

Journal of Electronic Imaging

JElectronicImaging.org

Perceptually Driven Visual Information Analysis

Mohamed-Chaker Larabi
Sanghoon Lee
Mohammed El Hassouni
Frédéric Morain-Nicolier
Rachid Jennane



Mohamed-Chaker Larabi, Sanghoon Lee, Mohammed El Hassouni, Frédéric Morain-Nicolier, Rachid Jennane, "Perceptually Driven Visual Information Analysis," *J. Electron. Imaging* **25**(6), 061601 (2017), doi: 10.1117/1.JEI.25.6.061601.

Perceptually Driven Visual Information Analysis

Mohamed-Chaker Larabi

Université de Poitiers
XLIM-SIC UMR 6172 CNRS
Boulevard Marie et Pierre Curie BP 30179
86962 Futuroscope, France

Sanghoon Lee

Yonsei University
Multidimensional Insight Lab
C703 MDI Lab 3rd Engineering Building
Seodaemun-gu 120-749, Republic of Korea

Mohammed El Hassouni

Mohammed V University of Rabat
LRIT URAC 29 Laboratory
Faculty of Sciences
Avenue Ibn Battouta, B.P. 1014 RP, Rabat, Morocco

Frédéric Morain-Nicolier

Université de Reims, CRESTIC
9, rue de Québec
10026 Troyes Cedex, France

Rachid Jennane

Université d'Orléans
I3MTO Laboratory
8, rue Léonard de Vinci, BP 6744
F-45072 Orléans, France

In the last decade, the field of visual information analysis and processing became active more than ever because of the large availability of capturing devices at very affordable prices, increasing the amount of exchanged/published visual content in both personal and professional contexts. It becomes difficult to find application fields not using or creating visual information. Therefore, the targeted applications are numerous—for instance medical imaging, security, manufacturing, gaming, etc. To handle such a plethoric amount of data, smart algorithms are designed with the aim of extracting useful information, often on the fly, for taking decisions or assisting experts for such a task. Algorithms are often expected to be computationally plausible, feasible, and reliable. Besides, it is more and more required to handle the data by taking into account the behavior of end-users in order to be as close as possible to the expected results.

When developing imaging solutions, it became natural to account for perceptual features to mimic the human visual system. For instance, the field of compression has proven the importance of using perceptual models when seeking a perceptual tradeoff between quality and bitrate. There is no doubt that, besides bitrate saving, the use of a perceptual strategy helps in distributing the bit-budget on the most visually important areas. Quality assessment has also taken benefit from the use of perceptual models to better target the area of the visual content, influencing, positively or negatively, the end-user judgment. Other fields, such as segmentation, security, fingerprinting, etc., have shown an enthusiasm for adopting perceptual models to achieve more improvement.

Several computational models of the human visual system are available even for use in real-time conditions. However, because of the diversity of targets, applications, fields, and

devices, the user should have a minimum knowledge before being able to select a given model or feature. By analyzing the recent literature, one can find that visual saliency is widely used and that huge efforts have been devoted to the development of comprehensive models for image/video, 2D/3D, gray-scale/color and even for virtual reality. The main explanation of this attractiveness to such a type of perceptual model is linked, as stated before, to the need to preserve resources (bit-budget in the case of compression for instance) and use them on appropriate parts of the content to have a better impact on our perception. Just-noticeable difference (JND) models have also regained interest with the advent of new applications and content types. One can list more than ten recently proposed JND models dedicated to 3D, for instance. The gain achieved with such models does not need to be demonstrated anymore. Besides the aforementioned models, one can refer to visual masking, contrast sensitivity, binocular fusion, to cite a few of them.

This special section presents some of the recent advances in perceptually driven visual information processing with applications to multimedia, consumer cameras, image and video coding, and other related topics. The papers of this special section can be grouped into five categories, according to the addressed subject.

The first category addresses image and video quality metrics. First, [Eddine et al.](#) proposed a no-reference image quality assessment (IQA) metric. The IQA index is built upon the fusion of multiple distortion measures, i.e. relevant metrics are selected with a particle swarm optimization, then the fusion is obtained thanks to a support vector regression (SVR). Extensive validations on the commonly used datasets showed better performance than the classical methods of quality assessment. Also relying on SVR for pooling features, [Cakir and Cetin](#) proposed an IQA framework using the 2D complex mel-cepstrum for feature extraction. Experimental

