Center-Emphasized Gradient-Feature Based Quality Assessment Method for Screen Content Images

Md. Abu Layek, Ngo Thien Thu, SoYeon Yu, TaeChoong Chung and Eui-Nam Huh$^{\boxtimes}$

Kyung Hee University Global Campus

Yongin, South Korea-17104

Email: {layek, thu.ngo, loverduck97, tcchung, johnhuh}@khu.ac.kr

Abstract—Screen content images (SCI) are composed of texts, graphics, images, emails, web pages, and other computer-generated components which make it different from the pure natural images. As a result, image quality assessment (IQA) methods designed for natural images are not always suitable for screen content images. However, as time goes, many screen contents are being generated which share several properties of natural images. As a result, IQA methods for SCI need to be more generic while giving priority to the prominent SCI features. The image gradient is a very important feature for all kinds of images, especially for SCI. In this paper, we proposed a novel image gradient based similarity index (CGSI) for screen content images where the gradient is used both as a quality map and extractor for another feature map. Our previous studies show that the center part of an image is visually the most important part and HVS is more sensitive to any distortion in the middle area. To address this issue, the center area of both gradient and feature similarity maps are raised by element-wise squaring. Eventually, the final quality score is calculated using a summation of standard deviation based pooling strategy. We evaluated our proposed method on three large scale SCI databases and compared with 8 other state-of-the-art IQA methods designed for both natural and screen content images. Results show that the proposed approach provides competitive performance and defeats all other methods with a large gap in overall performance. Also, the relatively faster running time makes it suitable for most of the real-time applications.

Index Terms—Image quality assessment, Screen contents images, Center-emphasized

I. INTRODUCTION

With the increasing development of mobile technologies, different portable and hand-held devices are becoming part and parcel of our daily life. Mobile cloud computing (MCC) [1] along with the thick and thing client solutions further enhanced the capabilities of these devices and enables them to perform heavier workloads. People are now can work, play games, surf web pages, advertise or stream from anywhere anytime. All these processes involve the huge generation and transmission of screen contents images (SCIs). As a result, service providers need to know the qualities of SCIs in order to maintain the user's quality of experience (QoE).

Image quality assessment (IQA) has gained a lot of researchers attention for decades and we have been witnessed several milestones works in the course of this time. However, SCI should be dealt with differently because it has several easily distinguishable characteristics[2]:

- SCIs contain various types of contents together such as texts, natural images, drawings, graph/charts, icons etc.
- SCIs are computer generated without using general image acquisition tools as a result source images does not contain any noises.
- Most of the SCIs have higher contrast, sharp edges and a large part with homogeneous regions.

While the difference between SCIs and natural images are clear, day by day SCIs is becoming more diverse as a result more general quality assessment methods are needed for SCI.

In the literature, IQA methods are classified in three broad categories: No-reference IQA where only the distorted images are given; Reduced-reference IQA where partial information of the reference images are given; and the Full-reference IQA where full images of both references and distorted images are available. The center focus of this paper is the full-reference IQA (FR-IQA).

Although the classical pixels-based IQA methods such as PSNR or MSE [3] have been widely used, they have failed to reflect the human visual system (HVS) and bear poor correlation with human evaluated subjective scores (MOS/DMOS). Then the milestone work of structural similarity index (SSIM) [4] brought IQA from pixels-level to structure levels which reflects HVS more than previous methods. Since then, scores of methods has been devised considering different features of images in relation to HVS and recent methods provide very good correlation with human evaluated scores.

Xue et al. proposed the gradient magnitude similarity deviation (GMSD) [5] considering the fact that image gradients are sensitive to image distortions. The mean deviation similarity index (MDSI)[6] modified the gradient similarity map through a fusion technique while Wang et al. in their multi-scale contrast similarity deviation (MCSD) [7] considered multi-scale root men square contrast and achieved better performances for natural images. All these three methods are very fast because they utilize the faster image gradient calculation followed by a standard deviation (SD) pooling. Reisenhofer et al. proposed a Haar wavelet-based perceptual similarity index (HPSI) [8] using a Haar wavelet decomposition to assess local similarities between two images. Sun et al. proposed superpixel-based similarity index for full-reference image quality assessment (SPSIM) [9] which they found better reflects HVS. Although these methods work well with satisfactory performance for

natural images, they don't behave in the same way with screen content images.

