkorian et al. [KAK12]. have shown that similarity of gaze location increased with age. The very young, assuming normal or corrected visual acuity, may rely more on bottom-up mechanisms, *i.e.,* low-level visual salience, than adults to inspect visual scenes. In 2014, Helo et al. [HPSR14] found out that fixation durations decreased while saccade lengths increased with age. They also observed that computational models of bottom-up visual attention is a better predictor for young people's eye movements than for adults.

The PERCEPT team is one of the few research team investigating successfully this objective. We already published key papers on this topic [ZLM18,LMCLR+17,LMCL+17].

[TCC+09] P. Tseng, R. Carmi, I. G. M. Cameron, D. Munoz, L. Itti, "Quantifying center bias of observers in free viewing of dynamic natural scenes", *Journal of Vision 9*, 7:4, July 2009, p. 1–16.

[MCMH04] S. Martinez-Conde, S. L. Macknik, D. H. Hubel, "The role of fixational eye movements in visual perception", *Nature Reviews Neuroscience 5*, 3, 2004, p. 229–240.

[Ray09] K. Rayner, "Eye movements and attention in reading, scene perception, and visual search", *The quarterly journal of experimental psychology 62*, 8, 2009, p. 1457–1506.

[KAK12] H. L. Kirkorian, D. R. Anderson, R. Keen, "Age differences in online processing of video: an eye movement study", *Child development 83*, 2, 2012, p. 497–507.

[HPSR14] A. Helo, S. Pannasch, L. Sirri, P. Rämä, "The maturation of eye movement behavior: Scene viewing characteristics in children and adults", *Vision research 103*, 2014, p. 83–91.

[ZLM18] T. Zhang, O. Le Meur, "How Old Do You Look? Inferring Your Age From Your Gaze", *in: 2018 25th IEEE International Conference on Image Processing (ICIP)*, IEEE, p. 2660–2664, 2018.

[LMCLR+17] O. Le Meur, A. Coutrot, A. Le Roch, A. Helo, P. Rämä, Z. Liu, "Age-

Inferring the task or the identity of an observer from his gaze. It is well known that the task influences the visual deployment. The first study dates back to 1967 with the seminal book of Yarbus [Yar67]. The second part of his book presents a variety of eye tracking experiments. Amongst them, he studied how the instructions given to an observer influence the observer's eye movement behaviour, when viewing the painting, *The Unexpected Visitor*.

Considering that the task influences the visual deployment, Henderson et al. [HSW+13] demonstrate that it is also possible to classify the task that a person is engaged in from their eye movements. Four tasks were considered: scene search, scene memorization, reading, and pseudo-reading. Results showed that tasks could be identified reliably above chance.

Influence the visual disease of an observer from his gaze. A recent study demonstrates that it is possible to detect whether we suffer from a visual disease or not [Itt15]. Autism Spectrum Disorder (ASD) affects one in 68 people in the US [AI14]. Given that there are limited clinical resources, it becomes urgent to design new diagnostic tools for screening the disease as early as possible. A new diagnostic paradigm recently appears that consists in analysing eye movements while observers watch a natural scene onscreen [JZ17]. In 2019, PERCEPT team will investigate this new field thanks to a new funded exploratory CominLabs project.

For the sake of completeness, we could add the inference of the emotional state and the inference of gender

2.2.3 Understanding and modelling the creator intent

In the context of PERCEPT team, we study only images that are created by humans in order to convey a mood, a feeling or a message. This excludes images that are only a capture of a scene, such CCTV images, without any aesthetic intent. Through the ages, images creators have defined a visual language based on aesthetics choices. These aesthetics choices concern: composition, framing, light, color, sharpness, texture, image

dependent saccadic models for predicting eye movements", *in : Image Processing (ICIP), 2017 IEEE International Conference on*, IEEE, p. 3740–3744, 2017.

[LMCL+17] O. Le Meur, A. Coutrot, Z. Liu, P. Rämä, A. Le Roch, A. Helo, "Your gaze betrays your age", *in : Signal Processing Conference (EUSIPCO), 2017 25th European*, IEEE, p. 1892–1896, 2017.

[Yar67] A. Yarbus, *Eye movements and vision*, Plenum Press: New York, 1967.

[HSW+13] J. M. Henderson, S. V. Shinkareva, J. Wang, S. G. Luke, J. Olejarczyk, "Predicting cognitive state from eye movements", *PloS one 8*, 5, 2013, p. e64937.

[Itt15] L. Itti, "New Eye-Tracking Techniques May Revolutionize Mental Health Screening", *Neuron 88*, 3, Nov 2015, p. 442–444.

[AI14] Autism, D. D. M. N. S. Y. . P. Investigators, "Prevalence of autism spectrum disorder among children aged 8 years-autism and developmental disabilities monitoring network, 11 sites, United States, 2010", *Morbidity and Mortality Weekly Report: Surveillance Summaries 63*, 2, 2014, p. 1–21.

[JZ17] M. Jiang, Q. Zhao, "Learning Visual Attention to Identify People with Autism Spectrum Disorder", *in : Proceedings of the IEEE International Conference on Computer Vision*, p. 3267–3276, 2017.

size, image orientation, etc. Our objective is to understand the aesthetic choices made by the image creator and to automatically compute an aesthetics report that describes her/his choice. We call this: the **aesthetic intent model**. To achieve this, we have first to take account of human visual perception and compute the appearance of image by means of an **image appearance model**. Indeed, pixel values do not reflect directly the human perception of image. For example, a perceived hue of a small patch depends on the neighbouring color.

Image appearance model [Fai97]**.** The perception of an image depends on the viewing conditions and the spatial structure of the image. An image appearance model computes, from pixels values and viewing conditions, perceptual quantities such as lightness, brightness, colorfulness, chroma and saturation. State of the art image appearance models are limited to standard display viewing conditions.

