Gradient-Weighted Structural Similarity for Image Quality Assessments

Qiaohong Li*†, Yuming Fang‡, Weisi Lin*, and Daniel Thalmann†
*School of Computer Engineering, Nanyang Technological University, Singapore, 639798
†Institute for Media Innovation, Nanyang Technological University, Singapore, 637553
‡School of Information Technology, Jiangxi University of Finance and Economics, China, 330013
Email: {QLI013,FA0001NG}@e.ntu.edu.sg, {WSLIN,DANIELTHALMANN}@ntu.edu.sg

***Abstract*—The goal of Image Quality Assessment (IQA) is to design computational models that can automatically predict the perceived image quality consistent with human subjective ratings. In this paper, we propose a full reference IQA metric gradient weighted structural similarity (GW-SSIM) by incorporating the gradient information to the well-known IQA metric SSIM. Experimental results demonstrate that GW-SSIM can greatly improve the quality prediction accuracy and achieve the best performance among the SSIM-based methods by addressing SSIM's shortcomings. Additionally, incorporating the proposed gradient weighting (GW) map into peak-signal-to-noise ratio (PSNR) also makes it quite competitive to state-of-the-art IQA models, and this is meaningful since PSNR is still a widely adopted metric.**

Keywords*—*image quality assessment, gradient weighting map, structural similarity (SSIM), GW-SSIM, GW-PSNR

I. Introduction

With the fast development of digital imaging and communication technologies, human experience and satisfactory has become one of the most important factors to the success of digital products and services [1]. Visual quality of images and video can be degraded under many circumstances, such as lossy compression, congested network transmission, digital watermarking, *etc*. It is crucial to quantify the impact of various distortions on the perceived image quality [1]. Subjective viewing test is an accurate and reliable way for IQA tasks. However, it is time consuming and expensive, and also cannot be embedded into real-time image/video quality monitoring and prediction applications. Objective IQA metrics which aim to replace the human judgment of perceived quality have attracted much research effort.

Generally, objective IQA metrics can be categorized into full-reference, reduced-reference and no-reference algorithms according to available information of the reference image [1]. For full-reference IQA methods, both the original and distorted images are available for quality evaluation. Full reference IQA models can be further classified into two main groups: the human vision model methods and the engineering methods [1]. The IQA methods based on human vision model aim to mimic the characteristics of the human visual system (HVS) from a bottom-up architecture. Two representative HVS models are visual difference predictor (VDP) [2] and the Sarnoff JND metric [3]. The engineering approach aims to calculate the structure or information content preserved in the distorted image compared with the original one. Many state-of-the-art IQA metrics fall under this category and can achieve great performance in terms of high correlation with the subjective quality scores. The well-known SSIM is designed based on the idea that the perceived quality can be quantified by the local structural distortions [4]. An Information Fidelity Criterion (IFC) metric is proposed to calculate the mutual information between the reference and distorted images [5]. The Visual Information Fidelity (VIF) criterion quantifies the loss of image information to the final quality score [6]. The Most Apparent Distortion (MAD) metric adaptively combines a detection based strategy and an appearance-based strategy to predict the quality score by including both the HVS properties and the structural difference calculation [7]. A Feature SIMilarity (FSIM) Index is proposed based on the assumption that the changes of low-level features can be a good indicator of the final quality [8].

SSIM predicts the visual image quality as the product of three components in the spatial domain, *i.e.* the luminance similarity, contrast similarity and structure similarity [4]. Although SSIM has shown a strong correlation with human perception and are currently used in many image processing applications, there are some built-in limitations for this metric and many modifications for SSIM have been proposed subsequently [9]. A multi-scale SSIM (MS-SSIM) is proposed in [10] to account for the different image resolution and viewing conditions. The gradient-based structural similarity (GSSIM) is proposed in [11] to improve the quality prediction accuracy for badly blurred images by calculating the contrast and structure similarities in gradient domain. A complex wavelet SSIM (CW-SSIM) is built to extend the quality metric to an image similarity measure which can handle small geometric distortions caused by ill-registration of the reference and distorted images [12]. A discrete wavelet transform based structural similarity (DWT-SSIM) is proposed in [13] for image processing algorithms in DWT domain. In [14], the performance of the three components of SSIM and their pairwise combinations are analyzed and the simplified SSIM is proposed by discarding the illumination component for computation efficiency. In [15], the edge and contrast similarity (ECSM) metric is built which replaces the structure similarity component of SSIM with an edge similarity measure. In [16], images are first classified into four categories (changed edge, preserved edge, texture regions, smooth regions) and then fixed weights are assigned to the four regions to obtain the final quality score. A geometric structural distortion (GSD) model is established for visual quality evaluation based on the contrast similarity component, edge direction and gradient magnitude similarities [17]. The information content weighting strategy based on advanced