In this scenario, several IQA methods specially designed for SCIs are also available. Ni et al. proposed a gradient similarity based SCI quality assessment [10]. They use gradient magnitude and an SD pooling to calculate the final score. The research group who developed the SIQAD [11] database, proposed RWQMS [12] which is trained on SIQAD.

In this paper, we propose a full-reference IQA for SCI named as CGSI using image gradient and a feature similarity extracted with the gradient map of the reference image. After computing both of the gradient similarity and feature similarity maps, the center parts are modified as suggested in [13]. Finally, a summation of SD pooling is used to calculate the CGSI score. The proposed IQA approach is compared with 8 state-of-the-art methods among them 6 are general IQA and 2 are especially proposed for SCI. To ensure the comparison as fair as possible, we evaluated all methods on three large scale and publicly available SCI databases, SIQAD[11], SCID[14] and QACS [12]. Results show that our proposed method gives satisfactory performance on all of the databases which stand top of the list in overall score prediction.

We organize the remaining of this paper as follow. Section II shows the proposed method overview with a brief mathematical definition. The experimental results are shown in Section III. Finally, we conclude our paper in Section IV.

II. PROPOSED IMAGE QUALITY ASSESSMENT METHOD
A. Overview

In Figure 1, we sketch out the overview of our proposed SCI quality assessment method. At first, the gradient maps of the reference (r) and distorted (d) SCIs calculated. We use the prewitt filters in horizontal and vertical directions p_x and p_y as defined in Equation 1.

$$p_x = \begin{bmatrix} 1/3 & 0 & -1/3 \\ 1/3 & 0 & -1/3 \\ 1/3 & 0 & -1/3 \end{bmatrix}, p_y = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 0 & 0 & 0 \\ -1/3 & -1/3 & -1/3 \end{bmatrix} \quad (1)$$

Then, the gradient maps in the form of gradient magnitude of r and d denoted by $G_r(x,y)$ and $G_d(x,y)$ respectively and are defined as

$$G_r(x,y) = \sqrt{(r \circledast p_x)^2 + (r \circledast p_y)^2}$$
$$G_d(x,y) = \sqrt{(d \circledast p_x)^2 + (d \circledast p_y)^2} \quad (2)$$

where $\circledast$ is the convolution operation. Using the gradient magnitudes, the gradient similarity map $GS(x,y)$ is computed as

$$GS(r,d) = \frac{2G_r(x,y) \odot G_d(x,y) + c_1}{G_r(x,y)^2 + G_d(x,y)^2 + c_1} \quad (3)$$

where $\odot$ is the element-wise multiplication and $c1$ is a positive constant to increase the calculation stability.

Now, the feature maps of r and d are computed by

$$F1_r(x,y) = G_r(x,y) \odot \left[r \odot \{Med(e) \times c2 + \epsilon\} \right]$$
$$F1_d(x,y) = G_r(x,y) \odot \left[d \odot \{Med(e) \times c2 + \epsilon\} \right] \quad (4)$$

where $c2$ is a positive constant, ϵ is a very small value used to avoid zero results, Med returns median value from a vector and e is the mean absolute error (MAE) defined as

$$e = \frac{1}{n} \sum_{i=1}^{n} | r1_i - d1_i | \quad (5)$$

where n is the number of elements in each column and $r1$ and $d1$ are the column-wise mean-subtracted reference and distorted images respectively.

Then, we compute a threshold value t which we will use to select important features

$$t = \sum_{i=1}^{m} (\sum_{j=1}^{n} F1_r^2) \times c3 \quad (6)$$

where m, n are the number of elements in each row and column respectively; and $c3$ is a positive constant. The feature maps are then calculated as

$$F_r(x,y) = F1_r(x,y) \geq t$$
$$F_d(x,y) = F1_d(x,y) \geq t \quad (7)$$

After computing both of the feature maps we calculate their similarity map as

$$FS(r,d) = \frac{2F_r(x,y) \odot F_d(x,y) + c_4}{F_r(x,y)^2 + F_d(x,y)^2 + c_4} \quad (8)$$

where $c4$ is a positive constant.

