OBJECTIVE QUALITY ASSESSMENT OF TONE-MAPPED VIDEOS

Hojatollah Yeganeh, Shiqi Wang, Kai Zeng, Mahzar Eisapour and Zhou Wang

Dept. of Electrical & Computer Engineering, University of Waterloo, Waterloo, ON, Canada
Email: {hyeganeh, s269wang, kzeng, zhou.wang}@uwaterloo.ca

ABSTRACT

With the fast advances in video acquisition, computational imaging, and display technologies, there has been a growing interest in high dynamic range (HDR) videos. Tone mapping operators (TMOs) that convert HDR content to low dynamic range (LDR) ones provide a practically useful solution for the visualization of HDR videos on standard LDR displays, where the user experience highly depends on the performance of the TMOs being used. Without an appropriate perceptual quality measure, different TMOs cannot be compared. Subjective experiments may be a reliable solution, but is time consuming, expensive, and difficult to be embedded into optimization processes. Here we make one of the first attempts to develop an objective quality assessment model for tone-mapped videos that incorporates structural fidelity, statistical naturalness and memory effect. Validation using subject-rated tone-mapped videos show that the proposed method is well-correlated with subjective scores.

Index Terms— High dynamic range video, tone-mapped videos, video quality assessment.

1. INTRODUCTION

Recently high dynamic range (HDR) images and videos have attracted a great deal of attention in the industry, thanks to the fast advancement of sensor, computational imaging and display technologies. HDR delivers high precision content closer to what is perceived from real scenes by recording a wide dynamic range of luminance, far beyond what standard acquisition and display devices can handle. Currently HDR TVs and monitors are still quite expensive. To visualize HDR content on standard displays, significant effort has been made to develop tone mapping operators (TMOs) that convert HDR contents to Low Dynamic Range (LDR) ones. Numerous TMOs for still images have been proposed in the last decades [1], but very few of them are meant to handle dynamic range reduction of HDR videos [2] [3]. The perceptual experience of HDR videos vary when different TMO techniques are adopted. Assessing the overall quality of tone-mapped videos is a critical and challenging problem.

So far the assessment of tone-mapped videos mostly relies on human subjective evaluations. Many subjective studies have been conducted to measure the performance of different TMOs and to investigate the effects of different perceptual attributes in perceived quality of tone-mapped videos [4] [2] [5]. Although subjective testing is often considered to be the most reliable method, also has fundamental limitations. Specifically, it is time consuming and expensive, and is difficult to be integrated into optimization frameworks.

Typical video quality assessment (VQA) models assume the reference and distorted videos to have the same dynamic range. However, the dynamic range of HDR and LDR videos are drastically different, and thus popular objective VQA models are not applicable. A dynamic range independent objective assessment method known as DRIVQM was proposed in [6], where the VQA model is an extension of the objective measure for tone-mapped images called DRIVDP [7] built upon the HDR-VDP [8]. These models are human visual system (HVS) based fidelity measures aiming to classify visible and invisible distortions. Subjective experiments suggested that DRIVDP and DRIVQM are well-correlated with human subjective assessment. However, they do not provide a single quality score for the entire image or video being tested.

In this paper we aim to develop an objective VQA model for tone-mapped videos that uses the corresponding HDR videos as references. We propose a quality index that adopts the concepts of multi-scale structural fidelity and statistical naturalness, and combines them with a measure to handle memory effect [9] [10] [11]. Subjective studies was carried out to calibrate the parameters and to evaluate the performance of the proposed measures. Our experiments show that the proposed measure is well-correlated with human subjective opinions.

2. PROPOSED METHOD

2.1. Intra Frame Structural Fidelity

It was hypothesized that the HVS uses the preservation of structural information to determine the perceived distortions of image and videos, and many successful objective measures such as SSIM are based on this hypothesis [12]. A structural fidelity measure for tone-mapped images was proposed in [13], and the benefit of using such a measure in optimizations was demonstrated in [13] [14]. Here we adopt the structural fidelity measure to derive an intra frame structural fidelity measure for tone-mapped videos.

