FSITM: A Feature Similarity Index For Tone-Mapped Images

Hossein Ziaei Nafchi, Atena Shahkolaei, Reza Farrahi Moghaddam, *Member, IEEE.*, and Mohamed Cheriet, *Senior Member, IEEE*

Abstract—In this work, based on the local phase information of images, an objective index, called the feature similarity index for tone-mapped images (FSITM), is proposed. To evaluate a tone mapping operator (TMO), the proposed index compares the locally weighted mean phase angle map of an original high dynamic range (HDR) to that of its associated tone-mapped image calculated using the output of the TMO method. In experiments on two standard databases, it is shown that the proposed FSITM method outperforms the state-of-the-art index, the tone mapped quality index (TMQI). In addition, a higher performance is obtained by combining the FSITM and TMQI indices. The MATLAB source code of the proposed metric(s) is available at https://www.mathworks.com/matlabcentral/fileexchange/59814.

Index Terms—Tone-mapping operator, Objective quality assessment, Mean phase, High dynamic range.

I. INTRODUCTION

THERE is increasing interest in high dynamic range (HDR) images, HDR imaging systems, and HDR displays. The visual quality of high dynamic range images is vastly higher than that of conventional low-dynamic-range (LDR) images, and the significance of the move from LDR to HDR has been compared to the momentous move from black-and-white to color television [1]. In this transition period, and to guarantee compatibility in the future, there has been a need to develop methodologies to convert an HDR image into its 'best' LDR equivalent. For this conversion, tone mapping operators (TMOs) have attracted considerable interest. Tone-mapping operators have been used to convert HDR images into their LDR associated images for visibility purposes on non-HDR displays.

Unfortunately, TMO methods perform differently, depending on the HDR image to be converted, which means that the best TMO method must be found for each individual case. A survey of various TMOs for HDR images and videos is provided in [2] and [3]. Traditionally, TMO performance has been evaluated subjectively. In [4], a subjective assessment was carried out using an HDR monitor. Mantiuk et al. [5] propose an HDR visible difference predictor (HDR-VDP) to estimate the visibility differences of two HDR images, and this

H. Ziaei Nafchi, A. Shahkolaei, R. Farrahi Moghaddam and M. Cheriet are with the Synchromedia Laboratory for Multimedia Communication in Telepresence, École de technologie supérieure, Montreal (QC), Canada H3C 1K3; Tel.: +1-514-396-8972; Fax: +1-514-396-8595; Emails: hossein.zi@synchromedia.ca, atena.shahkolaei.1@ens.etsmtl.ca, rfarrahi@synchromedia.ca, imriss@ieee.org, mohamed.cheriet@etsmtl.ca

tool has also been extended to a dynamic range independent image quality assessment [6]. However, the authors did not arrive at an objective score, but instead evaluated the performance of the assessment tool on HDR displays. Although subjective assessment provides true and useful references, it is an expensive and time-consuming process. In contrast, the objective quality assessment of tone mapping images enables an automatic selection and parameter tuning of TMOs [7], [8]. Consequently, objective assessment of tone-mapping images, which is proportional to the subjective assessment of the images, is currently of great interest.

Recently, an objective index, called the tone mapping quality index (TMQI) was proposed in [2] to objectively assess the quality of the individual LDR images produced by a TMO. The TMQI is based on combining an SSIM-motivated structural fidelity measure with a statistical naturalness:

$$\text{TMQI}(H, L) = a[S(H, L)]^\alpha + (1 - a)[N(L)]^\beta. \quad (1)$$

where S and N denote the structural fidelity and statistical naturalness, respectively. H and L denote the HDR and LDR images. The parameters α and β determine the sensitivities of the two factors, and a ($0 \leq a \leq 1$) adjusts their relative importance. Both S and N are upper bounded by 1, and so the TMQI is also upper bounded by 1 [8]. Although the TMQI clearly provides better assessment for tone-mapped images than the well-known image quality assessment metrics, like SSIM [9], MS-SSIM [10], and FSIM [11], its performance is not perfect. Liu et al. [12] replaced the pooling strategy of the structural fidelity map in the TMQI with various visual saliency-based strategies for better quality assessment of tone mapped images. They examined a number of visual saliency models and conclude that integrating saliency detection by combining simple priors (SDSP) into the TMQI provides better assessment capability than other saliency detection models.

