Visual Saliency Guided High Dynamic Range Image Compression

Tian Feng* , Charith Abhayaratne†
Department of Electronic and Electrical Engineering
The University of Sheffield
Sheffield, S1 3JD, United Kingdom
Email: *ashbringerft@gmail.com, †c.abhayaratne@sheffield.ac.uk

***Abstract*—Recent years have seen the emergence of the visual saliency-based image and video compression for low dynamic range (LDR) visual content. The high dynamic range (HDR) imaging is yet to follow such an approach for compression as the state-of-the-art visual saliency detection models are mainly concerned with LDR content. Although a few HDR saliency detection models have been proposed in the recent years, they lack the comprehensive validation. Current HDR image compression schemes do not differentiate salient and non-salient regions, which has been proved redundant in terms of the Human Visual System. In this paper, we propose a novel visual saliency guided layered compression scheme for HDR images. The proposed saliency detection model is robust and highly correlates with the ground truth saliency maps obtained from eye tracker. The results show a reduction of bit-rates up to 50% while retaining the same high visual quality in terms of HDR-Visual Difference Predictor (HDR-VDP) and the visual saliency-induced index for perceptual image quality assessment (VSI) metrics in the salient regions.**

I. Introduction

Various image/video processing applications employ saliency detection models to estimate regions of interest (ROIs). According to [1], [2], at least 70 visual saliency models were proposed, only 3 of them were proposed for high dynamic range (HDR) image content. The saliency detection algorithm proposed by Itti *et al.* [3] is the most widely used technique for low dynamic range (LDR) visual attention evaluation. The Itti *et al.* HDR image extension model fails to detect the median contrasts salient areas. The Contrast Features (CF) [4], Gao *et al.* [5], [6] and Dong [2] *et al.* models try to replace the color, intensity and orientation channels with more HDR adapted feature. However, they fail to validate on the HDR eye-tracking dataset.

HDR media records greater luminance range and wider color gamut than LDR media. Hence, the HDR contents are closer to the actual scene observed by Human Visual System (HVS). Thus, saliency detection of HDR content represents more meaningful attention regions of the natural scene. HDR applications could benefit from saliency analysis *e.g.,* tone-mapping [7] and compression [8]. Lossy HDR image compression JPEG-HDR [9], JPEG2000-HDR [10], JPEG-XT [11] and Mapping-free HDR image compression [12] have been proposed. These HDR image compression approaches fail to compress non-salient regions differently than salient regions because HVS perceives them differently. The works of [8], [13]–[15] suggested LDR content can be compressed by visual saliency *i.e.,* Region of Interest (ROI) in JPEG2000 [16]. By extending this concept to HDR content, the HDR visual saliency-guided compression can be realized.

In this paper, we propose a photographic visual saliency modeling for HDR images with application to compression. The saliency guided compression could reduce potentially 50% of the bit-rates cost. Since it is not optimal to compress the salient and non-salient regions with the same visual quality. The performance is evaluated by comparing the saliency map computed by computational models to the ground truth from eye fixation points acquired from eye tracker. The Pearson Correlation Coefficient (PCC) and Normalized Scanpath Saliency (NSS) quantitative metrics are chosen because they have the highest correlation to the human visual ground truth according to the latest MIT visual saliency benchmark [17], [18].

The main contributions of this paper are: A novel saliency estimation scheme for HDR images with high accuracy. The first saliency guided HDR image compression application is illustrated. The rest of this paper is organized as follows. Section 2 and section 3 present the proposed HDR saliency estimation mode and proposed saliency guided HDR image compression. Section 4 discussed performance evaluation and concluded in section 5.

II. Proposed HDR saliency estimation model

The proposed HDR image visual saliency model is based on the experiment that the CF is not enough to represent the full dynamic range of HDR content, and the salient regions of given HDR image after a stack of 3 brightness parameters tone-mapping operation should at least appeared once.