B. Center Emphasized Similarity

At this stage, we will increase the sensitivity of center parts in both of the similarity maps. If the original dimension is $(H \times W)$, then the corresponding dimension for the center block becomes $(H_{mid} \times W_{mid})$, where:

$$H_{mid} = \left\lceil \frac{H}{2} \right\rceil \quad \text{and} \quad W_{mid} = \left\lceil \frac{W}{2} \right\rceil. \quad (9)$$

The center block is defined as a rectangular area identified by two corner points (x_{min}, y_{min}) and (x_{max}, y_{max}), where:

$$x_{min} = \left\lceil \frac{H}{4} \right\rceil$$
$$y_{min} = \left\lceil \frac{W}{4} \right\rceil$$
$$x_{max} = \left\lceil \frac{H}{4} \right\rceil + H_{mid}$$
$$y_{max} = \left\lceil \frac{W}{4} \right\rceil + W_{mid} \quad (10)$$

If we denote the center part of the similarity maps as $GS(mid)$ and $FS(mid)$, then the updated center parts $CGS_{(mid)}$ and $CFS_{(mid)}$ respectively are defined as
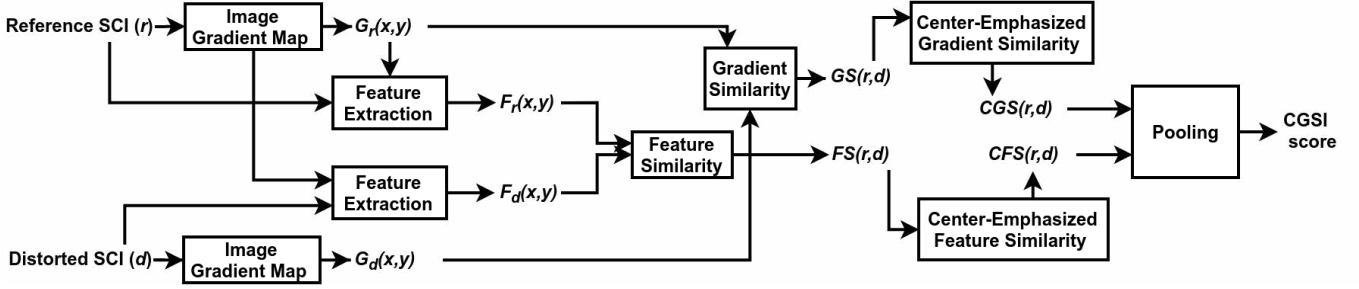
Fig. 1: Flow diagram of the proposed approach.

$$CGS(mid) = GS(mid) \odot GS(mid)$$
$$CFS(mid) = FS(mid) \odot FS(mid) \qquad (11)$$

where $\odot$ is the element-wise multiplication. With the updated center region we get the center-emphasized similarity maps denoted by $CGS(r,d)$ and $CFS(r,d)$.

C. Sum of Deviation based pooling strategy

Standard deviation based pooling achieves very good performance in specific cases and is adopted by several successful methods as mentioned in the introduction. We also adopted SD pooling strategy due to its' suitability with gradient-based methods. However, now we have two similarity metrics, as a result, we employ a summation of standard deviation based approach similar to [13]. Using this strategy the final quality score ($CGSI$) is calculated as:

$$CGSI = \left[W1 \times std\{CGS(r,d)\} + W2 \times std\{CFS(r,d)\} \right] \qquad (12)$$

where $W1$ and $W2$ are weighting factors that specify the importance of GS and FS, respectively. The standard deviations in the above equation are defined as:

$$std(CGS(r,d)) = \left\{ \frac{1}{M} \sum_{i=1}^{M} (CGS_i - \mu_{CGS})^2 \right\}^{\frac{1}{2}}$$
$$std(CFS(r,d)) = \left\{ \frac{1}{M} \sum_{i=1}^{M} (CFS_i - \mu_{CFS})^2 \right\}^{\frac{1}{2}} \qquad (13)$$

where M is the number of total elements in the similarity maps; CGS_i and CFS_i are the ith items; μ_{CGS} and μ_{CFS} are the mean values of the $CGS(r,d)$ and $CFS(r,d)$.