Nowadays these models require new psycho-physics studies in order to cope with new displays such as High Dynamic Range displays and Head Mounted Displays. Indeed High Dynamic Range imaging enables to capture the luminance range equal to human sensitivity range. Consequently, all details captured in a High Dynamic Range image can be perceived by human eye. Unfortunately, in most of case, High Dynamic Range images have to be tone mapped in order to be displayed on a screen which capability is limited in terms of dynamic range. Tone mapping algorithm objectives are consequently twofold: first, producing an image that look natural and artefact free, second, preserving the details visible by human visual system. Similarly, state of the art color appearance models do not cover head mounted display viewing condition. Further color appearance models that cover high Dynamic Range imaging and Head Mounted Display are required. That why in PERCEPT team we address these new viewing condition and propose color appearance models, quality metrics that could help to design new perception aware image processing algorithm and especially image retargeting algorithms. A common issue is the lack of visual content, for example high dynamic range omnidirectional content are few. The computer graphics background of PERCEPT team, and especially on global illumination algorithm, enable to compute images required in our studies.

The scientific methodology for the design of image appearance model is based on two main stages: first, psycho physics studies are performed to measure Just Noticeable Differences (JND) related to light or color perception; second, given these JND data, response functions can be easily determined.

Aesthetic intent model [Jol09]**.** The literature related to visual art provides many cognitive models of aesthetics. Most of these models are multidimensional and share common features. They point out that the main aesthetic dimension are: framing, composition, light , color palette, color harmony, image size and image ratio. Creator's choice relative to an aesthetics dimension, for example light, is classified as aesthetics style. For example the light style can be either high key, medium key key or low key. High key means that the image appears bright and low contrasted. A computational

[Fai97] M. D. Fairchild, *Color appearance models*, Addison-Wesley-Longman, 1997.
[Jol09] M. Joly, *Introduction Ã l'analyse de l'image*, Armand Colin, 2009.

aesthetic intent model computes the style chosen by the creator related to a number of aesthetics dimension. Its consists of an aesthetics reports. Consequently aesthetic intent model is a classification algorithm that map pixels values into aesthetics styles.

They are two main approaches for computing of aesthetics intent: supervised or unsupervised classification. A supervised approach requires image database with aesthetics labels. There are only few such databases and they do not cover enough aesthetics. Unsupervised approaches compute main aesthetics features directly from a huge amount of images. Due to their statistics bases, the main limitation of these methods is the few aesthetics they are able to identify.

2.2.4 Computational models of image aesthetics quality

Aesthetics quality assessment consists in automatically appraising the beauty of an image. There are many underlying assumptions in the works addressing this topic. The most important one is the assumption related to the objective nature of beauty. Although that there has been many debates on this topic and that beauty criteria have changed throughout ages, the agreement between people belonging to a homogeneous cultural group suggests that it is possible to predict successfully the aesthetics of a visual scene.

Beauty appraisal is subjective by nature: it is a personal judgment even if it reflects a common and shared taste. Consequently, image aesthetics assessment methods are supervised models that use different kinds of machine learning algorithms. The input is a database that consists of images and associated aesthetics scores. Thanks to this kind of database, a machine learning algorithm is able to compute a mapping function between image features and aesthetics score. Some aesthetics quality assessment method just classify image into two aesthetics class, such as *ugly* vs *beautifull*, or *amateur* vs *professional*. Other methods compute a continuous score.

Despite recent great achievements due to deep learning, these methods are not reliable for assessing the aesthetics quality of images of a photo collections that contains many similar images of the same scene. In addition, in the case of professional photography, these recent methods do not perform well. We tackle these two main limitations by investigating context-aware and aesthetics/style-aware methods.

2.2.5 Visual perception with various image modalities

Virtual Reality (VR). The basic idea of VR is easy to describe, but hard to achieve. VR could be defined as a realistic, immersive and interactive simulation of a 3D world. This simulation should match our physical world as closely as possible, by fooling our brain, letting us believe that we are in a real world. Understanding how our senses, brains, and bodies work is then crucial for designing VR systems. There are a wealth of studies dealing with many aspects of human visual perception such as motion perception, perception of the dynamic properties of objects. In the PERCEPT project, we aim to investigate both the influence of VR on both eye movements and the fidelity of virtual world with respect to the human visual system.

Omnidirectional image or 360° images [JRLMJ18]**.** A particular case of VR is related to the 360° images. An omnidirectional (or 360°) image describes the visual information converging to a given point from any direction. It enables to restitute, at a given position, any viewing angle desired by a user. Many applications benefit from this powerful format. In particular, the interest in omnidirectional imaging has exploded during the recent years due to the arrival of 360° cameras to the public market. The key feature of omnidirectional images is that they are not meant to be watched entirely at the same time, as opposed to classical 2D images. In other words, a user has the opportunity to choose the position of his *viewport*, *i.e.,* the frame on which the visible portion of the spherical image is projected. Some part of the omnidirectional data might even not be observed by users, or might be observed with a lower probability. Therefore, it would be greatly beneficial for coding and streaming algorithms to know *a priori* where a user is likely to watch within a 360° image [HS16,XCS17]. This might also be useful for retargeting algorithms aiming the best display of an omnidirectional content on a 2D screen.

High Dynamic Range (HDR) and Wide Color Gamut (WCG). High Dynamic Range imaging consists of image capture, coding and display that enable to capture and display the whole dynamic range to which human visual system is sensitive. HDR image capture can recover details in both shadows and highlights. With this technological breakthrough, new Tone Mapping and Tone Expansion operators have been proposed. Tone Mapping operator scale down the dynamic range of an HDR image to fit legacy displays capabilities, while Tone Expansion operators scale up the dynamic range of legacy images to HDR domain.

Even if human visual perception in dark or bright stimulus are well known, HDR image perception still needs to be furhter investigated. In addition, HDR imaging imaging enables, for image creator, a new field for designing new image aesthetics. In the PERCEPT project, we aim to investigate both the HDR image perception and HDR image aesthetics.

The PERCEPT team has pointed out that one crucial challenge during tone retargeting (tone mapping or tone expansion) is to preserve the aesthetics intent of the original image. We continue to investigate aesthetics challenges due to the HDR imaging breakthrough.

Computer generated images. Generating CG (computer generated) images is a difficult problem requiring to master a number of key algorithms such as global illumination, relighting, Bayesian Monte Carlo method (e.g. Quasi Monte Carlo).