978-1-4799-8391-9/15/$31.00 ©2015 IEEE

statistical models is proposed in [18] to emphasize image locations with high information content, and the resulting IW-SSIM has shown significant performance improvement. The gradient similarity (GS) model by combining the gradient information, illumination similarity and visual masking leads to an effective scheme for quality evaluation [9].

It is well accepted that human subjects are extremely sensitive to distortions around the edge region for quality assessment tasks [19]. SSIM has been reported to be less effective for badly blurred and noisy images as it underestimates the edge damage for quality prediction [9]. Although there are several existing works incorporating the gradient information to try to improve the performance of SSIM [9], [17], [11], they simply calculate the gradient similarity and combine it with other components to compute the final quality score. Different from the previous studies, we propose a gradient-weighted SSIM (GW-SSIM) by constructing a gradient weighting map to emphasize image locations with high gradient information. Experimental results have shown that the proposed GW-SSIM achieves consistently good performance for various visual distortions.

The reminder of this paper is organized as follows: Section 2 describes the proposed GW-SSIM metric, including GW map construction, the SSIM introduction and final quality score calculation. Section 3 presents the experimental results and analysis; we conclude the paper in the final section.

II. The Proposed GW-SSIM

A. Gradient Weighting Map

In this study, we propose to incorporate the gradient information into SSIM to construct a new GW-SSIM for visual quality evaluation. We calculate the gradient weighting map as follows. First, the gradient information is computed by convolving images with Prewitt filters defined as Eq. 1.

$$\mathbf{p}_x = \begin{bmatrix} 1/3 & 0 & -1/3 \\ 1/3 & 0 & -1/3 \\ 1/3 & 0 & -1/3 \end{bmatrix}, \mathbf{p}_y = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 0 & 0 & 0 \\ -1/3 & -1/3 & -1/3 \end{bmatrix} \tag{1}$$

The gradient magnitude of the input image is computed by horizontal and vertical gradients as follows:

$$\mathbf{g}_r(i) = \sqrt{(\mathbf{r} * \mathbf{p}_x)^2(i) + (\mathbf{r} * \mathbf{p}_y)^2(i)} \tag{2}$$

$$\mathbf{g}_d(i) = \sqrt{(\mathbf{d} * \mathbf{p}_x)^2(i) + (\mathbf{d} * \mathbf{p}_y)^2(i)} \tag{3}$$

where $\mathbf{r}$ and $\mathbf{d}$ denote the reference and distorted images and $\mathbf{g}_r$, $\mathbf{g}_d$ are their corresponding gradient magnitudes with i as the location index.

Since the gradient magnitudes of the reference and distorted images are calculated, a combined gradient magnitude map is constructed as the pixel-wise maximum between the reference gradient magnitude and distorted gradient magnitude.

$$\mathbf{g}_c(i) = \max(\mathbf{g}_r(i), \mathbf{g}_d(i)) \tag{4}$$

With the proposed combination method, not only the image locations with high gradient values in the reference image have been emphasized, the edges caused by JPEG compression or other distortions in the distorted image have also been assigned with high weights.