Let $\mathbf{X_t}$ and $\mathbf{Y_t}$ be the t-th frame of the HDR and tone-mapped LDR videos, respectively. Assuming $\mathbf{x_t}$ and $\mathbf{y_t}$ are two local patches extracted from $\mathbf{X_t}$ and $\mathbf{Y_t}$, respectively, a local structural fidelity measure is defined as

$$S_{\text{local}}(\mathbf{x_t}, \mathbf{y_t}) = \frac{2\tilde{\sigma}_{x_t}\tilde{\sigma}_{y_t} + C_1}{\tilde{\sigma}_{x_t}^2 + \tilde{\sigma}_{y_t}^2 + C_1} \cdot \frac{\sigma_{x_t y_t} + C_2}{\sigma_{x_t}\sigma_{y_t} + C_2}\,, \qquad (1)$$

where σ_{x_t}, σ_{y_t} and $\sigma_{x_t y_t}$ denote the local standard deviations and cross correlation between the two corresponding patches, respectively, and C_1 and C_2 are stability constants. The first term is a modification of the local contrast comparison component in SSIM [12], and the second term is the same as the structure comparison component in SSIM [12]. The local contrast comparison term is based on two considerations: First, as along as the contrast in the HDR and LDR patches are both significant or both insignificant, the contrast differences should not be penalized. Second, the measure should penalize the cases in which the contrast is significant in one of the patches but not in the other. To determine the significance of local contrast, the local standard deviation σ is passed through a nonlinear psychometric function given by [13]

$$\tilde{\sigma} = \frac{1}{\sqrt{2\pi}\theta_\sigma} \int_{-\infty}^{\sigma} \exp\left[-\frac{(t-\tau_\sigma)^2}{2\theta_\sigma^2}\right] dt\,, \tag{2}$$

where τ_σ is a threshold contrast and $\theta_\sigma = \tau_\sigma/3$. In [13], a single threshold contrast model was derived based on a contrast sensitivity function (CSF) for both HDR and LDR images. However, CSF curves depend on background luminance adaptation and thus using a single threshold contrast model for both LDR and HDR images is suboptimal. It was suggested in [15] that even if a single threshold contrast model for LDR images is acceptable, but such simplification results in significant inaccuracies for HDR images. HDR image or video may have been taken from an extremely low or extremely high luminance condition. Such luminance adaptation needs to be taken into account when determining the contrast threshold for HDR frames. Various models of CSF have been proposed in imaging applications. Here we use a well-known CSF function that is simple and empirically fitted to psychophysical data [16]. The model is defined as

$$\mathrm{CSF}(f,l) = c_0 f e^{c_1 f}(1+0.06e^{c_1 f})^{0.5}\,, \tag{3}$$

where

$$c_0 = \frac{540(1+0.7/l)^{-0.2}}{1+\frac{12(1+f/3)^{-2}}{w}}\,, \quad c_1 = 0.3(1+\frac{100}{l})^{0.15}\,, \tag{4}$$

and f is the spatial frequency of the stimulus, w is the stimulus size in degrees of visual angle, and l is the mean luminance of the stimulus in cd/m2. The threshold contrast τ_σ in (2) is the reciprocal of the CSF function in a given spatial frequency f, and thus a contrast versus intensity (CVI) function can be derived by computing $1/\mathrm{CSF}(f,l)$. The CVI function predicts the minimum distinguishable contrast at a particular adaptation level. Fig. 1 depicts a CVI curve where the spatial frequency equals 32 cycles/degree. The CVI

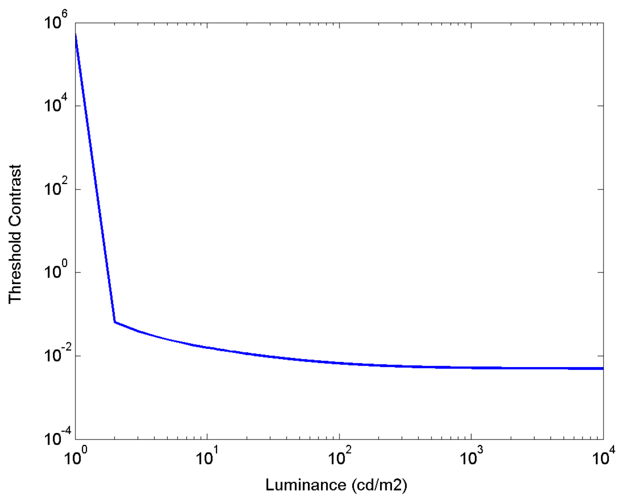


Fig. 1. Contrast versus intensity (CVI) function.

