A machine-learning framework to predict TMO preference based on image and visual attention features

Waqas Ellahi, Toinon Vigier, Patrick Le Callet

A machine-learning framework to predict TMO preference based on image and visual attention features

Waqas Ellahi, Toinon Vigier, Patrick Le Callet

A machine-learning framework to predict TMO preference based on image and visual attention features

Waqas Ellahi, Toinon Vigier, Patrick Le Callet
LS2N UMR CNRS 6003, Université de Nantes, Nantes, France
firstname.name@univ-nantes.fr

***Abstract*—Tone-mapping operator (TMO) plays a crucial role in the task towards displaying high dynamic range (HDR) contents on standard displays. Similarly, visual attention (VA) is a vital feature of the human visual system (HVS), while it has not been sufficiently i nvestigated i n p reference o f tone-mapped images. The potential benefits o f v isual a ttention-based features for quality assessment of tone-mapped images are studied in this paper. A novel framework is proposed for tone-mapped image quality assessment to predict image preference. The framework is evaluated on two different datasets. Experimental results illustrate the importance of visual attention for improving the performance of objective metrics. The proposed framework outperforms the existing methods and presents as a competitive alternative for tone-mapped images evaluation.**

***Index Terms*—Image quality, visual attention, machine learning framework, tone mapping operators, objective quality metrics**

I. Introduction and Related works

Recent advancements in multimedia applications aim at providing new immersive and realistic experiences to the end user by achieving depth perception (e.g. 3D TV), visual interaction (e.g. free viewpoint content), better resolution (Ultra High-Definition TV) o r m ore c ontrasts (High Dynamic Range and Wide Color Gamut technologies). All these new imaging technologies bring new challenges for the Quality of Experience community to reach the best visual quality at the end of the processing chain.

The main principle of High Dynamic Range (HDR) imaging technologies is to capture the accurate luminance values as much as possible from the real-world scene and reproduce them with the best quality on the display. HDR content can be rendered on dedicated hardware with a large dynamic range which are, unfortunately, expensive and not commonly available in the market. The display of HDR content on consumer-grade screens requires thus the compression of the dynamic range thanks to tone-mapping operators (TMOs). In the past years, numerous tone mapping algorithms have been proposed [1] that often require the selection of specific parameters considering the scene features [2]. A manual selection of the best adequate TMO and the ad-hoc parameters is not possible in many applications and the development of objective metrics to predict the quality of the tone-mapped images is still challenging.

Because TMO can impact the image quality differently (contrast distortions, color shifts, halos, etc.), different objective metrics have been proposed focusing on different aspects of the visual quality as contrast preservation based on human visual system (HVS) modeling [3], [4], image structural similarity [4], naturalness, colourfulness [5], phase information [6], aesthetic [7].

More recently, two different papers have proposed new approaches based on features fusion for the quality assessment of tone-mapped images. Hadizadeh and Bajic [8] developed a "bag of features" method based on 8 features to predict the perceived quality using a support vector regression model. One interest of this method is the consideration of nonlinearities and the facility to adapt the list of the used features. Krasula et al. [9] applied a selection algorithm on 60 features to identify the most relevant ones for the quality evaluation of TMOs. From that, a new objective metric, FFTMI, is trained as a linear combination of the 5 most relevant features.

In this paper, we propose to extend these two works by applying a machine learning framework using both image features and (VA) features to predict TMO preference. Firstly, this paper presents the prediction of the preference for tone-mapped images instead of a quality score. On contrast to the traditional method (MOS), Pairwise Comparison (PC) methodology reduces the subject uncertainty and provides more reliable subjective preferences [10]. Whereas, most of the image quality objective metrics predict a quality score (or mean opinion score), a few works have proposed methods to predict preference by using machine learning methods [11], [12], inspired by the development of rank learning in information retrieval [13]. These works focus on image distortions as compression, blur, noise, etc. and the application of this method on TMO content should be studied.

Secondly our paper investigates the interest of VA features for the prediction of the preference label. The influence of TMO on visual behavior has already been studied showing specific effects on visual saliency [14] and scanpath [15]. The use of VA for improving quality assessment is not new but it is mainly based on the weight of the visible distortion from saliency information [16], [17]. In our study, we propose to use VA as a proxy of difference in quality between two tone-mapped images.

The rest of the paper is organized as follows. The proposed

framework with the detail of the used features is explained in the next section. The evaluation of the predictive model and the contribution of the different features is presented in Section III. Finally, the concluding remarks are provided in Section IV.