Fig. 1 shows the proposed model's architecture. The proposed HDR saliency map detection method contains 3 parts: Tone-mapping, saliency detection, and saliency map fusion. HDR input image is first tone mapped into 3 LDR images by stack of brightness parameters of tone-mapping operator: p_1, p_2, p_3, where p_1 is the default value of tone-mapping operator, and $p_2 = \frac{1}{8}p_1$, $p_3 = 8p_1$ respectively, it will be justified in the later section. Then, the tone mapped LDR images are fed into CF model for 3 saliency maps. The HDR saliency map of given HDR image is the weighted combination of different tone-mapped versions. The HDR saliency map is then segmented into different visual saliency

ISBN 978-90-827970-1-5 © EURASIP 2018

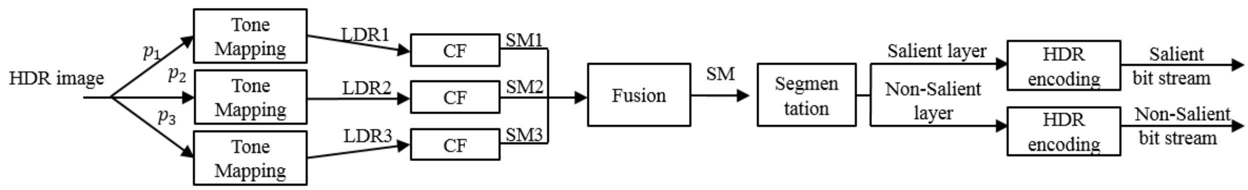


Fig. 1: The Architecture of the proposed model.

layers to guided the final compression of HDR image. The 3 exposure parameters set can be replaced by the gamma correction function with the similar performance.

A. *Saliency detection for HDR image*

The details of proposed visual saliency model description are followed:

1) Tone-mapping: The HDR image is tone-mapped into LDR image by tone-mapping operator [19], [20]:

$$I_{l_n} = T_{p,n}(I_h), \tag{1}$$

where I_l and I_h are LDR and HDR image respectively, the n is the corresponding picture index, T can be any kind of tone-mapping operators, p_n is the brightness parameter of the tone-mapping operator, p_1 is the default value of tone-mapping operator, and $p_2 = \frac{1}{8}p_1$, $p_3 = 8p_1$ respectively. The 3 exposures stack expands the 8 f-stops dynamic range of each tone mapped version to a total 14 f-stops collaboratively. Furthermore, the additional computational cost is minimized to just enough to cover the dynamic range of the given scene [2], [5].

2) Saliency detection: The saliency maps of 3 LDR images can be obtained by CF [4] model:

$$S_{l,n} = C(I_{l,n}), \tag{2}$$

where S is the saliency map, I_l is LDR image, C is the CF model. Fig. 2 shows the example of given HDR image and their tone mapped LDR version and CF saliency map results.

3) Saliency Fusion: Since proposed model tend to exploit visual saliency regions fully from 3 LDR image. The fusion step of proposed method is finished by take the highest saliency value for each pixel location among 3 saliency maps S see equation (4), which is according to our assumption that the salient regions of given HDR image after a stack of 3 brightness parameters tone-mapping operation should at least appeared once:

$$S_M(x,y) = max(S_n(x,y)). \tag{3}$$

where S_M is the final saliency map of input HDR image.

III. Proposed Saliency guided HDR image compression

Saliency based compressions [8] utilize saliency map as the ROI. The proposed method of compression application segments image into N layers by the density of the saliency map. In this paper, the input image is segmented into 2 layers to avoid the cost of overhead of too many layers. The

Fig. 2: (1) Tone mapped under exposure image and corresponding saliency map, (2) Tone mapped proper exposure image and corresponding saliency map, (3) Tone mapped over exposure image and corresponding saliency maps.

number of encoding layers is not limited by 2, it could be more than 2 depends on the requirement. The threshold (t) of segmentation is set to a given value from 0 to 1, when prediction visual saliency is lower than the given threshold (t), a lower compression quality setting is applied, when saliency is larger than the given threshold (t), a higher compression quality setting is applied. See Fig. 1.

The example of proposed encoding is illustrated by Fig. 3 Although the visual saliency threshold and compression quality are adjustable and independent, if the segmented salient region is greater than the half of the image size, the coding gain is constrained. Therefore, the top 30 salient regions are chosen for segmentation.

Fig. 4 shows the proposed model's decoding scheme. The proposed decoding process does not require the mask of the saliency map.