function basically suggests that the visual system is less sensitive to very low luminance condition and so the loss of contrast sensitivity is higher for low luminance levels. The local structural fidelity measure S_{local} is applied using a sliding window that runs across the frame, resulting in a map that reflects the variation of structural fidelity across space. By adopting the multi-scale approach presented in [19][20], the frames are iteratively low-pass filtered and downsampled to create a pyramid structure [13]. The local structural fidelity map is generated at each scale. In TQMI, at each scale, the local intra frame structural fidelity map is generated and then the map is pooled by averaging to provide a single score. However, It is shown that more accurate pooling strategies improves the performance of quality measures [20] [21]. Here, we use information content-weighted pooling proposed in [21]

$$w(\mathbf{x_t},\mathbf{y_t}) = \log\left[\left(1+\frac{\tilde{\sigma}_{\mathrm{x_t}}}{\mathrm{C}}\right)\left(1+\frac{\tilde{\sigma}_{\mathrm{y_t}}}{\mathrm{C}}\right)\right]\,, \tag{5}$$

The structural fidelity score in Scale l is then calculated as

$$S_l = \frac{\sum_{i=1}^{N_l} w(\mathbf{x}_{(\mathbf{t},i)},\mathbf{y}_{(\mathbf{t},i)}) S_{\mathrm{local}}(\mathbf{x}_{(\mathbf{t},i)},\mathbf{y}_{(\mathbf{t},i)})}{\sum_{i=1}^{N_l} w(\mathbf{x}_{(\mathbf{t},i)},\mathbf{y}_{(\mathbf{t},i)})}\,, \tag{6}$$

where $\mathbf{x}_{(\mathbf{t},i)}$ and $\mathbf{y}_{(\mathbf{t},i)}$ are the i-th patches in the t-th HDR and LDR frames being compared, respectively, and N_l is the number of patches in the l-th scale. Finally, the overall intra frame structural fidelity is calculated by combining scale level structural fidelity scores using

$$S_{\mathrm{intra}}(\mathbf{X_t},\mathbf{Y_t}) = \prod_{l=1}^{L} S_l^{\beta_l}\,, \tag{7}$$

where L is the total number of scales and β_l is the weight assigned to the l-th scale as in [19].

Computing threshold contrast for HDR frames based on the CVI function improves the accuracy of the structural fidelity term upon [13] specifically when the HDR frame is taken in low luminance condition. Fig. 2 illustrates how the use of CVI function in detecting significant contrast enhances the accuracy of structural fidelity assessment. Figures 2(a) and 2(d) are tone-mapped LDR images using Fattal's [17] and Reinhard's [18] TMOs. It can be observed that despite more structural details are visible in Fig. 2(d), it is very noisy specially near the borders of the image. By contrast, Fig. 2(a) contains fewer structural details but is more visually appealing. This is because the corresponding HDR image is taken in very low luminance condition where typically significant noise is recorded. Figures 2(b) and 2(e) are the structural fidelity maps while Figures 2(c) and 2(f) show the intra frame structural fidelity maps. By using the CVI function the region around the center is identified as having insignificant contrast, and thus Fig. 2(a) with insignificant contrast around the center has better fidelity score. However, the structural fidelity measure in [13] suggests that the entire HDR image to have significant contrast. Therefore, the fidelity maps in 2(c) and 2(f) provide more reasonable assessment than 2(b) and 2(e).