II. Proposed Machine-Learning Framework

In this section, we present the used image-based and VA-based features for the prediction of TMO preference. The proposed framework is based on a rank support vector machine (SVM) model and the details are as follows.

A. Image features

In order to reduce the number of studied features, only the ones identified as the most relevant features for TMO perceived quality in [9] are used. We briefly introduce each image feature here (details can be found in[9]).

Structural Similarity (SS)
The structural similarity feature is an adapted version of the SSIM index for the comparison of the structure of HDR and LDR images [4]. SSIM is modified by removing the comparison of luminance component and adapting the comparison between signal strength considering contrast sensitivity functions.

Feature Naturalness (FN)
This measure is developed based on the assumption that naturalness is mainly defined by the content's contrast, brightness, and colorfulness [2]. This feature is directly computed on the TMO content without reference to the HDR. FN is computed as a function of the product of three estimators, i.e. the global contrast factor [18], the mean intensity (MI), and the CQE1 colorfulness [19].

Feature Similarity (FS)
The Feature Similarity Index for Tone-Mapped Images (FSITM) compares the HDR and the tone-mapped images considering the phase congruency features based on the Locally Weighted Mean Phase Angle [6]. FSITM is computed per channel R, G and B leading to a set of three values.

B. Visual attention features

In this work, we focus on VA features aiming to compare visual behavior in LDR content tone-mapped with two different algorithms. The features can be separated in two main groups: the VA similarity metrics and features based on VA complexity.

In order to assess the interest to use VA features for the prediction of TMO preference, the proposed features in this paper are estimated based on real gaze data. We latter discuss how the most performing features can be estimated using computational methods.

1) Visual attention similarity: As it has been shown that TMOs can impact both visual saliency and visual behavior in images, different similarity metrics based either on the comparison of saliency maps [14], either on the comparison of scanpaths in two tone-mapped images [15] are used.

Pearson's Correlation Coefficient (CC)
CC is computed as the 2D linear correlation between the saliency map of the two tone-mapped images.

Kullback-Liebler divergence (KLD)
KLD measures the dissimilarity between the two normalized saliency maps seen as two 2D probability distributions.

Normalized Scanpath Saliency (NSS)
NSS determines a similarity score by computing the mean of normalized saliency validated at fixated points. This metric uses the saliency map of one tone-mapped image and the set of the fixation points for the second LDR conditions.

Fixation clusters based Hidden Markov Model (FCHMM)
This metric is based on a HMM-based framework to measure the similarity of visual behavior between two conditions of experiment [15]. This method uses a HMM to model the sequence of eye movements of one observer [20], [21]. The HMMs of each observer for one condition, i.e. one tone-mapped image, are then combined into a joint-HMM using a variational hierarchical expectation maximization algorithm [22], [23]. Finally, the similarity between two conditions is computed by considering likelihood between the joint HMM modeled based on eye data of condition 1 and the eye fixations of condition 2. Here, for each pair of TMO, the method is applied to each TMO as condition 1 and the mean value of the two likelihood scores is then calculated as the final similarity score.

2) Visual attention complexity: Visual attention complexity (VAC) reflects how a visual scene catches observers' attention and should be used to differentiate focused or exploratory contents. VAC is a proxy for the variation in observer fixations on a visual scene. The influence of TMOs on VA can also lead to impact VAC of a referent HDR scene. The possible difference of VAC in two tone-mapped images of the same HDR source can be used to speculate a difference in preference.

Different measurement methods have been proposed to estimate VAC based on the entropy of the saliency map [24] or inter-observer congruency (IOC), also named inter-observer agreement [25].

Entropy (ENT)
The entropy is directly computed on the saliency map as in [24] using $P_{max} = 1$.

Inter-observer congruency (IOC)
IOC is computed by comparing the behavior of one observer compared to the other observers. In most cases, a leave-one-out approach combined with the NSS similarity metric is used [25]. In this paper, in order to reduce the computational complexity, NSS is directly computed between a single observer eye data and the saliency maps of all observers. IOC is then estimated as the average of all observers' similarity scores.