A. *Adapting to block based compressions*

All standard lossy HDR image compressions JPEG-HDR, JPEG2000-HDR, JPEG-XT, and Mapping-Free can be extended to saliency guided compression since the nature of the proposed scheme is to quantize the non-salient regions with higher distortion. However, the most of compressions are micro-block based methods, the segmentation requires further coarsening to match the micro-block size. Otherwise,

ISBN 978-90-827970-1-5 © EURASIP 2018

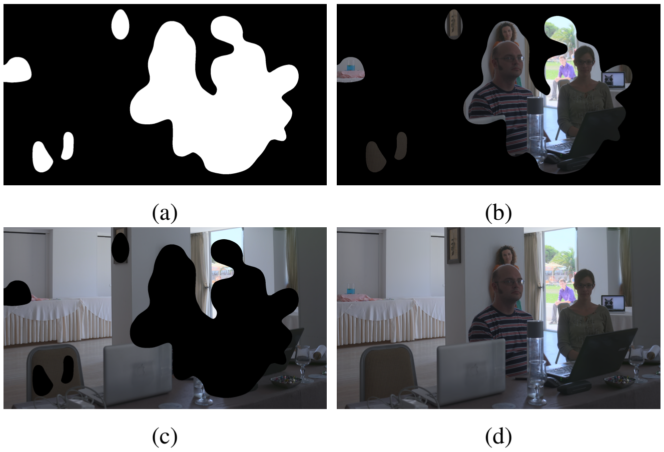


Fig. 3: (a) The saliency mask for segmentation, (b) Salient regions after segmentation, (c) Non-Salient regions after segmentation. (d) The summation of decoded Salient regions and Non-Salient regions

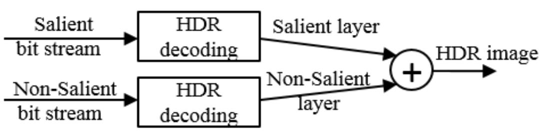


Fig. 4: The proposed model's decoding scheme.

the artifacts may occur among edges of segmentation after decoding.

IV. Performance Evaluation

The proposed saliency detection method is evaluated by comparing it to Itti *et al.* , Bremond (CF), Gao *et al.* and Dong *et al.* spatial model based on 2 quantitative metrics PCC and NSS. By comparing to the eye tracker dataset ground truth [21]. Reinhard *et al.* global [22] tone-mapping operator was chosen as an example TMO function since it provides the automatic parameters estimation.

To evaluate the performance of the proposed compression scheme, the HDR-Visual Difference Predictor (HDR-VDP) and the visual saliency-induced index for perceptual image quality assessment (VSI) are used. Since the traditional objectives image fidelity measurements fail in HDR field [23]–[26], and most of quality metrics fail to consider the visual saliency as measurement while VSI is capable [27]. Since the HDR-VDP did not include the visual saliency weighting, the same strategy is used of VSI for HDR-VDP, which is to weigh the quality by multiplying the saliency map pixel-wise.

A. Experiment Results and Discussion

In this section, experiment results are discussed.

1) Visual Saliency modelling: Fig. 5 (1) shows tone mapped HDR images of dataset No 14 and 33, the rest of Fig. 5 are the ground truth Fig. 5 (2) and corresponding saliency maps predicted by different models. The blue ball on the median left of the Fig.5 (1) (a) is detected as a salient object by Gao's Fig. 5 (5) (a) and the proposed methods Fig. 5 (7) (a) while other methods are less sensitive to low luminance area. The window on the center of the Fig. 5 (1) (b) is salient in the ground truth Fig. 5 (2) (b) and our prediction Fig. 5 (7) (b) while other predictions fail to detect.