Inspired by the previous findings that human beings are extremely sensitive to the distortions around the edge area [19], a low pass Gaussian filter with relative large standard deviation has been used to convolve with the combined gradient map to form the final GW map, as shown in Eqs. 5 and 6. Through this blur operation by Gaussian filter, the weights of small local region around the edge increase for their relative importance for quality assessment.

$$\mathbf{G}(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(x^2 + y^2)}{2\sigma^2}\right) \tag{5}$$

$$\mathbf{GW}(x, y) = \mathbf{g}_c(x, y) * \mathbf{G}(x, y) \tag{6}$$

Although the GW map construction is simple and straightforward, it has been proved to be an effective weighting map for the following reasons: (1) it's well known that gradient is an effective feature to capture local structures of images and the HVS is highly sensitive to it [9], thus more weights should be assigned to these locations; (2) the image distortion may bring in new edges that are extremely annoying for observers [19], which has also been considered in the proposed GW map; (3) the surrounding regions around the edge are also quite important for quality evaluation tasks. An illustration sample is given in Figure 1. Both the reference and distorted images are from LIVE database.

B. Gradient Weighted SSIM

SSIM predicts image quality as the product of three components in the spatial domain, *i.e.* the luminance similarity, contrast similarity and structure similarity [4]. Given two image patches $\mathbf{x}$ and $\mathbf{y}$ from the reference and distorted images, the three components are calculated as follows:

$$l(\mathbf{x}, \mathbf{y}) = \frac{2\mu_x\mu_y + c_1}{\mu_x^2\mu_y^2 + c_1} \tag{7}$$

$$c(\mathbf{x}, \mathbf{y}) = \frac{2\sigma_x\sigma_y + c_2}{\sigma_x^2\sigma_y^2 + c_2} \tag{8}$$

$$s(\mathbf{x}, \mathbf{y}) = \frac{\sigma_{xy} + c_3}{\sigma_x\sigma_y + c_3} \tag{9}$$

where μ_x, μ_y, σ_x^2, σ_y^2 and σ_{xy} represent the mean, variance and covariance of $\mathbf{x}$ and $\mathbf{y}$. c_1, c_2 and c_3 are constants to assure the numerical stability. By using a sliding window moving around the whole image, a SSIM map is constructed which can provide pixel-level quality evaluation. The overall image quality is computed as mean of the SSIM map. To account for different image resolution and viewing conditions, a multi-scale SSIM (MS-SSIM) has been developed by combining the SSIM scores at different scales where the contrast and structure similarities are computed at each scale while the illumination similarity is computed only at the coarsest scale [10].

By combining GW map with MS-SSIM, we define a gradient weighted SSIM measure (GW-SSIM). Define $\mathbf{x}_{i,j}$ and $\mathbf{y}_{i,j}$ as the ith image patches in jth scale reference and distorted images, $\mathbf{gw}_{i,j}$ is the gradient weight computed as the same location and scale, the jth scale GW-SSIM is computed as follows:

$$\text{GW-SSIM}_j = \frac{\sum_i \mathbf{gw}_{i,j} c(\mathbf{x}_{i,j}, \mathbf{y}_{i,j}) s(\mathbf{x}_{i,j}, \mathbf{y}_{i,j})}{\sum \mathbf{gw}_{i,j}} \tag{10}$$