[JRLMJ18] B. John, P. Raiturkar, O. Le Meur, E. Jain, "A Benchmark of Four Methods for Generating 360Â° Saliency Maps from Eye Tracking Data", *in : 2018 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, IEEE, p. 136–139, 2018.

[HS16] M. Hosseini, V. Swaminathan, "Adaptive 360 VR video streaming based on MPEG-DASH SRD", *in : IEEE International Symposium on Multimedia (ISM)*, IEEE, p. 407–408, San Jose, CA, USA, Dec. 2016.

[XCS17] J. C. X. Corbillon, A. Devlic, G. Simon, "Viewport-Adaptive Navigable 360-Degree Video Delivery", *in : Proc. of IEEE ICC*, May 2017.

QMC (Quasi Monte Carlo) are extensively used in rendering algorithms but among the many possible techniques proposed in the literature, it is very difficult to predict what will be their performances when applied to rendering. As the benefits of QMC compared to classic Monte Carlo depend on the integrand smoothness, this factor must necessarily be taken into account in a performance criterion. In PERCEPT team, we investigate a theoretical framework that enables to predict the performances of QMC techniques applied to global illumination to compute the illumination integral.

2.3 Application domains

The PERCEPT project mainly deals with applied-visual perception, which encompasses a number of applications. Our ambition is to improve existing applications and to define new and innovative ones. We draw up below a list of applications that could benefit from the computational models of visual perception:

- Eye-tracking-based applications:
  - → Inferring the age/gender/mood/emotion from the visual scanpath of an observer could be used in a number of practical applications. The main idea would be to monitor eye movements to infer the cognitive state as well as the profile of an observer. Such an information could be used as an input for intelligent human-computer graphical user-interfaces (GUI). These GUI would be dynamically tailored to the current cognitive state of an observer.
  - → Assistive applications to support people with impairment disabilities.
  - → Medical application such as early diagnoses. Eye movements provide a valuable complement to more conventional neuropsychiatric assessment that uses questionnaires and clinical evaluations, as demonstrated in [Itt15]. This kind of applications can significantly benefit from recent advances in machine learning (e.g. deep learning), to offer more sensitive and specific diagnoses as well as the possibility to provide early diagnoses.
- (Saliency-based) Content editing such as exemplar-based inpainting, refocusing...
- (Saliency-based) Content browsing such as automatic selection of the most beautiful pictures, detection of poor quality / redundant pictures.
- Media retargeting based on the visual salience: tone mapping and inverse tone mapping, reframing [MLML17], style transfer.

[Itt15] L. Itti, "New Eye-Tracking Techniques May Revolutionize Mental Health Screening", *Neuron 88*, 3, Nov 2015, p. 442–444.

[MLML17] T. Maugey, O. Le Meur, Z. Liu, "Saliency-based navigation in omnidirectional image", *in : Multimedia Signal Processing (MMSP), 2017 IEEE 19th International Workshop on*, IEEE, p. 1–6, 2017.

3 Scientific achievements

3.1 Deepcomics: Saliency estimation for comics [BJLM18]

Participants: Kévin Bannier, Olivier Le Meur, Eakta Jain.

Joint work with E. Jain (Univ. of Florida)

A key requirement for training deep learning saliency models is large training eye tracking datasets. Despite the fact that the accessibility of eye tracking technology has greatly increased, collecting eye tracking data on a large scale for very specific content types is cumbersome, such as comic images, which are different from natural images such as photographs because text and pictorial content is integrated. In this study, we show that a deep network trained on visual categories where the gaze deployment is similar to comics outperforms existing models and models trained with visual categories for which the gaze deployment is dramatically different from comics. Further, we find that it is better to use a computationally generated dataset on visual category close to comics one than real eye tracking data of a visual category that has different gaze deployment. These findings hold implications for the transference of deep networks to different domains.

3.2 A Benchmark of Four Methods for Generating 360° Saliency Maps from Eye Tracking Data [JRLMJ18]

Participants: Brendan John, Pallavi Raiturkar, Olivier Le Meur, Eakta Jain.

Joint work with B. John, P. Raiturkar, E. Jain (Univ. of Florida)

Modeling and visualization of user attention in Virtual Reality is important for many applications, such as gaze prediction, robotics, retargeting, video compression, and rendering. Several methods have been proposed to model eye tracking data as saliency maps. We benchmark the performance of four such methods for 360° images. We provide a comprehensive analysis and implementations of these methods to assist researchers and practitioners. Finally, we make recommendations based on our benchmark analyses and the ease of implementation.

[BJLM18] K. Bannier, E. Jain, O. Le Meur, "Deepcomics: Saliency estimation for comics", *in : Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*, 2018.

[JRLMJ18] B. John, P. Raiturkar, O. Le Meur, E. Jain, "A Benchmark of Four Methods for Generating 360Â° Saliency Maps from Eye Tracking Data", *in : 2018 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, IEEE, p. 136–139, 2018.

3.3 RGBD co-saliency detection via multiple kernel boosting and fusion [WLSLM18]

Participants: Lishan Wu, Zhi Liu, Hangke Song, Olivier Le Meur.

Joint work with L. Wu, Z. Liu, H. Song (Univ. Shanghai)

RGBD co-saliency detection, which aims at extracting common salient objects from a group of RGBD images with the additional depth information, has become an emerging branch of saliency detection. In this regard, this work proposes a novel framework via multiple kernel boosting (MKB) and co-saliency quality based fusion. First, on the basis of pre-segmented regions at multiple scales, the regional clustering by feature bagging is exploited to generate the base co-saliency maps. Then the clustering-based samples selection is performed to select the most similar regions with high saliency from different images in the image set. The selected samples are utilized to learn a MKB-based regressor, which is applied to all regions at multiple scales to generate the MKB-based co-saliency maps. Finally, to make full use of both MKB and clustering-based co-saliency maps, a co-saliency quality criterion is proposed for adaptive fusion to generate the final co-saliency maps. Experimental results on a public RGBD co-saliency detection dataset demonstrate that the proposed co-saliency model outperforms the state-of-the-art co-saliency models.

3.4 Image Selection in Photo Albums [KPC+18]

Participants: Dmitry Kuzovkin, Tania Pouli, Rémi Cozot, Olivier Le Meur, Jonathan Kervec, Kadi Bouatouch.