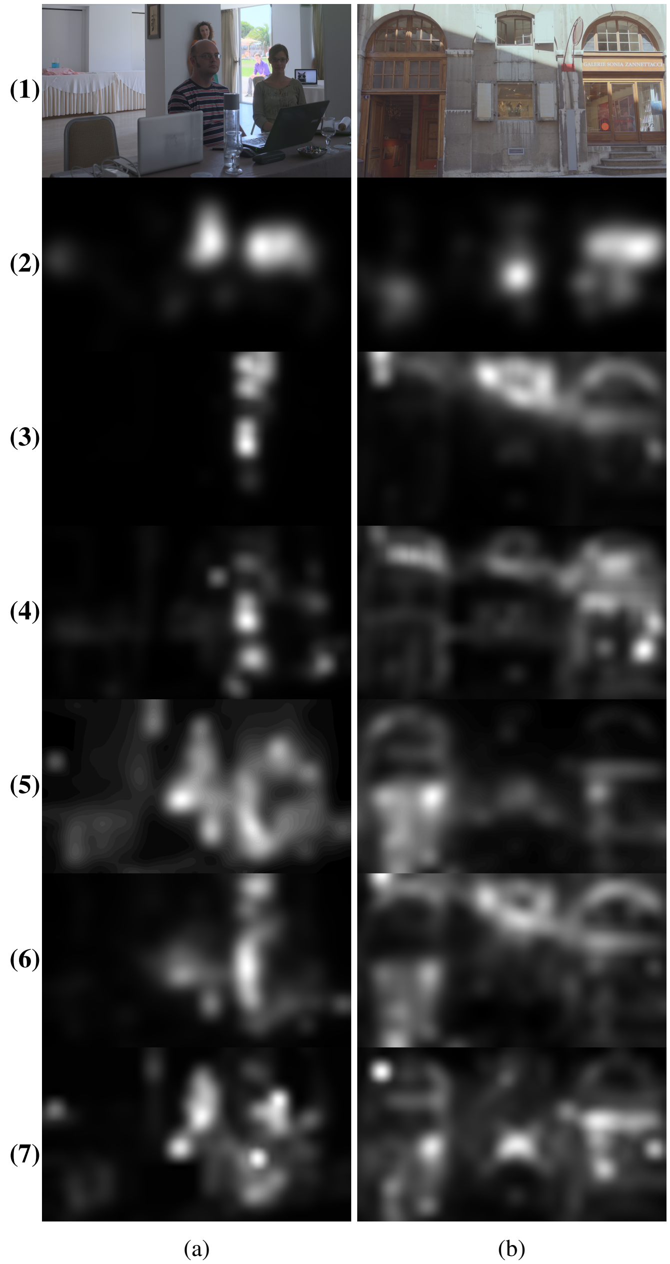
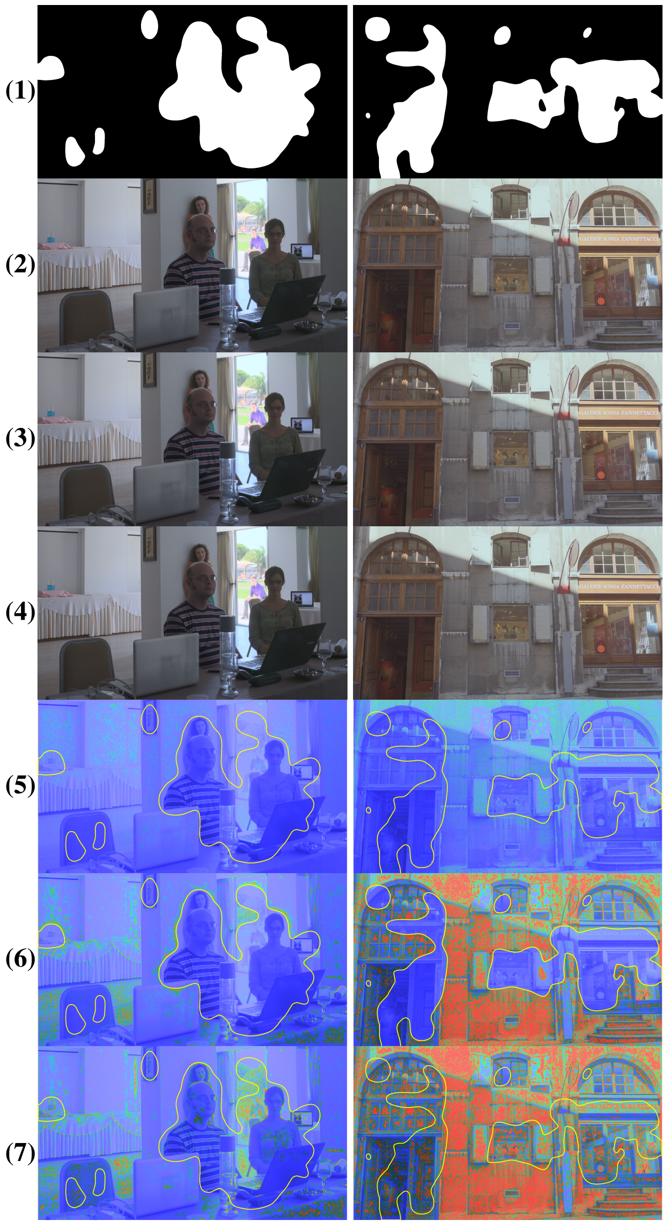


Fig. 5: Tone mapped image (1), Saliency maps of (2) ground truth, (3) Itti's, (4) CF, (5) Gao's, (6) Dong's and (7) proposed model from top to bottom. Column of left for No 14 image and right for No 33 image.