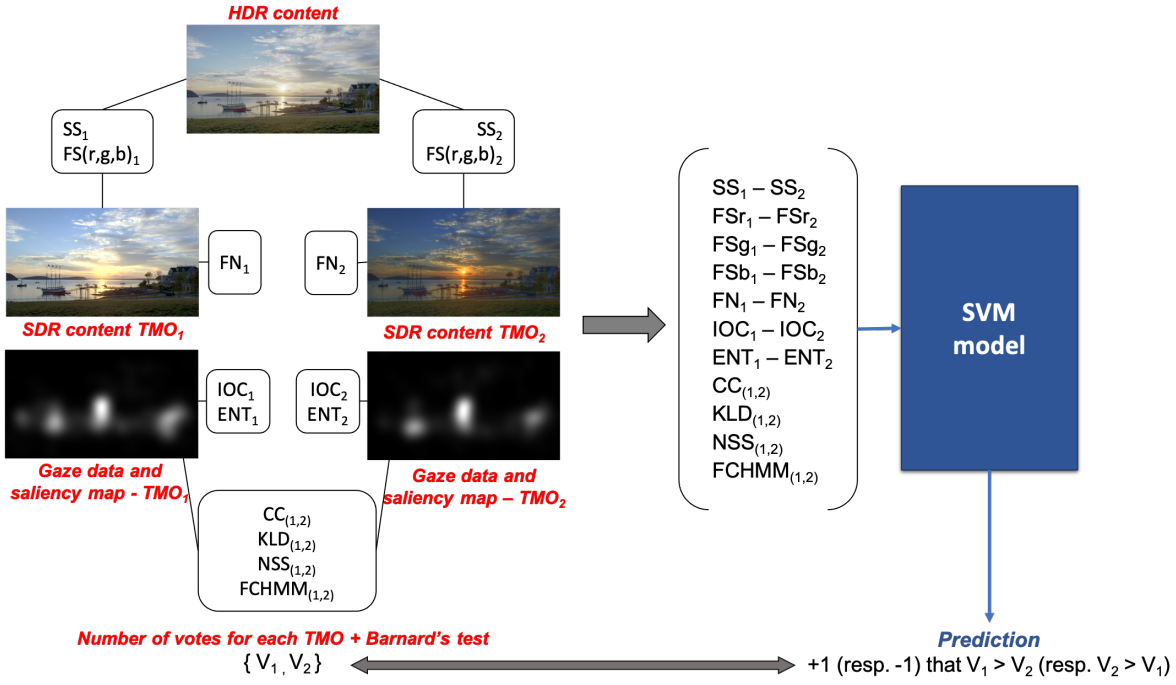


Fig. 1: Tone-mapped image preference prediction framework

C. Framework

In this section, we will present the complete machine-learning based framework proposed to predict TMO preference as an illustrated in in Figure 1.

For each pair of tone-mapped images $P_{i,j}$ from one HDR reference source, we have a set of image and VA based features $F_{i,j} = \{F^I_{i,j}, F^{VA}_{i,j}\}$.
F^I contains five features computed as the difference of SS, FN and FS_{rgb} between the two tone-mapped images:

$$F^I_{i,j} = \{F^I_i\} - \{F^I_j\} \text{ where } I = \{SS, FN, FR_r, FS_g, FS_b\}$$

F^{VA} contains the VA features divided in two categories F^{VAS} for features measuring the VA similarity between the two stimuli and F^{VAC} for features measuring the difference of VAC between the two tone-mapped images.

$$F^{VAS}_{i,j} = \{F^{VAS}_{(i,j)}\} \text{ where } VAS = \{CC, KLD, NSS, FCHMM\}$$

$$F^{VAC}_{i,j} = \{F^{VAC}_i\} - \{F^{VAC}_j\} \text{ where } VAC = \{IOC, ENT\}$$

The objective of our framework is to predict, from a pair of tone-mapped images $P_{i,j}$ represented by the features $F_{i,j}$, which image (I_i or I_j) is preferred. This goal can be expressed as a classification problem where the algorithm returns $+1$ (respectively -1) where I_i (resp. I_j) is preferred to I_j (resp. I_i) (in the following of the paper it will be written as $I_i > I_j$). In the literature, it has been identified as a pairwise learning-to-rank problem for which different machine-learning algorithms have been developed [11], [12].

In this work, we employ a SVM to train the predicted model on the vector of features $F_{i,j}$ along with its corresponding preference label $\{-1, +1\}$. SVM has been chosen here because the small amount of data. In our approach, SVM is used with a radial basis kernel function. After training the SVM, given any test pair image feature vector as input to the trained model, a preference label can be predicted.