The selection of the best photos in personal albums is a task that is often faced by photographers. This task can become laborious when the photo collection is large and it contains multiple similar photos. Recent advances on image aesthetics and photo importance evaluation has led to the creation of different metrics for automatically assessing a given image. However, these metrics are intended for the independent assessment of an image, without considering the possible context implicitly present within photo albums. In this work, we perform a user study for assessing how users select photos when provided with a complete photo album - a task that better reflects how users may review their personal photos and collections. Using the data provided by our study, we evaluate how existing state-of-the-art photo assessment methods perform relative to user selection, focusing in particular on deep learning based approaches. Finally, we explore a recent framework for adapting independent image scores to collections and evaluate in which scenarios such an adaptation can prove beneficial.

[WLSLM18] L. Wu, Z. Liu, H. Song, O. Le Meur, "RGBD co-saliency detection via multiple kernel boosting and fusion", *Multimedia Tools and Applications*, 2018, p. 1–15.

[KPC+18] D. Kuzovkin, T. Pouli, R. Cozot, O. Le Meur, J. Kervec, K. Bouatouch, "Image Selection in Photo Albums", *in : Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval, ICMR '18*, ACM, p. 397–404, New York, NY, USA, 2018, http://doi.acm.org/10.1145/3206025.3206077.

3.5 Aesthetics aware lighting [GLCC18]

Participants: Rémi Cozot.

Joint work with Quentin Galvane (IRISA/Univ. Rennes), Christophe Lino (LTCI, Telecom ParisTech, Paris-Saclay University), Marc Christie (IRISA/Univ. Rennes).

The placement of lights in a 3D scene is a technical and artistic task that requires time and trained skills. Most 3D modelling tools only provide a direct control of light sources, through the manipulation of parameters such as size, location, flux (the perceived power of light) or opening angle (the light frustum). Approaches have been relying on automated or semi-automated techniques to relieve users from such low-level manipulations at the expense of an important computational cost. In this work, guided by discussions with experts in scene and object lighting, we propose an indirect control of area light sources. We first formalize the classical 3-point lighting design principle (key-light, fill-lights and back/rim-lights) in a parametric model. Given a key-light placed in the scene, we then provide a computational approach to (i) automatically compute the position and size of fill-lights and back/rim-lights by analyzing the geometry of 3D character, and (ii) automatically compute the flux and size of key, fill and back/rim lights, given a sample reference image in a computationally efficient way. Results demonstrate the benefits of the approach on the quick lighting of 3D characters, and further demonstrate the feasibility of interactive control of multiple lights through image features.

3.6 Inferring age from eye movements

3.6.1 Your gaze betrays your age [LMCL+17]

Participants: Olivier Le Meur, Antoine Coutrot, Zhi Liu, Pia Rämä, Adrien Le Roch, Andrea Helo.

Joint work with A. Coutrot from LS2N (Nantes), Pr. Z. Liu (Shanghai Univeristy), R. Pia and A. Helo (Paris VI).

In this study, we present evidence that information derived from observers' gaze can be used to infer their age. This is the first study showing that simple features extracted from the ordered sequence of fixations and saccades allow us to predict the age of an observer. Eye movements of 101 participants split into 4 age groups (adults, 6-10 year-old, 4-6 year-old and 2 year-old) were recorded while exploring static images. The analysis of observers' gaze provides evidence of age-related differences in viewing patterns. Therefore, we extract from the scanpaths several features, including fixation durations and saccade amplitudes, and learn a direct mapping from those features to age using Gentle AdaBoost classifiers. Experimental results show that the proposed image-

[GLCC18] Q. Galvane, C. Lino, M. Christie, R. Cozot, "Directing the Photography: Combining Cinematic Rules, Indirect Light Controls and Lighting-by-Example", *Computer Graphics Forum*, 2018.

[LMCL+17] O. Le Meur, A. Coutrot, Z. Liu, P. Rämä, A. Le Roch, A. Helo, "Your gaze betrays your age", *in : Signal Processing Conference (EUSIPCO), 2017 25th European*, IEEE, p. 1892–1896, 2017.

blind method succeeds in predicting the age of the observer up to 92% of the time. The use of predicted salience does not further improve the classification's accuracy.

3.6.2 How old do you look? Inferring Your Age from Your Gaze [ZLM18]

Participants: Tianyi Zhang, Olivier Le Meur.

The exploration strategy of a visual scene, represented by a visual scanpath, depends on a number of factors. Among them, the age of the observer plays a significant role. For instance, young kids are making shorter saccades and longer fixations than adults. Based on these observations, we propose a new method for inferring the age of the observer from its scanpath. The proposed method is based on a 1D CNN network which is trained by real eye tracking data collected on five age groups. In order to boost the performance, the training dataset is augmented by predicting a high number of scanpaths thanks to the use of an age-dependent computational saccadic model. The proposed method brings a new momentum in this field not only by significantly outperforming existing methods but also by being robust to noise and data erasure.

3.7 High Dynamic Range

3.7.1 Quality Assessment of HDR/WCG Images Using HDR Uniform Color Spaces [RLMCD19]

Participants: Maxime Rousselot, Olivier Le Meur, Rémi Cozot, Xavier Ducloux.

High Dynamic Range (HDR) and Wide Color Gamut (WCG) screens are able to render brighter and darker pixels with more vivid colors than ever. To assess the quality of images and videos displayed on these screens, new quality assessment metrics adapted to this new content are required. Because most SDR metrics assume that the representation of images is perceptually uniform, we study the impact of three uniform color spaces developed specifically for HDR and WCG images, namely, IC_tC_p, $J_za_zb_z$ and HDR-Lab on 12 SDR quality assessment metrics. Moreover, as the existing databases of images annotated with subjective scores are using a standard gamut, two new HDR databases using WCG are proposed. Results show that MS-SSIM and FSIM are among the most reliable metrics. This study also highlights the fact that the diffuse white of HDR images plays an important role when adapting SDR metrics for HDR content. Moreover, the adapted SDR metrics does not perform well to predict the impact of chrominance distortions.

[ZLM18] T. Zhang, O. Le Meur, "How Old Do You Look? Inferring Your Age From Your Gaze", *in: 2018 25th IEEE International Conference on Image Processing (ICIP)*, IEEE, p. 2660–2664, 2018.