Table. I illustrated that the proposed method has the highest mean values for both of PCC and NSS metrics, which means our model is the closest to the ground truth from eye tracking data. The standard deviation of the proposed method is not the lowest, this is due to some salient regions appeared in

ISBN 978-90-827970-1-5



(a) (b)

Fig. 6: (1) Segmentation by threshold, (2) Decoded image of JPEG2K-HDR – the encoded file size are 607 KB, when compression ratio is 20, (3) Decoded image of Visual Saliency guided JPEG2K-HDR: the encoded file size are 363 KB, when compression ratio is 20 for salient regions, and 100 for non-salient regions, (4) Decoded image of JPEG2K-HDR: the encoded file size are 121 KB, when compression ratio is 100, (5) HDR-VDP map of decoded image of JPEG2K-HDR of (2), (6) HDR-VDP map of decoded image of Visual Saliency guided JPEG2K-HDR of (3), (7) HDR-VDP map of decoded image of JPEG2K-HDR for compression ratio 20.

the ground truth data are not detected across the prediction maps.

2) Application of HDR image compression: The Fig. 6 illustrate the proposed HDR image compression application guided by visual saliency map. The distortions of JPEG2K-HDR compression are measured by HDR-VDP metric in $(5)-(7)$ of Fig. 6. The visual saliency map guided JPEG2K-HDR compression obtained higher compression ratio while keeps same quality for salient regions which defined by the threshold. The color map used for HDR-VDP prediction maps showed in the Fig. 7 pixels in HDR-VDP closer to blue means lower distortion, and red means highest distortion.

TABLE I: The PCC and NSS of proposed features of Itti's model for HDR image calculated by 42 images [21].

Features		Itti [3]	CF [4]	Gao [5]	Dong [2]	Proposed
PCC	Mean	0.192	0.192	0.205	0.275	**0.312**
	Std	0.247	0.212	0.21	**0.165**	0.202
NSS	Mean	0.349	0.362	0.397	0.502	**0.579**
	Std	0.492	0.432	0.438	**0.376**	0.437

Fig. 7: The color map used for visualize HDR-VDP metric.

The proposed compression preserves good quality in salient regions at the expense of non-salient regions coded with a high quantization parameter. The HDR-VDP final result is a global result instead of detection rate per each pixel. It is not meant to show an RD curve with a single HDR-VDP value, therefore, two examples of the global and the salient regions HDR-VDP results and VSI results are displayed in Fig. 8. As the plots illustrated, the proposed compression obtained a higher compression quality with less bit-rate. The red and green curves are the normal compression with low and high-quality settings, and they belong to the same RD fitting curve. The blue and purple curves are the proposed method evaluated on only salient regions and full image. The proposed method obtained higher visual quality at same bit-rate compared to the normal compressions in both HDR-VDP and VSI metrics when evaluated with the visual saliency priority. A subjective comparison of the images compressed by the proposed method versus conventional methods could be conducted to further justify the improvements of the proposed method.

V. Conclusions

In this paper, a novel HDR saliency estimation model and the visual saliency guided HDR image compression scheme are proposed. The proposed HDR visual saliency prediction model has the best correlation to the ground truth from eye tracking fixations against existing methods, which leads to an efficient HDR image compression and evaluation. The proposed compression obtained a reduction of bit-rate up to 50% while retaining the same high visual quality in terms of HDR-VDP and VSI metric in the salient regions. The future work is to extend the proposed model to the HDR video, the saliency guided video compression could benefit more from visual attention model.