III. Experiments

In this section, we present the application of the proposed framework on two datasets and the obtained results.

A. Datasets

Two existing datasets with tone-mapped images and preference scores are used to evaluate our framework. They have been extended with gaze data collected in in-lab experiments to compute VA features. The main characteristics of these datasets are presented in Table I.

TABLE I: Characteristics of the two datasets (ET means Eye Tracking).

Features	Exp-TMO	PairTMO
# of SCR	20	10
# of HRC	4	9
Resolution	640x480	1920x1080
# of observers in voting Exp.	40	20
# of Significant pairs	92 (57%)	204 (65%)
# of observers in ET Exp. per image	9-19	27-28
Viewing duration in ET Exp. (sec)	5	
Viewing distance	6.75 H	3 H
# of Pairs	120	360

TABLE II: Performance of the framework for different input features on each dataset.

	Exp-TMO			PairTMO		
Input features	**Accuracy**	**Precision**	**Recall**	**Accuracy**	**Precision**	**Recall**
$\boldsymbol{F^I_{SS}}$	56.43	62.72	62.18	63.68	73.60	69.14
$\boldsymbol{F^I_{FN}}$	68.47	71.23	77.63	64.68	77.36	64.67
$\boldsymbol{F^I_{FS}}$	64.09	80.39	52.72	67.63	77.71	70.71
$\boldsymbol{F^I}$	73.97	86.00	68.18	68.59	77.61	72.93
$\boldsymbol{F^I + F^{VAS}_{KLD}}$	77.01	79.71	84.91	73.94	75.64	85.99
$\boldsymbol{F^I + F^{VAS}_{CC}}$	78.18	80.95	85.09	74.57	75.34	86.46
$\boldsymbol{F^I + F^{VAS}_{NSS}}$	80.52	89.03	75.81	76.06	78.67	86.95
$\boldsymbol{F^I + F^{VAS}_{FCHMM}}$	75.16	87.60	68.36	72.07	79.57	77.46
$\boldsymbol{F^I + F^{VAC}_{ENT}}$	75.59	84.60	72.18	70.62	78.55	75.92
$\boldsymbol{F^I + F^{VAC}_{IOC}}$	81.57	89.06	77.63	78.43	80.34	88.71

1) Exp-TMO Dataset: The Exp-TMO dataset has been recently published in [26]. This dataset contains originally 20 HDR sources processed with 4 different TMOs (see as Hypothetical Reference Circuits HRC) leading to a total of 120 pairs of tone mapped stimuli used in our experiment. All the pairs were rated by 40 observers. Eye tracking data were recorded with the Tobii Pro Fusion eye tracker with a frequency of 120 Hz. The stimuli were presented to participants during five seconds. 55 different observers took part to the complete experiment but, in order to avoid repetitions, only the first visualization of each source content was used. Thus, each tone-mapped stimuli have been viewed by 9 to 17 observers.

2) PairTMO Dataset: The PairTMO dataset has been published in [27]. This dataset contains originally 10 HDR sources processed with 9 different TMOs (see as HRC) leading to a total of 360 pairs of tone mapped stimuli used in our experiment. As explained in [27], an adaptive square design experimental plan was used with 20 observers leading to a number of votes per pairs between 2 and 19. The number of votes are in the range of 5 to 16 for maximum number of pairs, only one pair for 2 and 19 votes. Eye tracking data were recorded with the Tobii Pro Fusion eye tracker with a frequency of 120 Hz. The stimuli were presented to participants during five seconds. One experimental session for one observer was build considering half of the complete dataset to avoid too much repetitions of the same source processed with different TMOs. 55 observers took part in the experiment and each tone-mapped stimuli were seen by 27 or 28 different observers.

Raw gaze data from the two datasets were clustered in saccades and fixations using EyeMMV algorithm [28]. The final saliency maps were calculated considering a Gaussian filter of sigma equals to one visual degree (i.e 57 pixels).

B. Training methodology

The proposed framework was applied on significant pairs only. Indeed, the non-statistical significance between two tone-mapped stimuli can be due either due to a non-agreement between participants, either to a small number of observers. The question of the prediction of significant pairs is not raised in this paper but will be furthered studied by the authors. The significant pairs of each dataset were determined with the Barnard's test. 92 significant pairs were identified for Exp-TMO dataset and 204 ones for PairTMO.