[RLMCD19] M. Rousselot, O. Le Meur, R. Cozot, X. Ducloux, "Quality Assessment of HDR/WCG Images Using HDR Uniform Color Spaces", *Journal of Imaging 5*, 1, 2019, `http://www.mdpi.com/2313-433X/5/1/18`.

3.7.2 Impacts of viewing conditions on HDR-VDP2 [RAD+18]

Participants: Maxime Rousselot, Eric Auffret, Xavier Ducloux, Olivier Le Meur, Rémi Cozot.

HDR and WCG increase significantly quality of viewing experience by rendering impressive images and videos. Automatically assessing the quality of these HDR WCG images is one crucial objective in a broadcast process. Full-reference HDR metrics have been designed in the last years to achieve this objective: HDR-VDP2, HDR-VQM, PU-encoding metrics. Recent studies have pointed out that HDR-VDP2 is one of the best metric. Unfortunately, HDR-VDP2 is quite complex to use due to numerous and sometimes hard-to-know parameters such as display emission spectrum, surround luminance and angular resolution. In this study, we show that HDR-VDP2 does not require an accurate knowledge of the viewing condition parameters. For that, we not only test the impact of these parameters on existing image databases of subjective quality scores, but also we propose a new and complementary image database.

3.8 Color transfer methods

3.8.1 Transformation of the Beta distribution for color transfer methods [HMCB18]

Participants: Hristina Hristova, Olivier Le Meur, Rémi Cozot, Kadi Bouatouch.

We proposes a novel transformation between two Beta distributions. Our transformation progressively and accurately reshapes an input Beta distribution into a target Beta distribution using four intermediate statistical transformations. The key idea is to adopt the Beta distribution to model the discrete distributions of color and light in images. We designed a new Beta transformation which we apply in the context of color transfer between images. Experiments have shown that our method obtains more natural and less saturated results than results of recent state-of-the-art color transfer methods. Moreover, our results portray better both the target color palette and the target contrast.

3.8.2 Transformation of the multivariate generalized Gaussian distribution for image editing [HLMCB18b]

Participants: Hristina Hristova, Olivier Le Meur, Rémi Cozot, Kadi Bouatouch.

[RAD+18] M. Rousselot, E. Auffret, X. Ducloux, O. Le Meur, R. Cozot, "Impacts of viewing conditions on HDR-VDP2", *in : 2018 26th European Signal Processing Conference (EUSIPCO)*, IEEE, p. 1442–1446, 2018.

[HMCB18] H. Hristova, O. L. Meur, R. Cozot, K. Bouatouch, "Transformation of the Beta Distribution for Color Transfer", *in : Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 1: GRAPP,*, INSTICC, SciTePress, p. 112–121, 2018.

[HLMCB18b] H. Hristova, O. Le Meur, R. Cozot, K. Bouatouch, "Transformation of the multivariate generalized Gaussian distribution for image editing", *IEEE transactions on visualization and computer graphics 24*, 10, 2018, p. 2813–2826.

Multivariate generalized Gaussian distributions (MGGDs) have aroused a great interest in the image processing community thanks to their ability to describe accurately various image features, such as image gradient fields. However, so far their applicability has been limited by the lack of a transformation between two of these parametric distributions. In this study, we propose a novel transformation between MGGDs, consisting of an optimal transportation of the second-order statistics and a stochastic-based shape parameter transformation. We employ the proposed transformation between MGGDs for a color transfer and a gradient transfer between images. We also propose a new simultaneous transfer of color and gradient, which we apply for image color correction.

3.8.3 Multi-purpose bi-local CAT-based guidance filter [HLMCB18a]

Participants: Hristina Hristova, Olivier Le Meur, Rémi Cozot, Kadi Bouatouch.

In this study, we propose a new guidance filter, based on color perception through a chromatic adaptation model. Our method consists of a patch-wise linear transformation, which transfers details from a guidance image to an input image. The amount of transferred details is controlled by a novel chromatic adaptation transform (CAT), called bi-local CAT, embedded in our method. The bi-local CAT contributes to the detail recovery from the guidance image as well as to the preservation of the input reflections and shadows. Our bi-local CAT-based guidance filter is applied in various image processing domains, such as image denoising, image deblurring, texture transfer, detail enhancement, skin beautification, etc.

4 Software development

4.1 DeepSaliencyWebPage

Participants: Olivier Le Meur, Kévin Bannier.

Num. APP: IDDN.FR.001.280008.000.S.A.2018.000.21000
This software allows to predict the salience of webpages. An internal document presents the performance of the deep-based proposed approach. Experimental results show that the proposed method ranks first compared to a set of existing models.

4.2 Transformation of the multivariate generalized Gaussian distribution (MGGD) [HLMCB18b]

Participants: Hristina Hristova, Olivier Le Meur, Rémi Cozot, Kadi Bouatouch.

[HLMCB18a] H. Hristova, O. Le Meur, R. Cozot, K. Bouatouch, "Multi-purpose bi-local CAT-based guidance filter", *Signal Processing: Image Communication 65*, 2018, p. 141–153.

[HLMCB18b] H. Hristova, O. Le Meur, R. Cozot, K. Bouatouch, "Transformation of the multivariate generalized Gaussian distribution for image editing", *IEEE transactions on visualization and computer graphics 24*, 10, 2018, p. 2813–2826.

Num. APP: IDDN.FR.001.500017.000.S.P.2017.000.21000
The software performs a novel transformation between MGGDs (multivariate generalized Gaussian distribution), consisting of an optimal transportation of the second-order statistics and a stochastic-based shape parameter transformation. We employ the proposed transformation between MGGDs for a color transfer and a gradient transfer between images. We also propose a new simultaneous transfer of color and gradient, which we apply for image color correction.

5 Contracts and collaborations

5.1 International Initiatives

5.1.1 Spatio-temporal saliency model for unconstrained videos and its applications

Participants: Olivier Le Meur.