ISBN 978-90-827970-1-5

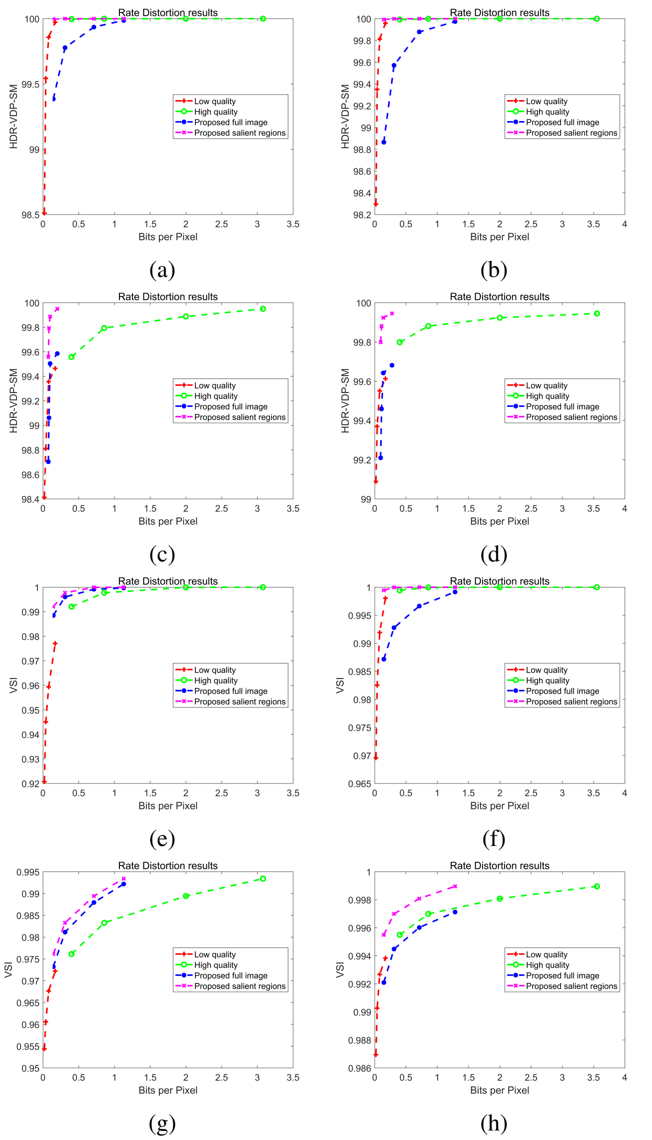


Fig. 8: RD results of saliency guided compression (a) and (b) HDR-VDP of JPEG2K-HDR, (c) and (d) HDR-VDP of JPEG-HDR, (e) and (f) VSI of JPEG2K-HDR, (g) and (h) VSI of JPEG-HDR. The results in the left column are for image 14, results in the right column are for image 33.

REFERENCES

[1] A. Borji and L. Itti, "State-of-the-art in visual attention modeling," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 35, no. 1, pp. 185–207, 2013.

[2] Y. Dong, M. T. Pourazad, and P. Nasiopoulos, "Human Visual System-based saliency detection for High Dynamic Range content," *IEEE Transactions on Multimedia*, vol. 18, no. 4, pp. 549–562, April 2016.

[3] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, no. 11, pp. 1254–1259, 1998.

[4] R. Brémond, J. Petit, and J.-P. Tarel, "Saliency maps of High Dynamic Range images," in *Trends and Topics in Computer Vision*. Springer, 2010, pp. 118–130.

[5] X. Gao, S. Brooks, and D. V. Arnold, "Virtual photograph based saliency analysis of High Dynamic Range images," in *Proceedings of the Symposium on Computational Aesthetics*. ACM, 2013, pp. 87–92.

[6] ——, "Automated parameter tuning for tone mapping using visual saliency," *Computers & Graphics*, vol. 52, pp. 171–180, 2015.

[7] W.-C. Lin and Z.-C. Yan, "Attention-based High Dynamic Range imaging," *The Visual Computer*, vol. 27, no. 6-8, pp. 717–727, 2011.

[8] C. Guo and L. Zhang, "A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression," *IEEE Transactions on Image Processing*, vol. 19, no. 1, pp. 185–198, Jan 2010.

[9] G. Ward and M. Simmons, "JPEG-HDR: A backwards-compatible, High Dynamic Range extension to JPEG," in *ACM SIGGRAPH 2006 Courses*. ACM, 2006, p. 3.