The MATLAB implementation of our approach will be available on-line at http://layek.khu.ac.kr/CGSI.

III. EXPERIMENTAL RESULTS

We performed evaluations and compared our proposed method with 8 state-of-the art methods SSIM [4], GMSD[5], MDSI[6], MCSD[7], HPSIe[8], SPSIM[9] , RWQMS[12] and SCIGSS[10].

A. Description of Databases

The objective IQA methods are evaluated on three large scale SCI databases SIQAD[11], SCID[14] and QACS [12]. The general description of the databases are given in Table I.

TABLE I: Basic information about the databases used for the experiments.

Dataset	Reference Images	Distorted Images	Distortion Types	No. of Subjects
SIQAD	20	980	7	96
SCID	40	1800	9	186
QACS	24	492	2	20

SIQAD is a state-of-the-art database for SCI which is used by many researchers. It consists of 20 references and 980 distorted SCIs obtained by adding 7 different types of noises. The distortion types are the Gaussian noise (GN), Gaussian blur (GB), motion blur (MB), contrast change (CC), JPEG compression (JPEG), JPEG2000 compression (J2K), and the layer segmentation-based coding (LSC).

SCID is the latest and the biggest database we found, it contains 40 reference and 1800 distorted images with 9 types of distortions; Gaussian noise (GN), Gaussian blur (GB), motion blur (MB), contrast change (CC), color saturation change (CSC), color quantization with dithering (CQD), JPEG compression (JPEG), JPEG2000 compression (J2K), and HEVC-SCC. The number of observers in SCID is 186 which makes it more reliable.

The QACS database is the simplest among these three, consisting of 24 reference images; HEVC and the SCC extension HEVC-SCC is used to prepare the 492 distorted images.

B. Evaluation Metrics

The performance is measured using some correlation measurements with the subjective scores or human evaluated values. However, to apply the linear correlation, the two compared values should be on the same scale and perfectly linearly correlated [15]. For this purpose, a logistic mapping function is used to convert the objective scores and we use the following nonlinear regression model as suggested by Sheikh [16].

$$mps = \beta_1 \left\{ \frac{1}{2} - \frac{1}{1 + \exp(\beta_2(obs - \beta_3))} \right\} + (\beta_4 \times obs) + \beta_5 \qquad (14)$$

where obs is the objective score, mps is the mapped value, and β_i are the 5 parameters that are tuned based on the relationship between objective and subjective scores. To

TABLE II: Performance comparison of IQA methods on three databases.