- Project type: National Science Foundation of China (Num. 61471230)
- Dates: 2015–2018
- PI institution: Univ. Shanghai (Pr. Z. Liu)

Spatiotemporal saliency model is becoming an international cutting-edge research topic in the recent years. However, the common drawback of the current spatiotemporal saliency models is their insufficiency for processing unconstrained videos. The complexity of spatial and temporal features in unconstrained videos results in a significant degradation on saliency detection performance of the current models. In order to overcome the drawback of current models, this project proposes a spatiotemporal saliency model based on region-level temporal and spatial features. The expected research results will not only enrich and promote the development of research on saliency model, but also advance the saliency-based video processing technology.

5.2 National Initiatives

5.2.1 ANR ASTRID DGA DISSOCIE

Participants: Olivier Le Meur, Anne-Flore Perrin.

- Project type: ANR ASTRID
- Dates: 2017–2020
- PI institution: IETR
- IETR, LS2N, IRISA
- Home page

The aerial surveillance, monitoring and observation with drone present major challenges in terms of defense, security and environment. For example, France and Great Britain

have agreed to invest 2 billion euros in a project to build next-generation multi-role drones capable of carrying out surveillance and observation missions, identifying targets and launching strikes on enemy territory for future operational capacity beyond 2030. However, the observation, targets identification and surveillance missions are currently being carried out by human operators who do not have the ability to fully and effectively exploit all available drone videos. The science and the technology of the eye-tracking study, visual attention modeling, human operator models, and intelligent compression opens up new possibilities to meet these challenges. In this context, the DISSOCIE project aims to develop automatic and semi-automatic operator models capable of detecting salient areas from the point of view of human operators, by considering the low-level characteristics of the salient content in the videos, geo-temporally localized contextual information, and the expertise and the detection strategies of human operators. Machine learning can be used at different levels of this modeling process. The new HEVC video compression standard and the scalable coding will also be exploited in this project to improve the efficiency when the experts rewatch the videos. The originality of the project lies in an innovative approach to jointly address these challenges based on the complementarity and the strengthening of the scientific expertise gathered in the consortium: especially on eye-tracking analysis, visual fixation prediction, visual attention modeling, salient object detection and segmentation, human observer modeling, and video compression.

5.2.2 ANR REVERY

Participants: Rémi Cozot, Ific Goudé.

- Project type: : ANR Société de l'information et de la communication (DS07) 2017
- Dates: 2017–2020
- PI institution: CRESTIC/Univ. Reims Champagne Ardenne
- IRISA, L2S (Laboratoire des Signaux et Systèmes), XD PRODUCTIONS SA
- Home page

The ReVeRY project will design a specific GRID OF CAMERAS, a cost-efficient system that acquires at once several viewpoints under several exposures and will convert a multiview, multiexposed, video stream into a high quality rich media. In the last two decades, industries and researchers proposed significant advances in media content acquisition systems in three main directions: increase of resolution and image quality with the new ultra-high-definition (UHD) standard; stereo capture for 3D content; and high-dynamic range (HDR) imaging. Compression, representation, and interoperability of these new media are active research fields in order to reduce data size and to be perceptually accurate. The creative industries face issues associated with limited possibilities offered by current media. The traditional acquisition pipeline uses and outputs video streams, either UHD, stereo or HDR. After shooting, possible changes on the media (i.e., viewpoint, framing, aperture, lighting) are limited, which may force a film director to shoot the scene again if any changes are necessary. For now, a pre-shooting choice is made according to the nature of the media to capture (2D, stereo, or HDR) depending on the foreseen use. However, these choices are currently never jointly offered.

A major breakthrough of the project is to jointly resolve HDR and depth reconstruction. The ReVeRY project UHD and HDR will endow the produced media with geometrical and color precision and quality surpassing current video media formats. Its very nature will furthermore conveniently allow new usages, such as reframing, free viewpoint, and relighting at both post-production and visualization stages and facilitate live adaptation to different types of displays: theatre, TV, Cloud, HDR or not, 2D or 3D. This later feature makes the new media truly versatile according to consumption conditions.

The huge market investments in virtual reality (VR) are positioning it as a plain new media. VR immersive screens are today very limited, only trapping the user in a sphere's center of 2D multiaxial images. The virtual cameras of virtual reality propose 360° viewing of a poor and only concave 3D transcriptions. A *liberated* VR requires the creation of convex 3D reconstruction tools using specific camera's organization for 3D scanning of real environments and moving actors, giving the user total freedom of visualization through unlimited positioning and adaptive resolutions of a true virtual cameras (stereoscopic or not).

ReVeRY wants to remedy the lack of a joint representation of the different media formats discussed above. This project aims at providing solutions to replace the traditional video media (mono, stereoscopic or 360°) by a richer data stream overcoming all previously listed bottlenecks. This richer media will jointly embed UHD, HDR, and Depth information in a dedicated format. The project will (i) propose this new media along with one or more suitable representations, (ii) develop a demonstrative prototype of a dedicated acquisition system, and (iii) prove their benefits to media creation industry. The project will enable versatile postproduction and re-usability of single shootings of live action under less-restrained range of lighting intensities (e.g., outdoor scenes). Its outcome is important with a twofold impact: (i) societal (for consumers and experts), by increasing content quality and (ii) economical, by reducing production costs. Ultimately, the creation approach will evolve to provide a completely new experience of screen viewing to the public.

5.3 Bilateral industry grants

5.3.1 Technicolor Phd

Participants: Dmitry Kuzokin, Rémi Cozot, Olivier Le Meur.

- Project type: Phd CIFRE
- Title: Spatiotemporal retargeting and recomposition based on artistic rules
- Dates: 2016–2019

The thesis topic is **Spatiotemporal Retargetting and Recomposition based on Artistic Rule**. In order to address the problem of photo recomposition, it is important to be able to evaluate the aesthetic quality of photos. After bench-marking state of the art methods, we figure out that existing approaches present a major limitation in their way of photo assessment: most of the methods perform evaluation of photographs in an independent manner, whereas in a real-life scenario photo assessment is largely affected

by the context (such as other photos present in the same album) or by the subjective factors defined by individual user preferences. Thus, we attempt to model the influence of other photos from a processed photo collection, as well as individual user preferences. We propose a new approach for context-based photo assessment, where the relevant photo context is extracted with help of clustering approach, and an individual photo quality metric is adapted in respect to the similar photos from the same scene.