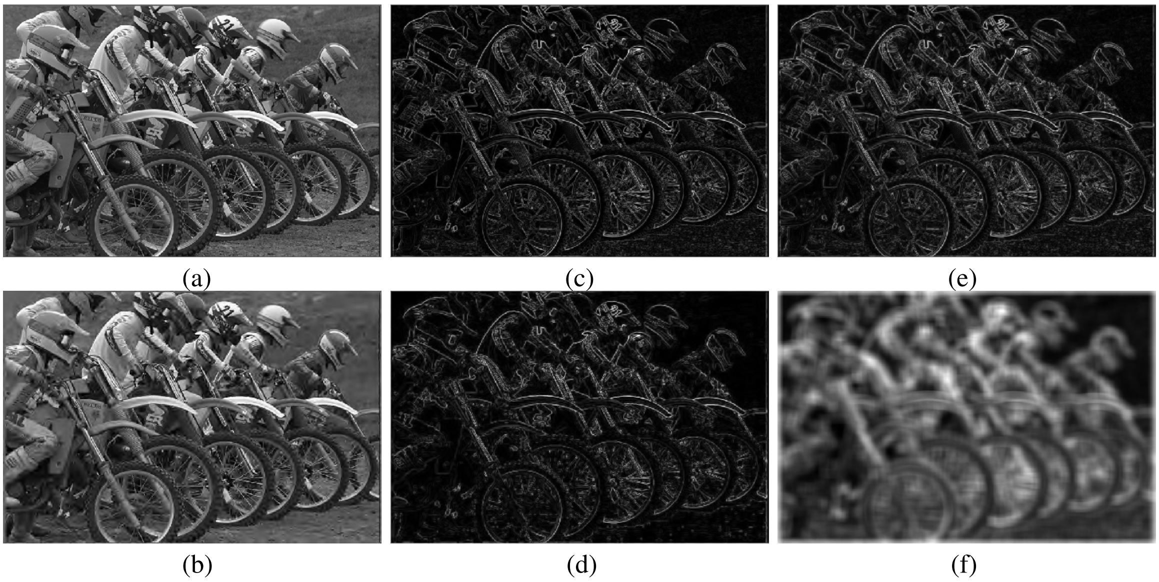

Fig. 1. Examples of GW map. (a) The reference image (bikes.bmp). (b) The JPEG 2000 compression image. (c) The gradient magnitude of the reference image. (d) The gradient magnitude of the JPEG 2000 compression image. (e) The combined gradient map. (f) The final GW map. Exemplar images are from LIVE database.

for $j = 1, ..., M-1$, and

$$\text{GW-SSIM}_M = \frac{1}{N_M}\sum_i l(\mathbf{x}_{i,M}, \mathbf{y}_{i,M})c(\mathbf{x}_{i,M}, \mathbf{y}_{i,M})s(\mathbf{x}_{i,M}, \mathbf{y}_{i,M}) \tag{11}$$

where N_M is the number of windows at the Mth scale. The overall GW-SSIM score is computed as the weighted product of GW-SSIM$_j$ as follows:

$$\text{GW-SSIM} = \prod_{j=1}^{M}(\text{GW-SSIM}_j)^{\beta_j} \tag{12}$$

where β_j is the weight at the jth scale.

A similar procedure has also been incorporated to the popular PSNR metric to form the GW-PSNR IQA metric. The parameters in GW-PSNR and GW-SSIM are set as the same with the MS-SSIM (including two constants, the windows size, five scale weights).

III. Experimental results

For performance evaluation, we compare the proposed GW-SSIM as well as GW-PSNR with eight state-of-the-art IQA metrics including PSNR, IW-PSNR [18], SSIM [4], MS-SSIM [10], IW-SSIM [18], GSSIM [11], GSD [17], and GS [9]. Among these IQA metrics, GSSIM [11], GSD [17], GS [9] are developed by including the gradient similarity into the SSIM framework, and IW-SSIM [18] utilizes an information content map to weight the MS-SSIM which shares a similar scheme with the proposed method. We tested all the IQA metrics on the LIVE database which includes 29 reference images and 779 distorted images corrupted by five types of distortions [20]. Three benchmark criteria are reported including Pearson linear correlation coefficient (PLCC), Spearman rank order correlation coefficient (SROCC) and root mean squared error (RMSE) between the subjective MOS and predicted scores [21]. A monotonic logistic function advised in [21] is used before performance comparison.

The performance on LIVE database, which consists of five types of distortions (JPEG2000 (JP2K) compression, JPEG compression, additive white Gaussian noise (AWN), Gaussian blur (GB), and a Rayleigh fast fading channel simulation (FF) noise), is shown in Table I to Table III.

TABLE I. SROCC comparison on the LIVE database. The best two IQA models are shown in boldface

IQA model	JP2K	JPEG	AWN	GB	FF	ALL
PSNR	0.895	0.881	**0.985**	0.782	0.891	0.876
IW-PSNR	0.962	0.967	**0.981**	0.937	0.833	0.933
SSIM	0.961	0.976	0.969	0.952	**0.956**	0.948
MS-SSIM	0.963	**0.981**	0.973	0.954	0.947	0.951
IW-SSIM	0.965	**0.981**	0.967	**0.972**	0.944	**0.957**
GSSIM	0.935	0.944	0.926	0.968	0.948	0.918
GSD	0.911	0.931	0.879	0.964	0.953	0.908
GS	**0.970**	0.978	0.977	0.952	0.940	0.956
GW-PSNR	0.955	0.966	0.978	0.920	0.900	0.940
GW-SSIM	**0.968**	**0.983**	0.977	**0.971**	**0.959**	**0.960**