Dataset	Metric	SSIM	GMSD	MDSI	MCSD	HPSIe	SPSIM	RWQMS	SCIGSS	CGSI
SIQAD	SROCC	0.5836	0.7318	0.5602	0.7331	0.7092	0.5844	**0.7815**	**0.8438**	**0.8238**
	KROCC	0.4235	0.5521	0.4073	0.5632	0.5304	0.4221	**0.5835**	**0.6488**	**0.6276**
	PLCC	0.5912	0.7426	0.5663	0.7432	0.7142	0.5909	**0.8103**	**0.8516**	**0.8363**
	RMSE	11.5450	9.5859	11.7977	9.5768	10.0189	11.5477	**8.3892**	**7.5032**	**7.8481**
SCID	SROCC	0.7146	0.8138	0.7814	**0.8147**	**0.8573**	0.7863	0.7162	0.7484	**0.8231**
	KROCC	0.5136	0.6183	0.5854	**0.6209**	**0.6606**	0.5859	0.5169	0.5417	**0.6316**
	PLCC	0.7328	0.8337	0.7883	**0.8339**	**0.8556**	0.7913	0.7353	0.7602	**0.8476**
	RMSE	9.6359	7.8204	8.7135	**7.8163**	**7.3314**	8.6594	9.5979	9.2006	**7.5157**
QACS	SROCC	0.8829	0.8769	0.8526	**0.8900**	0.8894	**0.9004**	0.8502	0.8621	0.8692
	KROCC	0.7072	0.7010	0.6709	**0.7175**	0.7172	**0.7282**	0.6604	0.6800	0.6879
	PLCC	0.8764	0.8746	0.8564	**0.8860**	**0.8893**	**0.8964**	0.8488	0.8653	0.8702
	RMSE	1.0684	1.0755	1.1456	**1.0289**	**1.0146**	**0.9835**	1.1730	1.1119	1.0932
OVERALL	SROCC	0.7007	0.7987	0.7259	**0.8016**	**0.8178**	0.7430	0.7559	0.7941	**0.8302**
	KROCC	0.5157	0.6109	0.5449	**0.6181**	**0.6301**	0.5582	0.5584	0.5946	**0.6389**
	PLCC	0.7120	0.8126	0.7320	**0.8146**	**0.8183**	0.7471	0.7748	0.8034	**0.8476**
	RMSE	8.9194	7.3350	8.4993	**7.3230**	**7.1865**	8.3703	7.9691	7.4759	**6.6495**

– For each row, the first, second and third-ranked performances are highlighted respectively in blue, red and black colors.

– For SROCC, KROCC and PLCC metrics, the higher the value, the better the method whereas for RMSE a lower score is better.

find the optimal parameters the built-in MATLAB function $nlinfit$ is used. The subjective scores are then used with these mapped scores to find the following Pearson's linear correlation coefficient (PLCC) which is defined as follows:

$$PLCC(q,s) = \frac{\sum_{i=1}^{m}(q_i - \mu_q)(s_i - \mu_s)}{\left\{\sum_{i=1}^{m}(q_i - \mu_q)^2\right\}^{\frac{1}{2}}\left\{\sum_{i=1}^{m}(s_i - \mu_s)^2\right\}^{\frac{1}{2}}} \quad (15)$$

where q and s are vectors of the objective and subjective scores, respectively; μ_o and μ_s are their mean scores, and m is the number of distorted images. The objective scores of q are actually the mapped scores using Equation (14). If we want to avoid this nonlinear mapping, rank order coefficients can be used. The popular Spearman's rank-order correlation coefficient (SROCC) is given as:

$$SROCC(q,s) = PLCC(rank(q), rank(s)) \quad (16)$$

The function $rank()$ returns a rank-vector when applied on a score vector. The i-th entry of a rank-vector contains the relative rank of the i-th item in the original score vector.

Another widely adopted rank order metric is the Kendall's rank-order correlation coefficient (KROCC), which is given as below:

$$KROCC(q,s) = \frac{C-D}{m(m-1)/2} \quad (17)$$

where C is the number of concordant pairs and D is the number of discordant pairs.

The root mean square error (RMSE) is defined as:

$$RMSE(q,s) = \left\{\frac{1}{m}\sum_{i=1}^{m}(q_i - s_i)^2\right\}^{\frac{1}{2}} \quad (18)$$

In the cases of PLCC, SROCC, and KROCC, a larger value is an indicator of superior method whereas a smaller value of the RMSE refers to a better IQA. Again, among these four metrics, SROCC is regarded as the most important correlation measurement.

C. Performance Comparison

TABLE III: Running time comparison of the IQA models.