5.3.2 Harmonic Phd

Participants: Maxime Rousselot, Rémi Cozot, Olivier Le Meur.

- Project type: Phd CIFRE
- Title: Encoder optimization for new High Dynamic Range and Wide Color Gamut video format
- Dates: 2016–2019

The thesis topic is **Optimizing the behavior of encoders for image compression in the face of new television standards**. Main feauture of HDR (High Dynamic Range) and WCG (Wide Color Gamut) images is greater luminance dynamics range and a wider range of colors. In order to able compression optimization, it is essential to correctly and automatically predict image quality. State of the art metrics take into account only the (achromatic) luminance component of the images; they are insensitive to chromatic artifacts. However, several recent works show that in the context of WCG images, compression can lead to severe and annoying chromatic artifacts. To address objectively this problem, we propose a new full-reference quality metrics for HDR/WCG images.

5.4 Collaborations

5.4.1 International

- IVP team (Home page, Univ. Shanghai): Z. Liu
- Jain Lab (Home page, Univ. Florida): E. Jain
- Bertalmio team (Home Page, Universitat Pompeu Fabra): M. Bertalmio
- Visual Computing Lab (Home Page, CNR-ISTI, Pisa): F. Banterle
- Visualisation team (Home Page, Univ. of Warwick): A. Chalmers

5.4.2 National

- SERPICO team (Home page, INRIA Rennes): P. Bouthemy and L. Maczyta
- MIMETIC team (Home page,INRIA Rennes): M. Christie
- VAADER team (Home page, IETR Rennes): L. Zhang, O. Deforges, K. Palma, W. Hamidouche
- IPI team (Home page, LS2N Nantes): V. Ricordel, M. Perreira
- ICONES team (Home page, XLIM Poitiers): C. Larabi
- EA4050, MSHB (Univ Rennes 2): M. Chérel

- LUTIN (Home page, Paris VIII): T. Baccino
- CRESTIC (Home page, Univ. Reims Champagne Ardenne) : C. Loscos
- Laboratoire des Signaux et Systèmes (L2S, UMR 8506), CNRS - CentraleSupelec (Home page, Université Paris-Sud) : F. Dufaux
- STORM research group, CNRS, IRIT (Home page, Univ. Paul Sabatier) : N. Mellado, D. Vanderhaeghe

6 Dissemination

6.1 Promoting scientific activities

6.1.1 Scientific Events Organisation

Co-Organizer of Thematic day of GDR IG-RV in computational photography.

Participants: Rémi Cozot.

The aim of this day is to bring together researchers working on computational photography to identify the diversity of activities in France on this research area, to exchange and communicate on recent advances in the field. This concerns the acquisition, decomposition, processing and digital / virtual modification of photographs, video sequences, or computer-generated images.
The related themes, without being exhaustive, are:

- Photogrammetry, photometry, decomposition into intrinsic lighting data, estimation of depth data,
- 3D video, 360 data composition, HDR imaging,
- Computational cameras,
- Holography,
- Taking into account user perception, style transfer,
- The organization of image / video collections,
- Increased / immersive digital content generation, video data navigation (FFV), texture generation,
- Specific capture systems (lightstages, 360 camera, HDR, 3D, camera grid, lightfield...)

6.1.2 Scientific Events Selection

Chair of Conference Program Committees

- Rémi Cozot was in the technical program committee of CGI, VISIGRAPP.
- Olivier Le Meur was in the technical program committee of EUSIPCO, QoMEX, ICME.
- Kadi Bouatouch was in the technical program committee of CGI, VISIGRAPP.

Reviewer

- Kadi Bouatouch serves as reviewer for SIGGRAPH.
- Rémi Cozot serves as reviewer for international conferences such as CGI, GI, Eurographics, Pacific.
- Olivier Le Meur serves as reviewer for international conferences such as EUSIPCO, QoMex, ICME, ETRA.

6.1.3 Journal

Member of the Editorial Boards.

- Rémi Cozot is associate editor for The Visual Computer Journal since 2016 (more than 30 papers has been processed).
- Olivier Le Meur is associate editor for IET (The Institution of Engineering and Technology) image processing journal since 2016 (more than 50 papers has been processed).

Reviewer - Reviewing Activities.

- Rémi Cozot serves as reviewer for journal such as IEEE Trans. on Image Processing, IEEE Trans. on Multimedia, Computer Graphics Forum, the Visual computer Journal, Journal of Imaging.
- Olivier Le Meur serves as reviewer for journal such as IEEE Trans. on Image Processing, IEEE Trans. on Multimedia, IEEE Trans. on TCSVT.
- Kadi Bouatouch serves as reviewer for ACM SIGGRAPH Conference, VISIGRAPP.

6.1.4 Invited Talks

CORESA 2018.

Participants: Olivier Le Meur.

Title: A guided tour of computational models of visual attention
Since the first computational model of visual attention, proposed in 1998 by Itti et al., a lot of progress has been made. Progress concern both the modelling in itself and the way we assess the performance of saliency models. Recently, new advances in machine learning, more specifically in deep learning, have brought a new momentum in this field. In this tutorial, we present saliency models as well as the metrics used to assess their performances. In particular, we will empathize new saliency models which are based on convolutional neural networks.We will present different deep architectures and the different loss functions used during the training process. We will conclude this presentation by introducing saccadic models which are a generalization of saliency models.

Huawei ISP Workshop 2018.

Participants: Rémi Cozot.

Title: taking into account aesthetics in image capture, editing and display.
The main topic of this talk is aesthetics intent and more precisely the challenges related to aesthetics intent when considering image capture, editing and display. First, we explain the crucial differences between image quality, image aesthetic quality and image aesthetics. Second, using our previous works on image tone retargeting, we focus on how to ensure aesthetic fidelity to an original content when retargeting this original content to a specific display. Finally, we open the scope to main challenges related to aesthetics intent in image editing.

6.1.5 Scientific Expertise

- Rémi Cozot serves as expert for ANR projects.
- Kadi Bouatouch serves as expert for F.R.S.-FNRS: Belgian research agency and KAUST university (King Abdullah University of Sciences and Technology).