2.2. Intra Frame Statistical Naturalness

Dynamic range reduction causes inevitable loss in structural details. On the other hand, preserving most of the structural details from HDR frames does not necessarily results in a high quality LDR frames, partially because the subjects do not see the original HDR content, and thus their judgment is somewhat blind. On the other hand, good looking tone-mapped frames should look natural. Although naturalness is a subjective concept, studies have shown that the most influential attributes in perceived naturalness are brightness and contrast [22]. We derive a statistical naturalness model from the statistics of about 7,000 grey-scale images representing many different types of natural scenes. Histograms of the means and standard

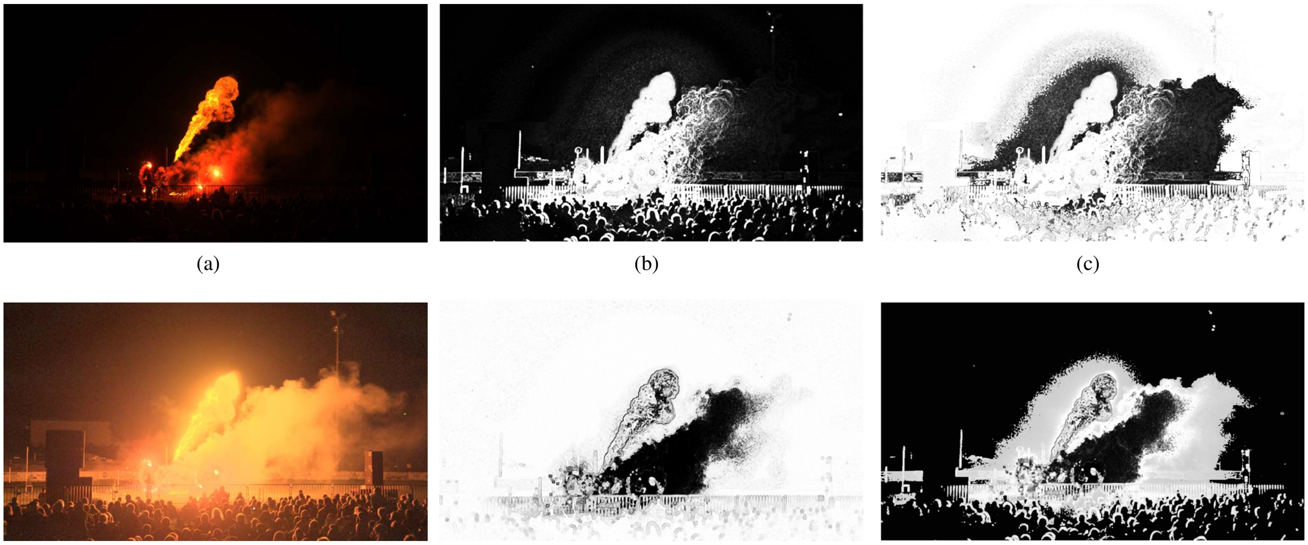


Fig. 2. (a) and (d): tone-mapped frame using TMO in [17] and [18], respectively. (b) and (e): first scale structural fidelity maps using TMQI in [13]. (c) and (f): first scale structural fidelity maps by the proposed method.

deviations of these images are shown in Fig. 3, which are useful indicators of the global intensity and contrast of images. We found that these histograms can be well fitted using a Gaussian probability density functions given by

$$P_m(m) = \frac{1}{\sqrt{2\pi}\sigma_m} \exp\left[-\frac{(m-\mu_m)^2}{2\sigma_m^2}\right] \tag{8}$$

and

$$P_d(d) = \frac{1}{\sqrt{2\pi}\sigma_d} \exp\left[-\frac{(d-\mu_d)^2}{2\sigma_d^2}\right] . \tag{9}$$

The model parameters are estimated by regression, and the best values we found are μ_m = 117.09 and σ_m = 34.88 in (8), and μ_d = 60.7 and σ_d = 12.15 in (9), respectively. Based on recent observations on the independence of image brightness and contrast [23], the statistical naturalness quantity is defined as the product of the two density functions

$$N_{\text{intra}}(\mathbf{Y_t}) = \frac{1}{K} P_m\, P_d\,, \tag{10}$$

where K is a normalization factor given by $K = \max\{P_m P_d\}$ to ensure that N_{intra} is upper bounded by 1.

2.3. Memory effect

It has been known for decades that subjective ratings of videos are highly influenced by their observation in the recent sections of the sequence [24]. The ratings are relatively high, if the recent section of the video is of good quality, and vice versa. It is as if subjects perform weighted temporal averaging of the instantaneous quality variations over time, where higher weights are given to the more recently seen sections of the video. To model such a recency memory effect, we recursively incorporate the memory component to intra frame structural fidelity and intra frame statistical naturalness at each time instant t by taking a weighted average of these features obtained over previous seconds, where the weight model is an exponential function given by

$$w_t(t) = \exp(-\lambda \mathrm{t})\,, \tag{11}$$

where λ controls the decaying rate over the past seconds, and is empirically set to be 0.5.