In this paper, we first propose a feature similarity index for tone-mapped images (FSITM) which is based on the phase information of images. It has been observed that phase information of images prevails its magnitude [13]. Also, physiological evidence indicates that the human visual system responds strongly to points in an image where the phase information is highly ordered [14]. Based on this assumption, several quality assessment metrics have been proposed [11], [15], [16]. In [11], the maximum moment of phase congruency covariance, which is an edge strength map, is used. Hassen et al. [15] used local phase coherence for image sharpness assessment. Saha et al. [16] proposed an image quality assessment using phase deviation sensitive energy features. Unfortunately, these

metrics do not provide a reliable assessment for tone mapped images.

The FSITM images proposed in this paper uses the phase-derived feature type of the images in a different way from that proposed in [11], [15], [16]. Our FSITM uses a locally weighted mean phase angle (LWMPA) [17], which is a feature map based on the local-phase. This phase-derived map is noise independent, and therefore there is no parameter to set for noise estimation. The proposed FSITM assesses both the appearance of the real world scene and the most pleasing image for human vision.

Given the FSITM and the TMQI, we also proposed a combined metric, FSITM_TMQI, which provides much better assessment of tone-mapped images. In the experiments, we compare the objective scores of our proposed similarity indices (FSITM, FSITM_TMQI), along with TMQI [2], on two major datasets [18], [19].

II. THE PROPOSED SIMILARITY INDEX

The proposed FSITM similarity index for tone-mapped images is based on a phase-derived feature map. As we mentioned before, phase-derived features have already been used successfully for quality assessment [11], [15], [16]. However, their results for evaluating tone-mapped images is not reliable similar to other popular quality assessment metrics like the SSIM and its variations [9], [10]. For this reason, we use the locally weighted mean phase angle (LWMPA) map in this paper, because it is a feature that is robust with respect to noise. Below, we briefly describe the theory and formulation of the LWMPA, and then discuss our proposed similarity index which is based on this feature map.

Let $M_{\rho r}^e$ and $M_{\rho r}^o$, which are known in the literature as quadratic pairs, denote the even symmetric and odd symmetric log-Gabor wavelets at a scale ρ and orientation r [20]. By considering $f(\mathbf{x})$ as a two-dimensional signal on the two-dimensional domain of $\mathbf{x}$, the response of each quadratic pair of filters at each image point $\mathbf{x}$ forms a response vector by convolving with $f(\mathbf{x})$:

$$\Big[e_{\rho r}(\mathbf{x}), o_{\rho r}(\mathbf{x})\Big] = \Big[f(\mathbf{x}) * M_{\rho r}^e, f(\mathbf{x}) * M_{\rho r}^o\Big]. \quad (2)$$

where the values $e_{\rho r}(\mathbf{x})$ and $o_{\rho r}(\mathbf{x})$ are real and imaginary parts of a complex-valued wavelet response at a scale ρ and an orientation r. We can now compute the local phase $\phi_{\rho r}(\mathbf{x})$ of the transform at a given wavelet scale ρ and orientation r:

$$\phi_{\rho r}(\mathbf{x}) = \text{arctan2}\Big(o_{\rho r}(\mathbf{x}), e_{\rho r}(\mathbf{x})\Big), \quad (3)$$

where $\text{arctan2(x,y)} = 2\arctan\frac{x}{\sqrt{x^2+y^2}+y}$. The locally-weighted mean phase angle $ph(\mathbf{x})$ is obtained using the summation of all filter responses over all the possible orientations and scales:

$$ph(\mathbf{x}) = \text{arctan2}\Big[\sum_{\rho,r} e_{\rho r}(\mathbf{x}), \sum_{\rho,r} o_{\rho r}(\mathbf{x})\Big]. \quad (4)$$

The pixels of $ph(\mathbf{x})$ take values between $-\pi/2$ (a dark line), $+\pi/2$ (a bright line), and 0 for steps. This classification of step and line features has been further studied in [21].

Algorithm 1 The feature similarity index for tone-mapped images (FSITM).