IQA	Running Time (ms)	Images Per Second
SSIM	23.00	43.38
GMSD	14.21	70.37
MDSI	33.47	29.88
MCSD	20.97	47.68
HPSIe	115.78	8.64
SPSIM	154.65	6.47
RWQMS	141.81	7.05
SCIGSS	125.43	7.97
CGSI	45.82	21.82

Table II shows the results on three benchmark databases among different IQA models for all of the four metrics mentioned in Section III-B. The top three values are boldfaced where the best performance is marked with blue color, then red and the third highest is black. However, in the case of RMSE, the lowest value is colored in blue, because a lower RMSE refers to a better method. We see that three different methods perform the best for the three different databases; SCIGSS performed best on SIQAD, HPSIe on SCID and SPSIM on QACs. However, the proposed CGSI performs consistently good on all of the databases and achieves the highest position for all metrics in the overall score. The overall score is calculated by weighted averages of the SROCC, KROCC, PLCC, and RMSE using the number of distorted images, as
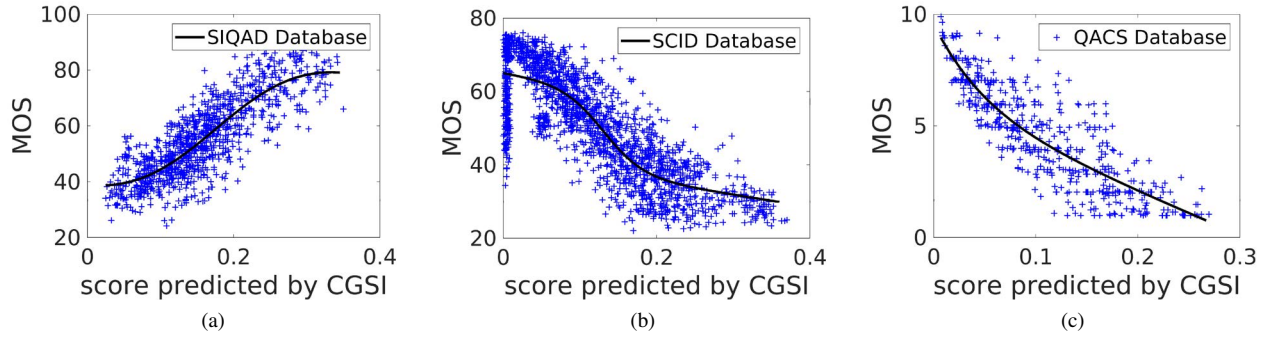
Fig. 2: Scatter plot of the subjective scores versus scores predicted by CGSI on three databases. The black curves are obtained using the nonlinear fitting of Equation 14.

suggested in [17]. From the table, we can order the top 5 SCI quality assessment methods as [CGSI, HPSIe, MCSD, GMSD, SCIGSS]. SCIGSS performed very well on SIQAD but below average on the most diverse database SCID. On the other hand, CGSI make use of the two most powerful SCI features i.e. image gradient and feature similarity which makes it a more general SCI assessment method. Moreover, the center-emphasis further improved the performance. On the overall score, compared to the RWQMS and SCIGSS, our CGSI has shown better prediction accuracy with (7.43%, 3.61%)-point, (8.04%, 4.43%)-point and (7.27%, 4.42%)-point higher overall SROCC, KROCC and PLCC values, respectively. This can be termed as a very big improvement for SCI quality assessment methods.

The scatter plots in Figure 2 demonstrates the predicted scores for CGSI on the three databases. The black curves are obtained using the nonlinear fitting of Equation 14 . We see that our proposed approach is quite consistent in predicting scores and fitted nicely for SIQAD and QACS. On SCID database, we see that several scores are mispredicted as 0 by our method. SCID is a big and diverse database with a large number of human observers and our method predicts wrongly for a few SCIs. If these issues can be corrected, it will be improved more and we left this as our future task.

Although the principal focus of designing an IQA model is the performance of its prediction, a low computational cost is also a desirable feature, especially for a real-time system. In the case of SCI, sometimes a low computational IQA is much necessary to provide a low-response service to the client in some applications such as desktop as a service, real-time adaptive streaming etc. We performed the run-time comparison on various SCI-QA models with MATLAB R2017b using a computer equipped with an Intel(R) Core(TM) i5-4670 CPU with a 3.40GHz processor and 16GB of RAM. The MATLAB codes provided by the authors were used and elapsed time was recorded using the $tic - toc$ functions; Table III shows the results. We see that CGSI needed 45.82 milliseconds to process a single image with a rate of 21.82 images per second. This rate meets the need for most of the real time applications.