2.4. Quality Model

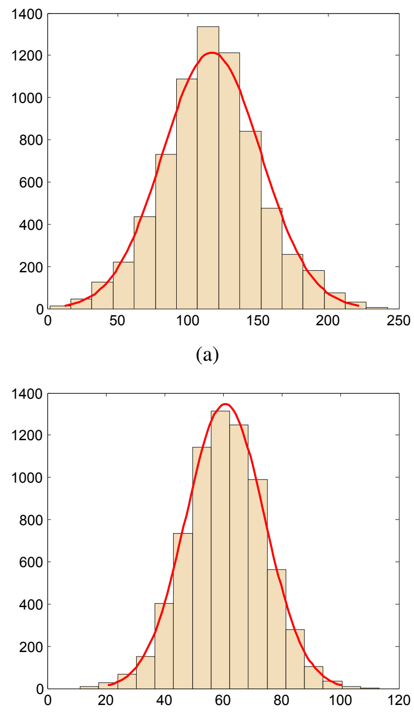
The intra frame structural fidelity and the intra frame statistical naturalness measures characterize different aspects of the quality of the tone-mapped videos. In many practical applications, users desire to obtain a single score that indicates the overall quality of the video. We combine these components by

$$Q = \frac{1}{T}\sum_{t=1}^{T}[w_s {S_{\text{intra}}}^{\alpha}(\mathbf{X_t},\mathbf{Y_t}) + (1-w_s){N_{\text{intra}}}^{\beta}(\mathbf{Y_t})] \quad , \tag{12}$$

where w_s and w_n adjust the relative importance of the components, and α, β and γ determine their sensitivities, respectively. Since all three components are upper-bounded by 1, the overall quality measure is also upper-bounded by 1. The parameters in (12) are to be determined to best fit the subjective evaluations. The details are given in Section 3.

3. VALIDATION

To tune the parameters and validate the proposed quality model, we carried out a subjective quality assessment experiment. Ten video sequences from MPEG [25] were used and tone-mapped using four TMOs developed by Reinhard et al. [18], Fattal et. al. [17], Ashikhmin et. al. [26], and Durand & Dorsey [27] with a temporal coherency method proposed in [28] and [29]. Thirty subjects aged between 20 and 30 participated in the experiment. They were given a brief introduction before being asked to assign an overall quality score to each tone-mapped video. After the subjective test, a statistical analysis as recommended in ITU. BT 500 was performed and outliers were removed from the subjective data. The subjective scores for each video were averaged to a mean opinion score



(b)

Fig. 3. Histograms of (a) means and (b) standard deviations.

(MOS). The MOS value obtained for each video is treated as the "ground truth" and is employed to compare against any prediction method. We calculate Spearman rank order correlation (SRCC) and Pearson linear correlation coefficient (PLCC) to quantify the level of agreement between MOS and the quality prediction method being tested.

Before assessing the objective quality models, we first evaluate how each individual component is correlated with subjective data. Table 1 provides the SRCC and PLCC performance of the intra frame structural fidelity and the intra frame structural naturalness. The results suggest that the natural-looking videos may be more appealing to subjects compared to the ones that better preserve structural details but do not look natural. The optimal parameters

Table 1. Performance evaluation of individual components

Feature	SRCC	PLCC
Intra frame structural fidelity	0.4	0.31
Intra frame statistical naturalness	0.57	0.70

in (12) are determined by solving a simple gradient decent optimization problem where the goal is to achieve maximum correaltion between Q and the subjective scores. Specifically, we randomly divide all videos to a training group and a testing group, each of which includes half of the data. The optimal weight parameters are then obtained using the training group, followed by a test performed against the testing group. We repeat the process multiple times, each with a different random split between the training and testing groups. The mean and std of the parameters obtained for each trial is given in Table 2, together with the corresponding mean SRCC and PLCC performance test results over 100 trails. It can be seen that the weights resulted from all 100 trials are fairly close to each other, and thus we use their mean values as the final model parameters. To the best of our knowledge, there is only one other objective