Finally, the framework was assessed on each dataset separately considering different combinations of features. For each dataset, a five-fold cross validation with a division of 80% data for training and 20% for testing was used.

C. Results and Discussion

In this part, we present and discuss the results of the framework considering different combinations of features. The interest of using VA features, mainly IOC, for TMO preference prediction is examined and the final results are compared with the state-of-the-art FFTMI metric [9].

1) Comparison of features' performance: The performance of the framework with different combination of input features is presented in Table II for both Exp-TMO and PairTMO datasets. The influence of each image features is presented. Results show that combination of the 5 image features leads to a better performance than each feature separately but with a moderate accuracy (around 70% for the two datasets) still. Then the interest of adding VA information is presented per feature. The combination of different VA features is not presented here because it did not improve the performance. Results show that saliency-based VA similarity features and IOC improves the performance of the framework for the two datasets. The best performance is achieved considering IOC with an increased accuracy of 7.6% for Exp-TMO and 9.84% for PairTMO. Finally, we achieve with the best features an accuracy around 80%.

Even if only significant pairs have been tested here, we investigate the influence of the votes' distribution on the classification performance assuming that more the observers agree that a tone-mapped image is better than another, more accurate is the model. To test this hypothesis, we plot in Figure 2 the number of classified and misclassified pairs across the percentage of votes obtained by the preferred stimuli of

TABLE III: Performance comparison with FFTMI.

	Exp-TMO			PairTMO		
Input features	**Accuracy**	**Precision**	**Recall**	**Accuracy**	**Precision**	**Recall**
$\boldsymbol{FFMTI}$	55.08	63.64	52.36	58.78	77.35	51.85
$\boldsymbol{F^I}$	73.97	86.00	68.18	68.59	77.61	72.93
$\boldsymbol{F^I + F_{IOC}^{VAC}}$	81.57	89.06	77.63	78.43	80.34	88.71

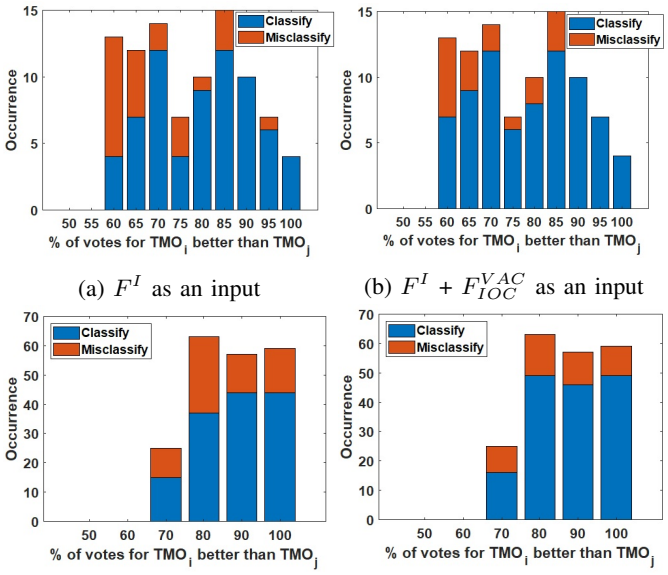
each pair. We can see that with and without VA features, the number of misclassified pairs is reduced from 90% of votes for the same image. This result is more pronounced for ExpTMO dataset. Additionally, the number of misclassified pairs is reduced for all bins when IOC is used as an input feature. For the ExpTMO, we can even observe the absence of misclassified pairs with IOC feature when the percentage of votes for the best TMO is above 90%.

2) Comparison with FFTMI: The proposed method is compared with the recently published tone mapped image preference objective metric FFTMI [9]. This metric uses the same image features as our method but in a linear combination in order to produce a quality score. In order to predict preference for a pair Pi, j with FFTMI, the scores obtained for stimuli i and stimuli j are computed and the image with the higher score is considered as the preferred one. Results, presented in Table III, show that our framework outperforms FFTMI metric here. The accuracy jumps from around 55% and 58% to 73% and 68% for Exp-TMO and PairTMO dataset respectively when non linearity is introduced using SVM with FFTMI features set. However, results should be considered with prudence because the parameters of the FFTMI metric have been learned on another dataset whereas our framework has been trained and tested on same datasets.

3) General discussion and future works: The obtained results are very promising about the use of a SVM model with image and VA based features for the prediction of TMO preference, achieving an average performance in accuracy around 80%. Moreover, the model outperforms the state-of-the-art FFTMI objective metric. However, some future works are still required to confirm these results and develop a robust objective metric for TMO prediction without ground-truth eye data.