6.1.6 Research Administration

- Rémi Cozot is member of IRISA/INRIA virtual reality platform steering committee.

6.2 Teaching, supervision

6.2.1 Teaching

- Master: Rémi Cozot, Computer Graphics, 24h, M1, ESIR Rennes, France
- Master: Rémi Cozot, Global Illumination, 24h, M1, ESIR Rennes, France
- Master: Rémi Cozot, Video Games, 24h, M2, ESIR Rennes, France
- Master: Rémi Cozot, Special Effects, 24h, M2, ESIR Rennes, France
- Master: Rémi Cozot, Introduction to Computer Graphics, 20h, M2, ISTIC Rennes, France
- Master: Rémi Cozot, Research in Computer Graphics, 20h, M2, ISTIC Rennes, France
- Master: Olivier Le Meur, Traitement avancés des images, 14h (CM), 8h (TP), M1, ESIR Rennes, France
- Master: Olivier Le Meur, Analyse et Compression Vidéo, 22h (CM), 14h (TP), M1, ESIR Rennes, France
- Master: Olivier Le Meur, Compression Audio-Visuelle, 24h (CM), 24h (TP), M2, ESIR Rennes, France
- Master: Olivier Le Meur, Image Representation, Editing and Perception, 8h (CM), M2 research, SIF, ISTIC Rennes, France

6.2.2 Supervision

- PhD: Dmitry Kusovkin, 2016-2019, **Spatiotemporal retargeting and recomposition based on artistic rules**. supervisors: Rémi Cozot, Olivier Le Meur.
- PhD: Maxime Rousselot, 2016-2019, **Encoder optimization for new High Dynamic Range and Wide Color Gamut**. Supervisors: Rémi Cozot, Olivier Le Meur.
- PhD: Lé Maczyta, 2017-2020, **Saillance visuelle dynamique dans des séquences d'images**. Supervisors: Patrick Bouthemy, Olivier Le Meur.
- PhD: IfiC Goudé, 2018-2021, **Rendering of 3D point clouds**. Supervisor: Rémi Cozot.
- PhD: Alexandre Bruckert, 2018-2021, **A Perceptual-driven Approach to Film Editing**. Supervisor: Olivier Le Meur.

6.2.3 Jury

PhD Emin Zerman (Télécom Paris Tech): Assessment and analysis of high dynamic range video quality.

Participants: Rémi Cozot (Reviewer).

PhD Debmalya Sinha (University of Warwick): Temporal incident light fields.

Participants: Kadi Bouatouch (External Examiner).

7 Bibliography

Major publications by the team in recent years

[1] K. Bannier, E. Jain, O. Le Meur, "Deepcomics: Saliency estimation for comics", *in: Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*, 2018.

[2] C. Bist, R. Cozot, G. Madec, X. Ducloux, "Tone expansion using lighting style aesthetics", *Computers Graphics 62*, 2017, p. 77–86.

[3] R. Boitard, R. Cozot, D. Thoreau, K. Bouatouch, "Zonal Brightness Coherency for Video Tone Mapping", *Image Commun. 29*, 2, February 2014, p. 229–246, `http://dx.doi.org/10.1016/j.image.2013.10.001`.

[4] M. Ebdelli, O. Le Meur, C. Guillemot, "Video inpainting with short-term windows: application to object removal and error concealment", *IEEE Transactions on Image Processing 24*, 10, 2015, p. 3034–3047.

[5] Q. Galvane, C. Lino, M. Christie, R. Cozot, "Directing the Photography: Combining Cinematic Rules, Indirect Light Controls and Lighting-by-Example", *Computer Graphics Forum*, 2018.

[6] C. Guillemot, O. Le Meur, "Image inpainting: Overview and recent advances", *IEEE signal processing magazine 31*, 1, 2014, p. 127–144.

[7] H. Hristova, O. Le Meur, R. Cozot, K. Bouatouch, "Multi-purpose bi-local CAT-based guidance filter", *Signal Processing: Image Communication 65*, 2018, p. 141–153.

[8] H. Hristova, O. Le Meur, R. Cozot, K. Bouatouch, "Transformation of the multivariate generalized Gaussian distribution for image editing", *IEEE transactions on visualization and computer graphics 24*, 10, 2018, p. 2813–2826.

[9] B. John, P. Raiturkar, O. Le Meur, E. Jain, "A Benchmark of Four Methods for Generating 360° Saliency Maps from Eye Tracking Data", *in : 2018 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, IEEE, p. 136–139, 2018.

[10] O. Le Meur, T. Baccino, "Methods for comparing scanpaths and saliency maps: strengths and weaknesses", *Behavior research methods 45*, 1, 2013, p. 251–266.

[11] O. Le Meur, A. Coutrot, Z. Liu, P. Rämä, A. Le Roch, A. Helo, "Visual attention saccadic models learn to emulate gaze patterns from childhood to adulthood", *IEEE Transactions on Image Processing 26*, 10, 2017, p. 4777–4789.

[12] O. Le Meur, A. Coutrot, "Introducing context-dependent and spatially-variant viewing biases in saccadic models", *Vision research 121*, 2016, p. 72–84.

[13] O. Le Meur, Z. Liu, "Saccadic model of eye movements for free-viewing condition", *Vision Research*, February 2015, `https://hal.inria.fr/hal-01204682`.

[14] C. B. R. Marques, K. Bouatouch, "Optimal sample weights for hemispherical integral quadratures", *Computer Graphics Forum*, 2018.

[15] H. Song, Z. Liu, H. Du, G. Sun, O. Le Meur, T. Ren, "Depth-aware salient object detection and segmentation via multiscale discriminative saliency fusion and bootstrap learning", *IEEE Trans. Image Process 26*, 9, 2017, p. 4204–4216.

[16] I. Thirunarayanan, K. Khetarpal, S. Koppal, O. L. Meur, J. Shea, E. Jain, "Creating segments and effects on comics by clustering gaze data", *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM) 13*, 3, 2017, p. 24.

[17] A. T. Zhang, B. O. Le Meur, "How Old Do You Look? Inferring Your Age From Your Gaze", *in : 2018 25th IEEE International Conference on Image Processing (ICIP)*, IEEE, p. 2660–2664, 2018.