[10] R. Xu, S. N. Pattanaik, and C. E. Hughes, "High-Dynamic-Range Still-Image Encoding in JPEG 2000," *IEEE Comput. Graph. Appl.*, vol. 25, no. 6, pp. 57–64, Nov. 2005. [Online]. Available: http://dx.doi.org/10.1109/MCG.2005.133

[11] A. Artusi, R. Mantiuk, T. Richter, P. Korshunov, P. Hanhart, T. Ebrahimi, and M. Agostinelli, "JPEG XT: A Compression Standard for HDR and WCG Images [Standards in a Nutshell]," *IEEE Signal Processing Magazine*, vol. 33, no. 2, pp. 118–124, 2016.

[12] T. Feng, S. Pal, and C. Abhayaratne, "Mapping-free high dynamic range image compression," in *Proceedings of the 2015 IEEE 5th International Conference on Consumer Electronics-Berlin (ICCE-Berlin)*. IEEE, 2015, pp. 172–175.

[13] N. Ouerhani, J. Bracamonte, H. Hugli, M. Ansorge, and F. Pellandini, "Adaptive color image compression based on visual attention," in *2001. Proceedings. 11th International Conference on Image Analysis and Processing*. IEEE, 2001, pp. 416–421.

[14] L. Itti, "Automatic foveation for video compression using a neurobiological model of visual attention," *IEEE Transactions on Image Processing*, vol. 13, no. 10, pp. 1304–1318, 2004.

[15] H. Hadizadeh and I. V. Bajić, "Saliency-aware video compression," *IEEE Transactions on Image Processing*, vol. 23, no. 1, pp. 19–33, 2014.

[16] L. Liu and G. Fan, "A new JPEG2000 region-of-interest image coding method: Partial significant bitplanes shift," *IEEE Signal Processing Letters*, vol. 10, no. 2, pp. 35–38, 2003.

[17] Z. Bylinskii, T. Judd, A. Borji, L. Itti, F. Durand, A. Oliva, and A. Torralba, "Mit saliency benchmark," http://saliency.mit.edu/.

[18] Z. Bylinskii, T. Judd, A. Oliva, A. Torralba, and F. Durand, "What do different evaluation metrics tell us about saliency models?" *arXiv preprint arXiv:1604.03605*, 2016.

[19] P. Debevec, "Recovering High Dynamic Range radiance maps from photographs," in *Proc. SIGGRAPH'97*, 1997.

[20] E. Reinhard, "Parameter estimation for photographic tone reproduction," *Journal of Graphics Tools*, vol. 7, no. 1, pp. 45–51, 2002.

[21] H. Nemoto, P. Korshunov, P. Hanhart, and T. Ebrahimi, "Visual attention in LDR and HDR images," in *9th International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM)*, 2015.

[22] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic tone reproduction for digital images," *ACM Transactions on Graphics (TOG)*, vol. 21, no. 3, pp. 267–276, 2002.

[23] R. Mantiuk, K. J. Kim, A. G. Rempel, and W. Heidrich, "HDR-VDP-2: a calibrated visual metric for visibility and quality predictions in all luminance conditions," in *ACM Transactions on Graphics (TOG)*, vol. 30, no. 4. ACM, 2011, p. 40.

[24] P. Hanhart, M. V. Bernardo, P. Korshunov, M. Pereira, A. M. Pinheiro, and T. Ebrahimi, "HDR image compression: a new challenge for objective quality metrics," in *2014 Sixth International Workshop on Quality of Multimedia Experience (QoMEX)*. IEEE, 2014, pp. 159–164.

[25] M. Azimi, A. Banitalebi-Dehkordi, Y. Dong, M. T. Pourazad, and P. Nasiopoulos, "Evaluating the performance of existing full-reference quality metrics on High Dynamic Range video content," in *International Conference on Multimedia Signal Processing (ICMSP)*, 2014.

[26] P. Hanhart, M. V. Bernardo, M. Pereira, A. M. Pinheiro, and T. Ebrahimi, "Benchmarking of objective quality metrics for HDR image quality assessment," *EURASIP Journal on Image and Video Processing*, vol. 2015, no. 1, p. 1, 2015.

[27] L. Zhang, Y. Shen, and H. Li, "Vsi: A visual saliency-induced index for perceptual image quality assessment," *IEEE Transactions on Image Processing*, vol. 23, no. 10, pp. 4270–4281, Oct 2014.