1: **procedure** FSITM(**H**, **L**, **C**) **Start**
2:    **H**: HDR, **L**: LDR, **C**∈{R, G, B}
3:    **LogH** $= log(\mathbf{H})$;
4:    Calculate ph($\mathbf{x}$) for **C** channel of images **L**, **H** and **LogH**.
5:    FSITM$^C = \alpha F^C(H,L) + (1-\alpha)F^C(logH, L)$;
6: **return** FSITM^C
7: **end procedure**

There are a few parameters to be considered in the calculation of ph($\mathbf{x}$). In our set of experiments, we determine the best fixed values for this operation (see section III). Unlike the phase-derived edge map and local phase that are used in other research [11], [15], the locally weighted mean phase angle ph($\mathbf{x}$) provides a good representation of image features, including the edges and shapes of objects. Since ph($\mathbf{x}$) indicates both dark and bright lines, it can be used to assess color changes, which is a popular feature of the TMOs. Moreover, the LWMPA is noise-independent, unlike the phase derived features used in [11], [15], [16], which are sensitive to noise, and therefore require an estimation of the noise. Some examples of ph($\mathbf{x}$) outputs are shown in Fig. 1.

We use only ph($\mathbf{x}$) to calculate the FSITM. First, the HDR (H) image is converted into its LDR (L) by simply taking logarithm of its values (log(H)). This rough LDR image is used as one of the reference images for computing the FSITM. Another reference image is the HDR image itself. The details of the FSITM calculations are provided below.

Given the input images H and L, and LogH=log(H) image, the ph($\mathbf{x}$) for each channel C of these three images is calculated using equation (4). The FSITM is based on the simple fact that the features in the two corresponding channels should have remained the same in their ph($\mathbf{x}$) maps. The FSITM is equal to 1 if all the feature types are the same, and 0 if they are all different. First, we define the feature similarity index for a channel C used in calculation of the FSITM:

$$\text{F}^C(I_1, I_2) = |P_1^C(\mathbf{x}) \cap P_2^C(\mathbf{x})| / (row \times col), \quad (5)$$

where P($\mathbf{x}$) denotes a binary image of ph($\mathbf{x}$):

$$P(\mathbf{x}) = U\big(ph(\mathbf{x})\big), \quad (6)$$

where $U(\cdot)$ is the unit-step function. For the case of tone-mapped images, the FSITM for a channel C is defined as:

$$\text{FSITM}^C = \alpha F^C(H,L) + (1-\alpha)F^C(\text{LogH}, L). \quad (7)$$

where α ($0 \leq \alpha \leq 1$), controls the impact factor of H and LogH in the calculation of the FSITM. Algorithm 1 lists all the steps in the process of calculating our proposed FSITM.

We also found that combining the FSITM and the TMQI provides a better assessment of tone-mapped images. Therefore, we proposed a combined index of the FSITM and the TMQI based on the following equation:

$$\text{FSITM}^C\_\text{TMQI} = (\text{FSITM}^C + \text{TMQI}) / 2 \quad (8)$$

In most of the cases, the different properties of these two indices cause them to moderate similarity estimation mistakes of each other.

(b) subjective score = 2.00   (b) subjective score = 5.95   (c) subjective score = 6.65   (d) subjective score = 7.8