Table 2. Statistics of model parameters and SRCC and PLCC performance from 100 trials

mean/ std of w_s	mean/ std of α	mean/ std of β	SRCC	PLCC
0.4/0.1	0/0.001	0.05/0.003	0.70	0.78

VQA known as DRIVQM [30] that is directly applicable to the scenario we are interested in. However, DRIVQM does not provide a single quality score for an entire tone-mapped video, making it impossible to be validated with our subjective data. Therefore, we compare its performance against the TMQI model that is designed for tone-mapped images [13]. The original TMQI parameters were determined using a different database of tone-mapped images, and thus for fair comparison we tune the model parameters in TMQI using our new subjective data for HDR videos, and the resulting model is denoted by TMQI_v.

In addition, to observe how an average subject performs in evaluating the overall quality of tone-mapped videos, we compute the SRCC and PLCC values between MOS and the scores given by each individual subject. When this is done for all 30 subjects, we take the mean and the standard deviation of the SRCC and PLCC values across all subjects. These values give useful information about the performance of an "average subject", and provide a meaningful anchor for our study.

Table 3. SRCC and PLCC performance evaluation

Measure	SRCC	PLCC
TMQI	0.51	0.50
TMQI_v	0.62	0.76
Proposed	0.70	0.79
Mean/stds of subject performance	0.76/0.06	0.79/0.04

The performance and comparison results are summarized in Table 3. It can be seen that the proposed method outperforms TMQI and TMQI_v. Moreover, the SRCC and PLCC values are within the range of ± 1 standard deviation from those of the mean over all subjects. This indicates that the proposed measure behaves quite similarly to an average human subject.

4. CONCLUSION

We developed a novel objective model to assess the quality of tone-mapped videos by combining intra frame structural fidelity, intra frame statistical naturalness, and memory effect. Our experiments show that the proposed method is well-correlated with subjective opinions. Future work includes using the proposed measure in the design and optimization of novel video tone mapping algorithms.

5. REFERENCES

[1] E. Reinhard, G. Ward, S. Pattanaik, P. Debevec, W. Heidrich, and K. Myszkowski, *High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting*, Morgan Kaufmann Publishers Inc., 2010.

[2] G. Eilertsen, R. Wanat, R. K. Mantiuk, and J. Unger, "Evaluation of tone mapping operators for HDR-video," *Computer Graphics Forum*, vol. 32, no. 7, pp. 275–284, 2013.

[3] Ronan Boitard, Rémi Cozot, Dominique Thoreau, and Kadi Bouatouch, "Zonal brightness coherency for video tone mapping," *Signal Processing: Image Communication*, vol. 29, no. 2, pp. 229–246, 2014.

[4] J. Petit and R. 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.

[5] M. Melo, M. Bessa, K. Debattista, and A. Chalmers, "Evaluation of hdr video tone mapping for mobile devices," *Signal Processing: Image Communication*, vol. 29, no. 2, pp. 247–256, 2014.

[6] T. O. Aydin, M. Čadík, K. Myszkowski, and H. Seidel, "Video quality assessment for computer graphics applications," in *ACM SIGGRAPH Asia 2010 Papers*. 2010, pp. 161:1–161:12, ACM.

[7] T. O. Aydm, R. Mantiuk, K. Myszkowski, and H. . Seidel, "Dynamic range independent image quality assessment," in *SIGGRAPH'08: International Conference on Computer Graphics and Interactive Techniques, ACM SIGGRAPH*, 2008.

[8] R. Mantiuk, S. J. Daly, K. Myszkowski, and H. Seidel, "Predicting visible differences in high dynamic range images: model and its calibration," in *Electronic Imaging 2005*. International Society for Optics and Photonics, 2005, pp. 204–214.

[9] H. Yeganeh, R. Kordasiewicz, M. Gallant, D. Ghadiyaram, and A. C. Bovik, "Delivery quality score model for internet video," in *Image Processing (ICIP), 2014 IEEE International Conference on*. IEEE, 2014, pp. 2007–2011.

[10] T. Hoßfeld, S. Biedermann, R. Schatz, A. Platzer, S. Egger, and M. Fiedler, "The memory effect and its implications on web qoe modeling," in *Proceedings of the 23rd International Teletraffic Congress*. International Teletraffic Congress, 2011, pp. 103–110.