As can be seen from these tables, the proposed GW-SSIM can effectively predict quality degradation caused by different types of distortions. It performs better than SSIM and SSIM variations under most cases. Among the SSIM-based metrics exploiting gradient inforamtion (including GSSIM, GSD, GS), GW-SSIM outperforms other exsiting metrics under most cases with only one exception that GS slightly outperforms GW-SSIM for JPEG 2000 compression. Compared with IW-SSIM that shares the same weighting strategy, GW-SSIM acheives better performance under all cases expect that both metrics perform similarly for Gaussian blur. Both the performance on individual distortion types and the overall performance on the whole LIVE database have validated the effectiveness and robustness of the proposed GW-SSIM. PSNR has always been criticized for its inconsistence with human vision perception. However, combining the GW map with PSNR has made the GW-PSNR a competitive IQA model.

TABLE II. PLCC COMPARISON ON THE LIVE DATABASE. THE BEST TWO IQA MODELS ARE SHOWN IN BOLDFACE

IQA model	JP2K	JPEG	AWN	GB	FF	ALL
PSNR	0.900	0.890	**0.988**	0.784	0.890	0.872
IW-PSNR	0.966	0.982	**0.985**	0.942	0.849	0.933
SSIM	0.967	0.979	0.983	0.948	0.955	0.945
MS-SSIM	0.969	0.983	0.984	0.956	0.947	0.949
IW-SSIM	0.973	0.982	0.982	**0.977**	0.944	**0.952**
GSSIM	0.942	0.943	0.951	0.970	0.951	0.920
GSD	0.918	0.931	0.899	0.970	**0.959**	0.913
GS	**0.976**	**0.984**	0.980	0.945	0.932	0.951
GW-PSNR	0.961	0.981	0.981	0.925	0.901	0.938
GW-SSIM	**0.975**	**0.984**	0.983	**0.973**	**0.958**	**0.955**

TABLE III. RMSE COMPARISON ON THE LIVE DATABASE. THE BEST TWO IQA MODELS ARE SHOWN IN BOLDFACE

IQA model	JP2K	JPEG	AWN	GB	FF	ALL
PSNR	11.01	14.55	**4.33**	11.46	13.01	13.36
IW-PSNR	6.49	5.93	4.87	6.20	15.03	9.84
SSIM	6.42	6.50	5.14	5.86	8.43	8.95
MS-SSIM	6.27	5.87	4.92	5.39	9.18	8.62
IW-SSIM	5.80	5.98	5.32	**3.98**	9.39	**8.35**
GSSIM	8.49	10.63	8.67	4.50	8.77	10.74
GSD	10.00	11.66	12.27	4.53	**8.11**	11.15
GS	**5.52**	**5.67**	5.57	6.03	10.29	8.43
GW-PSNR	7.01	6.15	5.50	7.02	12.35	9.50
GW-SSIM	**5.65**	**5.64**	**4.77**	**4.27**	**8.20**	**8.09**

IV. CONCLUSION

In this work, we have proposed a simple yet effective gradient weighting strategy to improve the performance of existing metrics for visual quality evaluation. The promising performance of GW-SSIM and GW-PSNR demonstrates that including the gradient structural information in a final weighting procedure for IQA is consistent with human perception for quality evaluation. In the proposed metric, three factors are considered into the GW map: the reference image gradient, the gradient caused by various distortions and the small local region around the edges. In the future, we would like to exploit the applications of the proposed GW map in other existing and emerging IQA metrics.

ACKNOWLEDGMENT

This work was funded by the Ph.D. Grant from the Institute for Media Innovation, Nanyang Technological University, Singapore.

REFERENCES

[1] W. Lin and C.-C. Jay Kuo, “Perceptual visual quality metrics: A survey,” *Journal of Visual Communication and Image Representation*, vol. 22, no. 4, pp. 297–312, 2011.