TMQI=0.9191, FSITM^R=0.8355   TMQI=0.8800, FSITM^R=0.7825   TMQI=0.7673, FSITM^R=0.7808   TMQI=0.7622, FSITM^R= 0.7514
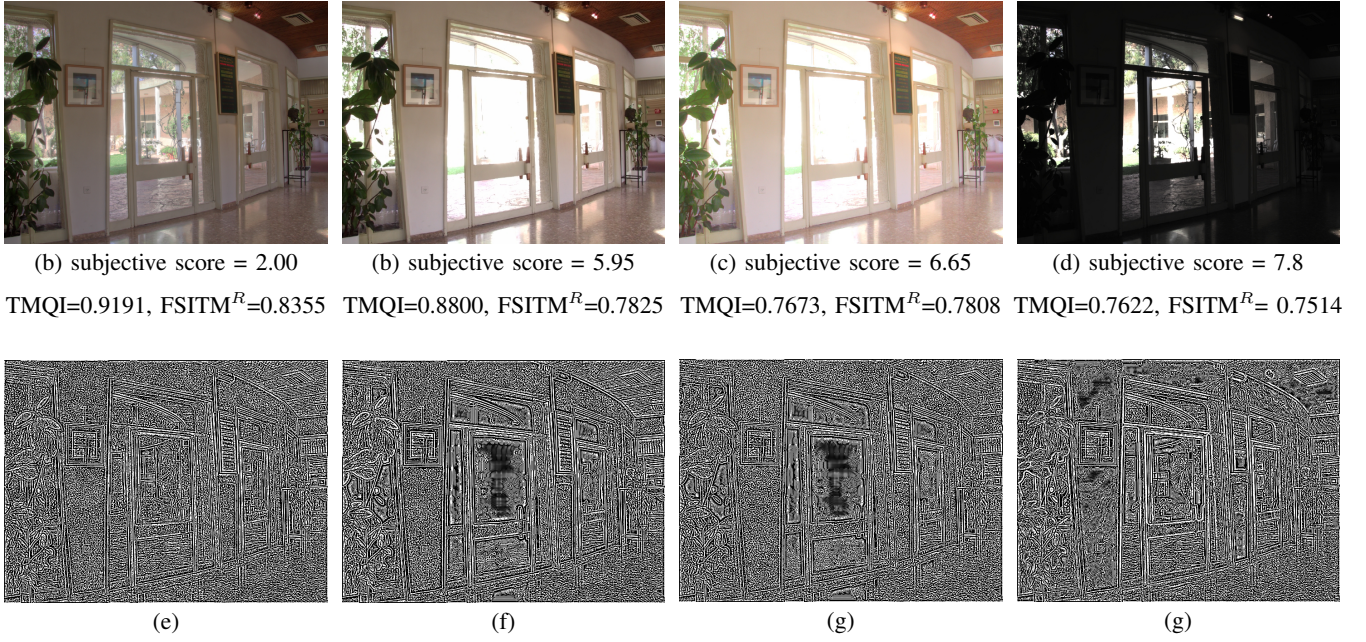

(e)   (f)   (g)   (g)

Fig. 1. (a)-(d) Some LDR images using different TMOs [2], along with their corresponding TMQI and FSITM scores for each. (e)-(g) The associated LWMPA maps of their red channel.

TABLE I
PERFORMANCE COMPARISON OF THE PROPOSED QUALITY INDICES AND TMQI [2], [18] ON THE DATASET A INTRODUCED IN [2], [18].

SRCC						
Index	TMQI [2], [18]	FSITM^R	FSITM^R_TMQI	FSITM^G	FSITM^G_TMQI	FSITM^B	FSITM^B_TMQI
Min	0.6826	0.6190	**0.7143**	0.5476	**0.7143**	0.1796	0.5509
Median	0.7857	0.8095	**0.8571**	0.8333	**0.8571**	**0.8571**	**0.8571**
Average	0.8058	0.8145	**0.8559**	0.8178	0.8424	0.7183	0.8097
STD	0.1051	0.1214	**0.0863**	0.1310	**0.0886**	0.2536	0.1229
KRCC						
Min	0.5455	0.5000	**0.5714**	0.3571	**0.5714**	0.2143	0.4001
Median	0.6429	0.7143	0.7143	0.7143	**0.7857**	0.7143	0.7143
Average	0.6840	0.7126	**0.7508**	0.6935	0.7317	0.5979	0.6838
STD	0.1221	0.1423	**0.1083**	0.1711	**0.1078**	0.2711	0.1436

TABLE II
PERFORMANCE COMPARISON OF THE PROPOSED QUALITY INDICES AND TMQI [2], [18] ON THE DATASET B INTRODUCED IN [19], [22].