IV. CONCLUSION

We proposed a novel IQA method for screen content images using image gradient and feature similarity where the features are extracted using the gradient map of the reference images. Extensive evaluations are done on three large scale public databases of screen content images and the proposed method is compared with 8 state-of-the-art approaches. Results suggest that the proposed CGSI performs satisfactorily on all databases and overall it outperforms all of the compared methods. The running time is also suitable for most of the real-time applications. The scatter plots with subjective scores show that, for SIQAD and QACS databases, the proposed approach predicted the scores nicely and consistently, however, for SCID database, CGSI mispredicted some scores near to 0. We will investigate more on this issue and try to improve this method in our next work.

ACKNOWLEDGMENT

REFERENCES

[1] D. Huang *et al.*, "Mobile cloud computing," *IEEE COM-SOC Multimedia Communications Technical Committee (MMTC) E-Letter*, vol. 6, no. 10, pp. 27–31, 2011.

[2] S. Shi, X. Zhang, S. Wang, R. Xiong, and S. Ma, "Study on subjective quality assessment of screen content images," in *2015 Picture Coding Symposium (PCS)*. IEEE, 2015, pp. 75–79.

[3] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? a new look at signal fidelity measures," *IEEE signal processing magazine*, vol. 26, no. 1, pp. 98–117, 2009.

[4] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to struc-

tural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[5] W. Xue, L. Zhang, X. Mou, and A. C. Bovik, "Gradient magnitude similarity deviation: A highly efficient perceptual image quality index," *IEEE Transactions on Image Processing*, vol. 23, no. 2, pp. 684–695, 2014.

[6] H. Z. Nafchi, A. Shahkolaei, R. Hedjam, and M. Cheriet, "Mean deviation similarity index: Efficient and reliable full-reference image quality evaluator," *IEEE Access*, vol. 4, pp. 5579–5590, 2016.

[7] T. Wang, L. Zhang, H. Jia, B. Li, and H. Shu, "Multiscale contrast similarity deviation: An effective and efficient index for perceptual image quality assessment," *Signal Processing: Image Communication*, vol. 45, pp. 1–9, 2016.

[8] R. Reisenhofer, S. Bosse, G. Kutyniok, and T. Wiegand, "A haar wavelet-based perceptual similarity index for image quality assessment," *Signal Processing: Image Communication*, vol. 61, pp. 33–43, 2018.

[9] W. Sun, Q. Liao, J.-H. Xue, and F. Zhou, "Spsim: A superpixel-based similarity index for full-reference image quality assessment," *IEEE Transactions on Image Processing*, vol. 27, no. 9, pp. 4232–4244, 2018.

[10] Z. Ni, L. Ma, H. Zeng, C. Cai, and K.-K. Ma, "Gradient direction for screen content image quality assessment," *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1394–1398, 2016.

[11] H. Yang, Y. Fang, and W. Lin, "Perceptual quality assessment of screen content images," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 4408–4421, 2015.

[12] S. Wang, K. Gu, X. Zhang, W. Lin, L. Zhang, S. Ma, and W. Gao, "Subjective and objective quality assessment of compressed screen content images," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 6, no. 4, pp. 532–543, 2016.

[13] M. Layek, A. Uddin, T. P. Le, T. Chung, E.-N. Huh *et al.*, "Center-emphasized visual saliency and a contrast-based full reference image quality index," *Symmetry*, vol. 11, no. 3, p. 296, 2019.

[14] Z. Ni, L. Ma, H. Zeng, J. Chen, C. Cai, and K.-K. Ma, "Esim: Edge similarity for screen content image quality assessment," *IEEE Transactions on Image Processing*, vol. 26, no. 10, pp. 4818–4831, 2017.

[15] Y. Ding, "General framework of image quality assessment," in *Visual Quality Assessment for Natural and Medical Image*. Springer, 2018, pp. 45–62.

[16] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Transactions on image processing*, vol. 15, no. 11, pp. 3440–3451, 2006.

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