First, the model is used with the 5 image features considered as the best ones to predict tone-mapped image quality in [9]. Image characteristics and aesthetics can play different role when observers have to assess overall visual quality or vote for the preferred image. The influence of other features as the ones used in [8] could be tested. The interest of using IOC for TMO preference prediction is clearly demonstrated in this paper. This result could be deeper studied in order to better interpret what is the role of IOC in image preference. This work should be extended with the use of computational IOC values in order to provide an objective metric (without the need of eye tracking data) for the prediction of preference between two tone-mapped images [25], [29].

In this work, the SVM model is trained and tested on the same datasets. Thus, the influence of content characteristics as resolution is limited. The robustness of the method for cross-dataset preference prediction should be studied. Moreover, the work is focused on significant pairs from the subjective experiment only and the question of predicting if a pair is significantly different has not been raised. Finally, a SVM model has been selected here because of a small amount of data. A deep-learning method could raise better results if the size of the dataset can be increased for the training phase.



(a) F^I as an input (b) F^I + F_{IOC}^{VAC} as an input

(c) F^I as an input (d) F^I + F_{IOC}^{VAC} as an input

Fig. 2: Plots for classifier performance across percentage of difference of votes. Top row for Exp-TMO data and bottom row for PairTMO dataset.

IV. Conclusion

In this paper, we have proposed a machine learning based framework for preference prediction between pair of tone mapped images. By combining image features with VA ones, we have also investigated the interest of using VA in the study of tone mapped image quality assessment. Results show that the use of a machine learning method and the integration of inter observer congruency in visual salience leads to a good accuracy. In order to develop a new objective metric usable on any dataset, this work will be further extended with the use of a computational method to estimate visual inter

observer agreement and the framework will be assessed on other datasets.

ACKNOWLEDGMENT

The work in this paper was funded from the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No 765911, European Training Network on Real Vision project.

REFERENCES

[1] J. Petit and R. K. Mantiuk, "Assessment of video tone-mapping: Are cameras' s-shaped tone-curves good enough?" *Journal of Visual Communication and Image Representation*, vol. 24, no. 7, pp. 1020–1030, 2013.

[2] L. Krasula, M. Narwaria, K. Fliegel, and P. Le Callet, "Rendering of hdr content on ldr displays: An objective approach," in *Applications of Digital Image Processing XXXVIII*, vol. 9599. International Society for Optics and Photonics, 2015, p. 95990X.

[3] T. O. Aydin, R. Mantiuk, K. Myszkowski, and H.-P. Seidel, "Dynamic range independent image quality assessment," *ACM Transactions on Graphics (TOG)*, vol. 27, no. 3, pp. 1–10, 2008.

[4] H. Yeganeh and Z. Wang, "Objective quality assessment of tone-mapped images," *IEEE Transactions on Image Processing*, vol. 22, no. 2, pp. 657–667, 2012.

[5] L. Krasula, M. Narwaria, K. Fliegel, and P. Le Callet, "Preference of experience in image tone-mapping: Dataset and framework for objective measures comparison," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 1, pp. 64–74, 2016.

[6] H. Z. Nafchi, A. Shahkolaei, R. F. Moghaddam, and M. Cheriet, "Fsitm: A feature similarity index for tone-mapped images," *IEEE Signal Processing Letters*, vol. 22, no. 8, pp. 1026–1029, 2014.

[7] M. Narwaria, M. P. Da Silva, P. Le Callet, and R. Pepion, "Tone mapping based hdr compression: Does it affect visual experience?" *Signal Processing: Image Communication*, vol. 29, no. 2, pp. 257–273, 2014.

[8] H. Hadizadeh and I. V. Bajić, "Full-reference objective quality assessment of tone-mapped images," *IEEE Transactions on Multimedia*, vol. 20, no. 2, pp. 392–404, 2017.

[9] L. Krasula, K. Fliegel, and P. Le Callet, "Fftmi: Features fusion for natural tone-mapped images quality evaluation," *IEEE Transactions on Multimedia*, vol. 22, no. 8, pp. 2038–2047, 2019.

[10] E. Zerman, V. Hulusic, G. Valenzise, R. K. Mantiuk, and F. Dufaux, "The relation between mos and pairwise comparisons and the importance of cross-content comparisons," *Electronic Imaging*, vol. 2018, no. 14, pp. 1–6, 2018.