[2] S. J. Daly, “Visible differences predictor: an algorithm for the assessment of image fidelity,” vol. 1666, Conference Proceedings, pp. 2–15, 10.1117/12.135952. [Online]. Available: http://dx.doi.org/10.1117/12.135952

[3] J. Lubin and D. Fibush, “Sarnoff jnd vision model,” pp. 97–612, 1997.

[4] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *Image Processing, IEEE Transactions on*, vol. 13, no. 4, pp. 600–612, 2004.

[5] H. R. Sheikh, A. C. Bovik, and G. de Veciana, “An information fidelity criterion for image quality assessment using natural scene statistics,” *Image Processing, IEEE Transactions on*, vol. 14, no. 12, pp. 2117–2128, 2005.

[6] H. R. Sheikh and A. C. Bovik, “Image information and visual quality,” *Image Processing, IEEE Transactions on*, vol. 15, no. 2, pp. 430–444, 2006.

[7] E. C. Larson and D. M. Chandler, “Most apparent distortion: full-reference image quality assessment and the role of strategy,” *Journal of Electronic Imaging*, vol. 19, no. 1, pp. 011 006–011 006–21, 2010.

[8] L. Zhang, L. Zhang, X. Mou, and D. Zhang, “Fsim: A feature similarity index for image quality assessment,” *Image Processing, IEEE Transactions on*, vol. 20, no. 8, pp. 2378–2386, 2011.

[9] A. Liu, W. Lin, and N. Manish, “Image quality assessment based on gradient similarity,” *Image Processing, IEEE Transactions on*, vol. 21, no. 4, pp. 1500–1512, 2012.

[10] Z. Wang, E. P. Simoncelli, and A. C. Bovik, “Multiscale structural similarity for image quality assessment,” in *Signals, Systems and Computers, 2004. Conference Record of the Thirty-Seventh Asilomar Conference on*, vol. 2, Conference Proceedings, pp. 1398–1402 Vol.2.

[11] G.-H. Chen, C.-L. Yang, and S.-L. Xie, “Gradient-based structural similarity for image quality assessment,” in *Image Processing, 2006 IEEE International Conference on*, Conference Proceedings, pp. 2929–2932.

[12] M. P. Sampat, W. Zhou, S. Gupta, A. C. Bovik, and M. K. Markey, “Complex wavelet structural similarity: A new image similarity index,” *Image Processing, IEEE Transactions on*, vol. 18, no. 11, pp. 2385–2401, 2009.

[13] C.-L. Yang, W.-R. Gao, and L.-M. Po, “Discrete wavelet transform-based structural similarity for image quality assessment,” in *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*, Conference Proceedings, pp. 377–380.

[14] D. M. Rouse and S. S. Hemami, “Understanding and simplifying the structural similarity metric,” in *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*, Conference Proceedings, pp. 1188–1191.

[15] W. Fu, X. Gu, and Y. Wang, “Image quality assessment using edge and contrast similarity,” in *Neural Networks, 2008. IJCNN 2008. (IEEE World Congress on Computational Intelligence). IEEE International Joint Conference on*, Conference Proceedings, pp. 852–855.

[16] C. Li and A. C. Bovik, “Content-partitioned structural similarity index for image quality assessment,” *Signal Processing: Image Communication*, vol. 25, no. 7, pp. 517–526, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0923596510000354

[17] G. Cheng, J. Huang, C. Zhu, Z. Liu, and L. Cheng, “Perceptual image quality assessment using a geometric structural distortion model,” in *Image Processing (ICIP), 2010 17th IEEE International Conference on*, Conference Proceedings, pp. 325–328.

[18] Z. Wang and Q. Li, “Information content weighting for perceptual image quality assessment,” *Image Processing, IEEE Transactions on*, vol. 20, no. 5, pp. 1185–1198, 2011.

[19] W. Lin, L. Dong, and P. Xue, “Visual distortion gauge based on discrimination of noticeable contrast changes,” *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 15, no. 7, pp. 900–909, 2005.

[20] H. Sheikh, Z. Wang, L. Cormack, and A. Bovik, “Live image quality assessment database release 2 (2005).”

[21] VQEG, “Final report from the video quality experts group on the validation of objective models of video quality assessment,” 2000.