SRCC						
Index	TMQI [2], [18]	FSITM^R	FSITM^R_TMQI	FSITM^G	FSITM^G_TMQI	FSITM^B	FSITM^B_TMQI
Min	0.7198	0.7363	0.8901	0.7692	**0.9231**	0.7637	0.8462
Average	0.7985	0.7692	0.9102	0.8461	**0.9267**	0.8241	0.8901
KRCC						
Min	0.5385	0.5897	0.6923	0.6154	**0.7692**	0.5385	0.6410
Average	0.6410	0.6410	0.7692	0.7265	**0.8119**	0.6410	0.7264

III. EXPERIMENTAL RESULTS

To evaluate the proposed FSITM index, we used the dataset (dataset A) introduced in [2] and [19]. The first dataset contains 15 HDR images, along with 8 LDR images for each HDR image. The HDR images were produced using different TMOs. The quality of LDRs is ranked from 1 (best quality) to 8 (worst quality). The ranks were obtained based on a subjective assessment of 20 individuals. The second HDR dataset (dataset B) used is also available along with subjective ranks for LDR images [19], [22]. That dataset contains three HDR images, and 14 LDR images for each HDR image.

To objectively evaluate the performance of the various similarity indices considered in our experiments, we use the Spearman rank-order correlation coefficient (SRCC) and the Kendall rank-order correlation coefficient (KRCC) metrics.

The proposed similarity indices (FSITM^C, FSITM^C_TMQI) are compared with the TMQI [2]. The results are listed in Tables I and II. The performance of the TMQI is listed based on the scores obtained by running the Matlab source code provided by Yeganeh and Wang in [18]. The FSITM^G outperforms the TMQI in terms of SRCC and KRCC for both datasets. In general, there is less variation in TMQI performance than in FSITM performance. In contrast, the FSITM^R_TMQI and FSITM^G_TMQI are more robust,

and also they outperform the FSITM and TMQI in terms of both the SRCC and KRCC scores.

It is worth to report the available results of other indices which have recently been proposed in the literature [12], [23]. In [12], the authors reported the SRCC performance of their proposed index for the dataset A [18]. Their minimum and average SRCC performance is 0.6905 and 0.8408, respectively. Their standard deviation of SRCC scores is reported as 0.0907. For the same dataset, the median performance of the ref. [23] is reported as follows: SRCC=0.8106 and KRCC=0.5865.

A number of parameters impact the quality of the locally weighted mean phase angle ph($\mathbf{x}$), namely the number of filter scales N_ρ, the wavelength of the smallest scale filter wLen, and the scaling factor between successive filters mult. In the experiments, these parameters were set to $N_\rho = 2$, $wLen = 2$, and $mult = 2$ for the LogH image, while they were set to $N_\rho = 2$, $wLen = 8$, and $mult = 8$ for the original HDR image. The rational for using two different set of parameters is that the size of the image features could be different. Overall, it is the three parameters of ph($\mathbf{x}$) along with the value of α that influence the performance of the proposed indices.

In this work, we only used the original HDR image and its logarithm image LogH. It is worth mentioning that we have tried the same strategy used in defining FSITM in RGB color space in other color spaces, such as Lab and Yxy color spaces. However, we did not get a good performance.

We evaluated the run time of the FSITM and the TMQI as follows: our experiments were performed on a Core i7 3.4 GHz CPU with 16 GB of RAM. The FSITM algorithm was implemented in MATLAB 2012b running on Windows 7. The TMQI and the FSITM took 1.95 and 3.36 seconds respectively to assess images of size $1200{\times}1600$, while the run time for the FSITM^C_TMQI is simply obtained by adding the TMQI and FSITM^C run-times.

IV. CONCLUSION

We have proposed an objective index, called the feature similarity index for tone-mapped images (FSITM), which is based on the local phase similarity of the original HDR and the target converted LDR image. Unlike other studies in which different phase-derived feature maps are used, we have used the locally weighted mean phase angle, which is a robust and noise-independent feature map. The performance of the proposed similarity index is compared with the state-of-the-art TMQI on two datasets, and has been found to be promising. The proposed FSITM and the TMQI have been then combined to obtain a more accurate quality assessment. Further studies are required to develop more comprehensive HDR datasets, along with their subjective scores. Such datasets would allow us to develop better performing indices.