results showed that promising IQA performances are obtained when introducing image phase information. Tackling video quality assessment, He et al. proposed a multiscale metric motivated by the biological evidence of visual motion perception. To produce motion perception quality, a motion energy model is derived for spatiotemporal slice images. Spatial and motion perception qualities are pooled using the random forests technique. The results proved that the metric ensures a higher consistency with human judgment and a higher generalization capability. Finding that color information has not been fully utilized in the stereoscopic image quality assessment literature, Xu et al. proposed a metric based on learning binocular manifold color visual properties. In the training stage, a feature detector is first created using non-negative matrix factorization manifold regularization by considering color information. Then, in the quality estimation stage, some specific regions are selected and feature vectors are extracted by using the developed feature detector. The final quality score is obtained by combining the scores of the binocular images, which are obtained from some defined metric. The authors evaluated the approach on commonly used databases, which shows a consistency with subjective quality assessment. Finally, Medina et al. proposed a colorimetric validation using radiometric-calibrated photographs of the scene taken according to a set of well-met conditions, as ground-truth information to measure the amount of realism of the results. Obtained results validated the ability to render a computer simulation of a real scene with a minimal number of perceptual differences.

Visual saliency driven applications have been the focus of a large set of papers. Vargic et al. proposed to take advantage of visual saliency obtained from a combination of low- to high-level features map with the aim of improving the performance of the well-known lossy image compression (SPIHT). The saliency information is used to weight the wavelet coefficients. To preserve the perceptually important defocus depth cue and important regions during compression, Khanna et al. proposed an opportunistic bit allocation using visual saliency information comprising both image features and defocus-based depth cue. Quantization values are assigned on the basis of saliency values over a frame. Experimental results showed good results over H.264 as well as pure and defocus saliency methods. Oakes et al. introduced a motion compensated wavelet-based visual attention model (VAM) including the spatial wavelet coefficients as spatial cues and local and global motion for the temporal cues. The proposed VAM is used to develop a video watermarking algorithm by generating a two-level watermarking weighting parameter map to be embedded into the host image according to the visual attentiveness of each region. By avoiding higher strength watermarking in visually attentive regions, the resulting watermarked video achieves high perceived visual quality while preserving high robustness. Saliency has also been exploited for abnormal event detection in video-surveillance applications by Shi et al. They propose to ignore nonsalient regions of the video and use region wise modeling to save time and computing resources while improving detection accuracy. Similarly, Ramadan and Tairi proposed an application of visual saliency for the sake of moving objects segmentation in video. The spatio-temporal saliency map allows extraction of a moving region of interests (ROI). The segmentation is obtained by an active contour seeded by the extracted ROI. Finally, a target search method based on salience mechanism and imaging model is proposed for rough 3D-modeling scenes by Wang and Hu. It generates a search path in which each node is a salient object with respect to its search region. The method solves ambiguities with a speed of search improved by over 50%.

Several papers dealt with perceptually driven segmentation. For instance, Behlim et al. introduced an image representation that encodes structural constraints via local binary patterns (LBP). The obtained segmentation intends to reproduce the segmentation as perceived by the human due to the consideration of the local structures. To perform object recognition, Jung et al. proposed the bipolar edge detection that provides depth via shape from shading information. The perceptual aspect comes from the use of the human visual system (HVS) low-level detectors. The bipolar edges are compared to binary edges in a face recognition task. Another approach of edge detection is proposed by Arora et al. using information set theory. The proposed method targeting color images proved good performance in finding robust edges especially in the presence of impulse noise. Hamid and Khan proposed an algorithm for merging for perceptually accurate line segments. They also proposed a method for quantitative comparison of line segment detection algorithms. Results on the York Urban dataset show that their merged line segments are closer to human-marked ground-truth line segments compared to competing methods. The paper of Chai and et al. deals with the problem of eliminating character-resembled blob or blobs on a detected region from the plate detection stage of automated license plate recognition system. The proposed methodology consists of emphasizing the blob differently in accordance with its location using the reference point that approximates the representative value of true signal properties. The method is evaluated for solving certain types of anomalies. Moving object detection was addressed using an enhanced codebook algorithm to optimize foreground information extraction complexity by Mousse et al. The purpose of the adaptive strategy is to reduce the computational complexity while maintaining the global accuracy. A super-pixel segmentation approach to model the spatial dependencies between pixels is used. The proposed algorithm gives a good performance rate during the foreground detection. On the same topic, ElHarrouss et al. proposed a background subtraction approach where the background modeling approach analyzes the illumination change problem. To achieve high accuracy for motion detection, the authors proposed a threshold function to compute a binary motion mask. Thorough experimental results showed that their method outperforms state-of-the-art models.