[11] A.K. Moorthy, Lark Kwon Choi, A.C. Bovik, and G. de Veciana, "Video quality assessment on mobile devices: Subjective, behavioral and objective studies," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 6, no. 6, pp. 652–671, Oct 2012.

[12] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Proc.*, vol. 13, pp. 35–44, 2004.

[13] Hojatollah Yeganeh and Zhou Wang, "Objective quality assessment of tone-mapped images," *Image Processing, IEEE Transactions on*, vol. 22, no. 2, pp. 657–667, 2013.

[14] K. Ma, H. Yeganeh, K. Zeng, and Z. Wang, "High dynamic range image compression by optimizing tone mapped image quality index," *Image Processing, IEEE Transactions on*, vol. 24, no. 10, pp. 3086–3097, Oct 2015.

[15] R. Mantiuk, K. Myszkowski, and H. Seidel, "Visible difference predicator for high dynamic range images," in *Systems, Man and Cybernetics, 2004 IEEE International Conference on*. IEEE, 2004, vol. 3, pp. 2763–2769.

[16] Peter G. J. Barten, *Contrast sensitivity of the human eye and its effects on image quality*, SPIE Optical Engineering Press, Washington, 1999.

[17] R. Fattal, D. Lischinski, and M. Werman, "Gradient domain high dynamic range compression," in *Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '02*, 2002, pp. 249–256.

[18] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic tone reproduction for digital images," *in Proc. of 29th annual Conference on Computer Graphics and Interactive Techniques, ACM SIGGRAPH*, vol. 21, pp. 267–276, 2002.

[19] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," *in Proc. of 37th Asilomar Conf. Signals, Systems and Computers*, 2003.

[20] Z. Wang and Qiang Li, "Information content weighting for perceptual image quality assessment," *To appear in IEEE Trans. Image Proc.*, 2011.

[21] Z. Wang and X. Shang, "Spatial pooling strategies for perceptual image quality assessment," in *Image Processing, 2006 IEEE International Conference on*, Oct 2006, pp. 2945–2948.

[22] Martin Čadík and Pavel Slavík, "The naturalness of reproduced high dynamic range images," in *IV '05: Proceedings of the Ninth International Conference on Information Visualisation*, Washington, DC, USA, 2005, pp. 920–925, IEEE Computer Society.

[23] V. Mante, R. Frazor, V. Bonin, W. Geisler, and M. Carandini, "Independence of luminance and contrast in natural scenes and in the early visual system," *Nature Neuroscience*, vol. 8, no. 12, pp. 1690–1697, 2005.

[24] R. Aldridge, J. Davidoff, M. Ghanbari, D. Hands, and D. Pearson, "Recency effect in the subjective assessment of digitally-coded television pictures," in *Image Processing and its Applications, 1995., Fifth International Conference on*. IET, 1995, pp. 336–339.

[25] "Call for evidence (cfe) for hdr and wcg video coding," MPEG adhoc group on support for HDR and XYZ, N15083, 111th meeting of the Moving Picture Experts Group (MPEG), Geneva , 2015.

[26] M. Ashikhmin, "A tone mapping algorithm for high contrast images," in *Proceedings of the 13th Eurographics workshop on Rendering*, 2002, pp. 145–156.

[27] F. Durand and J. Dorsey, "Fast bilateral filtering for the display of high-dynamic-range images," in *ACM Transactions on Graphics*, 2002, vol. 21, pp. 257–266.

[28] R. Boitard, K. Bouatouch, R. Cozot, D. Thoreau, and A. Gruson, "Temporal coherency for video tone mapping," in *SPIE Optical Engineering+ Applications*. International Society for Optics and Photonics, 2012, pp. 84990D–84990D.

[29] R. Boitard, R. Cozot, D. Thoreau, and K. Bouatouch, "Zonal brightness coherency for video tone mapping," *Signal Processing: Image Communication*, vol. 29, no. 2, pp. 229–246, 2013.

[30] M. Cadik, T. O. Aydin, K. Myszkowski, and H. Seidel, "On evaluation of video quality metrics: an hdr dataset for computer graphics applications," 2011.