ACKNOWLEDGMENTS

REFERENCES

[1] E. Reinhard, G. Ward, S. Pattanaik, and P. Debevec, *High Dynamic Range Imaging: Acquisition, Display, and Image-based Lighting*. Morgan Kaufmann, 2010.
[2] H. Yeganeh and Z. Wang, "Objective quality assessment of tone-mapped images," *IEEE Transactions on Image Processing*, vol. 22, no. 2, pp. 657–667, Feb 2013.
[3] 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. [Online]. Available: http://dx.doi.org/10.1111/cgf.12235
[4] P. Ledda, A. Chalmers, T. Troscianko, and H. Seetzen, "Evaluation of tone mapping operators using a high dynamic range display," *ACM Trans. Graph.*, vol. 24, no. 3, pp. 640–648, Jul. 2005.
[5] R. Mantiuk, S. Daly, K. Myszkowski, and S. Seidel, "Predicting visible differences in high dynamic range images-model and its calibration," *Proc. SPIE*, vol. 5666, pp. 204–214, 2005.
[6] T. O. Aydin, R. Mantiuk, K. Myszkowski, and H.-P. Seidel, "Dynamic range independent image quality assessment," *ACM Trans. Graph.*, vol. 27, no. 3, pp. 69:1–69:10, 2008.
[7] H. Yeganeh and Z. Wang, "Objective assessment of tone mapping algorithms," in *17th IEEE International Conference on Image Processing (ICIP)*, Sept 2010, pp. 2477–2480.
[8] K. Ma, H. Yeganeh, K. Zeng, and Z. Wang, "High dynamic range image tone mapping by optimizing tone mapped image quality index," in *IEEE International Conference on Multimedia & Expo*, 2014.
[9] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, April 2004.
[10] Z. Wang, E. Simoncelli, and A. Bovik, "Multiscale structural similarity for image quality assessment," in *Thirty-Seventh Asilomar Conference on Signals, Systems and Computers*, vol. 2, Nov 2003, pp. 1398–1402.
[11] L. Zhang, D. Zhang, X. Mou, and D. Zhang, "FSIM: A Feature Similarity Index for Image Quality Assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2378–2386, Aug 2011.
[12] X. Liu, L. Zhang, H. Li, and J. Lu, "Integrating visual saliency information into objective quality assessment of tone-mapped images," in *LNCS 8588 (International Conference on Intelligent Computing, ICIC'14)*, D.-S. H. et al., Ed. Taiyuan, Shanxi, China: Springer, 2014, pp. 376–386.
[13] A. V. Oppenheim and J. S. Lim, "The importance of phase in signals," *Proceedings of the IEEE*, vol. 69, pp. 529–541, 1981.
[14] M. Morrone and D. Burr, "Feature detection in human vision: a phase-dependent energy model," *Royal Society of London*, vol. B 235, pp. 221–245, 1988.
[15] R. Hassen, Z. Wang, and M. Salama, "Image sharpness assessment based on local phase coherence," *IEEE Trans. on Image Processing*, vol. 22, no. 7, pp. 2572–2581, 2013.
[16] A. Saha and Q. J. Wu, "Perceptual image quality assessment using phase deviation sensitive energy features," *Signal Processing*, vol. 93, no. 11, pp. 3182–3191, 2013.
[17] P. Kovesi, "Image features from phase congruency," *Videre: Journal of Computer Vision Research*, vol. 1, pp. 1–26, 1999.
[18] H. Yeganeh and Z. Wang, "Subject-rated image database of tone-mapped images," Online: https://ece.uwaterloo.ca/~z70wang/research/tmqi/, 2013.
[19] "Evaluation of tone mapping operators," Online: http://cadik.posvete.cz/tmo.
[20] G. Papari and N. Petkov, "Edge and line oriented contour detection: State of the art," *Image and Vision Computing*, vol. 29, pp. 79–103, 2011.
[21] P. Kovesi, "Edges are not just steps," in *Asian Conference on Computer Vision*, 2002, pp. 1–6.
[22] M. Čadík, "Perceptually based image quality assessment and image transformations," Ph.D. dissertation, Czech Technical University in Prague, January 2008.
[23] K. Gu, G. Zhai, M. Liu, X. Yang, and W. Zhang, "Details preservation inspired blind quality metric of tone mapping methods," in *Circuits and Systems (ISCAS), 2014 IEEE International Symposium on*, June 2014, pp. 518–521.