Classification is an important stage in various applications to achieve image understanding. Azzakhnini et al. proposed a face classification from RGB-Depth images. In particular, gender and ethnicity are identified from face images based on features (shape and texture) extracted with LBP, Gabor filter, histogram of oriented gradients descriptor, and SIFT. The classification is built upon SVM with AdaBoost. Following a similar goal, Bukar et al. proposed to improve the conventional active appearance model (AAM), by the use of partial least squares regression instead of PCA. The novel feature extraction model is then applied to the problems of age estimation and gender classification. The performance evaluation operated on the FGNET-AD benchmark database showed that the proposed strategy has a better predictive power than conventional AAM. Huang et al. adopted heterogeneous pulse coupled neural networks (HPCNN) for developing an image quantization algorithm where each neuron corresponds to a pixel. The parameters of the HPCNN model are estimated automatically according to different categories. Using mirror symmetry as a general purpose and biologically motivated prior, Michaux et al. proposed an approach to figure/ground organization. Based on the fact that the human visual system makes use of symmetry in producing 3D percepts of objects, they proposed a general-purpose

method for finding 3D symmetry correspondence by pairing the problem with the two-view geometry of the binocular correspondence problem. On a very industry-oriented problem, [Zheng and Wei](#) proposed an online automatic vision-based system for coupler yoke for freight trains. The achieved fault inspection rate and the average processing time of an image show high inspection accuracy and good real-time performance.

Decompositions are often the first step to analyzing visual data and are often based on some characteristics of the HVS. [Zaouali et al.](#) summarized the state-of-the-art of multiscale geometric decomposition (MGD) in a detailed review. The focus was put on studying the use of MGD in the remote sensing context. While addressing the computational problem, [Mesbah et al.](#) proposed a fast method for computing 3D moments Hahn by extending the notion of symmetry of Hahn polynomials, allowing reduction of the complexity of calculation by a factor of eight.

Several people deserve to be acknowledged for the success of this special section. We would like to thank the authors for submitting their high-quality papers, and the reviewers for dedicating their time and expertise to selecting the best papers for publication. Special thanks go to the Editor-in-Chief, Dr. Karen Egiazarian, and the editorial staff for their full support throughout the process of this special section.

Mohamed-Chaker Larabi received his PhD from the University of Poitiers in 2002. He is currently associate professor at the same university. His scientific interests deal with quality of experience and bio-inspired processing/coding/optimization of images and videos, 2D, 3D, and HDR. He is a member MPEG and JPEG committees. He served as the chair of the JPEG Advanced Image Coding (AIC) and the Test & Quality subgroup. He is a senior member of IEEE, and a member of the CIE and IS&T.

Sanghoon Lee received a BS in E.E. from Yonsei University in 1989 and an MS in E.E. from Korea Advanced Institute of Science and Technology in 1991. From 1991 to 1996, he worked for Korea Telecom. He received his PhD in E.E. from the University of Texas at Austin in 2000. From 1999 to 2002, he worked for Lucent Technologies on 3G wireless and multimedia networks. In March 2003, he joined the faculty of the Department of Electrical and Electronics Engineering, Yonsei University, where he is a full professor.

Mohammed El Hassouni received a PhD in image and video processing from the University of Burgundy in 2005. He joined Mohammed V University of Rabat as an assistant professor since 2006 and an associate professor since 2012. He also received the Habilitation from Mohammed V University in 2012. He was a visiting professor at several universities (Bordeaux, Orléans, Dijon, and Konstanz). He is member of IEEE, IEEE Signal Processing Society, and several conference program committees. He is also co-chair of QUAMUS workshop. His research focuses on image analysis, quality assessment, and mesh processing.

Frédéric Morain-Nicolier received an MS in applied physics in 1996 from Université de Bourgogne, France. From 1996 to 2001, he worked in Le2I lab at Le Creusot and received his PhD in 2000 from Université de Bourgogne. In 2001, he joined the CReSTIC lab of Université de Reims-Champagne-Ardenne and worked as associate professor in the Dept. of E.E. of Institut Universitaire de Technologie de Troyes. In 2010, he received his HDR and became full professor. His research interests include medical image processing, historical studies, image forensics, content-based image indexing, local and nonmetric similarities, and perceptual similarities.

Rachid Jennane is a full professor at the University of Orleans (France). He has been the principle investigator of several research projects. He spent the 1998 academic year as a visiting researcher at the EE Department of the University of Rhode Island (USA). He supervised more than 20 PhD and master's students in the area of signal and image processing. His current research interests include the processing of 2D/3D/nD medical images.