[11] F. Gao, D. Tao, X. Gao, and X. Li, "Learning to rank for blind image quality assessment," *IEEE transactions on neural networks and learning systems*, vol. 26, no. 10, pp. 2275–2290, 2015.

[12] K. Ma, W. Liu, T. Liu, Z. Wang, and D. Tao, "dipiq: Blind image quality assessment by learning-to-rank discriminable image pairs," *IEEE Transactions on Image Processing*, vol. 26, no. 8, pp. 3951–3964, 2017.

[13] Z. Fu, X. Sun, Q. Liu, L. Zhou, and J. Shu, "Achieving efficient cloud search services: multi-keyword ranked search over encrypted cloud data supporting parallel computing," *IEICE Transactions on Communications*, vol. 98, no. 1, pp. 190–200, 2015.

[14] M. Narwaria, M. P. Da Silva, P. Le Callet, and R. Pépion, "Effect of tone mapping operators on visual attention deployment," in *Applications of Digital Image Processing XXXV*, vol. 8499. International Society for Optics and Photonics, 2012, p. 84990G.

[15] W. Ellahi, T. Vigier, and P. Le Callet, "Hmm-based framework to measure the visual fidelity of tone mapping operators," in *2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. IEEE, 2020, pp. 1–6.

[16] G. Zhai and X. Min, "Perceptual image quality assessment: a survey," *Science China Information Sciences*, vol. 63, pp. 1–52, 2020.

[17] P. Le Callet and E. Niebur, "Visual attention and applications in multimedia technologies," *Proceedings of the IEEE*, vol. 101, no. 9, pp. 2058–2067, 2013.

[18] K. Matkovic, L. Neumann, A. Neumann, T. Psik, and W. Purgathofer, "Global contrast factor-a new approach to image contrast." *Computational Aesthetics*, vol. 2005, no. 159-168, p. 1, 2005.

[19] K. Panetta, C. Gao, and S. Agaian, "No reference color image contrast and quality measures," *IEEE transactions on Consumer Electronics*, vol. 59, no. 3, pp. 643–651, 2013.

[20] T. Chuk, A. B. Chan, and J. H. Hsiao, "Understanding eye movements in face recognition using hidden markov models," *Journal of vision*, vol. 14, no. 11, pp. 8–8, 2014.

[21] C. A. McGrory and D. Titterington, "Variational bayesian analysis for hidden markov models," *Australian & New Zealand Journal of Statistics*, vol. 51, no. 2, pp. 227–244, 2009.

[22] E. Coviello, G. R. Lanckriet, and A. B. Chan, "The variational hierarchical em algorithm for clustering hidden markov models," in *Advances in neural information processing systems*, 2012, pp. 404–412.

[23] E. Coviello, A. B. Chan, and G. R. Lanckriet, "Clustering hidden markov models with variational hem," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 697–747, 2014.

[24] W. Zhang, R. R. Martin, and H. Liu, "A saliency dispersion measure for improving saliency-based image quality metrics," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 6, pp. 1462–1466, 2017.

[25] O. Le Meur, T. Baccino, and A. Roumy, "Prediction of the inter-observer visual congruency (iovc) and application to image ranking," in *Proceedings of the 19th ACM international conference on Multimedia*, 2011, pp. 373–382.

[26] A. Ak, M. Abid, M. P. da Silva, and P. Le Callet, "On spammer detection in crowdsourcing pairwise comparison tasks: Case study on two multimedia qoe assessment scenarios." in *ICME 2021-First International Workshop on Quality of Experience in Interactive Multimedia*, 2021.

[27] L. Krasula, M. Narwaria, K. Fliegel, and P. Le Callet, "Influence of hdr reference on observers preference in tone-mapped images evaluation," in *2015 Seventh International Workshop on Quality of Multimedia Experience (QoMEX)*. IEEE, 2015, pp. 1–6.

[28] V. Krassanakis, V. Filippakopoulou, and B. Nakos, "Eyemmv toolbox: An eye movement post-analysis tool based on a two-step spatial dispersion threshold for fixation identification," *Journal of Eye Movement Research*, vol. 7, no. 1, 2014.

[29] A. Bruckert, Y. H. Lam, M. Christie, and L. Olivier, "Deep learning for inter-observer congruency prediction," in *2019 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2019, pp. 3766–3770.