REDUCED-REFERENCE QUALITY METRIC FOR SCREEN CONTENT IMAGE

Zhaohui Che[†], *Guangtao Zhai*[†], *Ke Gu*[‡], *and Patrick Le Callet*[§]

[†]Insti. of Image Commu. and Infor. Proce., Shanghai Jiao Tong University, China
[‡]Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China
[§]Luman Université, Université de Nantes, IRCCyN UMR CNRS 6597, Polytech Nantes, France
Email: chezhaohui@sjtu.edu.cn

ABSTRACT

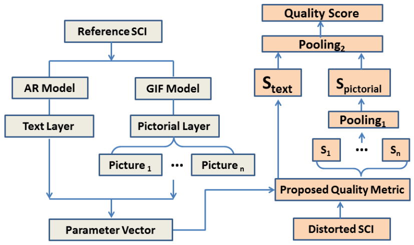
With the prevalence of digital products like cellphone, tablet and personal computer, the screen content image (SCI) consisting of text, graphic, and natural scene picture becomes a significant media in various communication scenarios. Consequently, we proposed a reduced-reference quality metric dedicated for SCI. The main contribution includes 2 aspects : **1)** we innovatively proposed a layer-based segmentation method to divide SCI into text layer and pictorial layer; **2)** we designed respective quality metrics dedicated for text and pictorial layers with a novel pooling strategy considering human visual saliency for SCI. Furthermore, exhaustive experimental results indicate that the proposed metric is highly comparative compared with state-of-the-art full-reference quality metrics.

Index Terms— Screen content, IQA, free energy

1. INTRODUCTION

Recently, the ubiquitous screen content images (SCIs) play a critical role in diverse scenarios such as remote conferencing [1], cloud transformation [2], etc. However, most consumer-type SCIs are captured by amateurish devices which corrupt the SCIs with various distortions, so that the quality of SCI becomes an attractive issue for consumers.

A plenty of image quality assessment (IQA) metrics have been proposed in the past decades. Considering the complex components of SCIs, i.e. text area, graphic area, and natural scene picture area, the traditional IQA metrics fail to evaluate these informative compound images. Notably, a few new quality metrics dedicated for SCIs were proposed in recent years. Specifically, Yang *et al.* proposed a full-reference quality metric SPQA [3] in 2015. The SPQA developed the classic SSIM metric [4], and used it to estimate the text regions and pictorial regions separately. Wang *et al.* came up with another full-reference SCI quality metric $\mathrm{Q_s}$ [5] based on SSIM. Shao *et al.* proposed a blind quality predictor [22] for SCIs using local and global sparse representation. Gu *et al.* proposed the full-reference SQMS [6] and no-reference BIQME [7] and SINE [8] in recent two years.



(a)

Fig. 1. Flowchart of the proposed quality metric.

Considering the drawbacks and flaws of the existing SCI quality metrics, we proposed a reduced-reference SCI quality assessment metric in this paper. The main contribution is divided into two aspects. Firstly, we proposed a novel layer-based segmentation algorithm to divide SCI into text and pictorial layers, so as to extract the accurate location, size and inclination angle of each picture located at the same SCI. Secondly, we designed two different reduced-reference quality metrics dedicated for pictorial and text layers. Specifically, incorporating with prior information of human visual saliency when viewing SCI, we proposed a novel pooling strategy and obtained the ultimate quality score of the SCI.

The rest of this paper is organized as follows. In section 2, we elaborated the proposed segmentation algorithm and quality metric in detail. The exhaustive experimental results were exhibited in section 3 for validating performance. We drew conclusions and gave out future directions in section 4.

2. REDUCED-REFERENCE QUALITY ASSESSMENT METRIC FOR SCREEN CONTENT IMAGE

2.1. Layer-based Segmentation Method

Above all, SCIs usually contain text areas and pictorial areas. In addition, most SCIs always contain several natural scene pictures scattered at arbitrary positions. Notably, text areas contain thin edges and small base colour numbers, while pictorial areas have thick boundaries and abundant colour. Accordingly, human visual perceptions of text and pictorial areas depend on different features. Driven by designing appropria-

978-1-5090-2175-8/17/$31.00 ©2017 IEEE

tive quality measurements for different areas while keeping the integrity of each picture located at the same SCI, we innovatively proposed a layer-based segmentation method.

The SCI segmentation methods have been investigated in recent years [10, 11, 12, 13] especially for SCI compression [21]. However, block-based methods [10, 11] destroy the integrity of pictures, while layer-based methods [12, 13] are not able to differentiate textual details of pictorial and text areas. Herein, we proposed a new segmentation strategy to divide a SCI into five layers, i.e. smooth background layer (SBL), smooth pictorial layer (SPL), textural pictorial layer (TPL), smooth text layer (STL), and textural text layer (TTL). Subsequently, we adopted spatial filtering techniques to extract TPL, TTL and SBL as follows.

Considering the heuristic information that text layer contains many short steep edges while pictorial layer contains chaotic texture details and thick boundaries, we adopted the autoregressive (AR) model [14, 15] and guided image filter (GIF) [16] to process the SCI respectively, because the AR model has good texture-preserving ability while the GIF is good at preserving edges. The AR model specifies that the output depends linearly on its own previous variable value and on a stochastic term. In digital image processing, this relationship can be expressed by equation 1.

$$y_i = \boldsymbol{\alpha} \times \boldsymbol{\gamma}^k(y_i) + \varepsilon_i \tag{1}$$

where y_i is the pixel value to be processed; $\boldsymbol{\alpha} = \{\alpha_1, ..., \alpha_k\}$ is the vector of AR coefficients; $\boldsymbol{\gamma}^k(y_i)$ means the k member neighborhood vector of y_i; ε_i is the difference between ground truth and predicted value. The parameter $\boldsymbol{\alpha}$ can be solved via the linear system:

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} ||\boldsymbol{y} - \boldsymbol{Y}\boldsymbol{\alpha}||_2 \tag{2}$$

where $\boldsymbol{Y(i,:)} = \gamma^k(y_i)$ and $\boldsymbol{y} = (y_1, y_2, ..., y_k)$. We can solve this linear system by least square method and obtain the approximate solution as $\hat{\boldsymbol{\alpha}} = \boldsymbol{(Y^TY)}^{-1}\boldsymbol{Y^Y y}$. The AR model can protect pictorial details well, but it performs poorly on steep edges of text. On the other hand, the GIF can generate output according to the guide image. GIF behaves as an efficient edge-preserving smoothing operator when the guide image is identical to the original input image.

After obtaining AR model filtering result I_{ar} and GIF filtering result I_{gif}, we calculated the coarse $\overline{TTL}$ and $\overline{TPL}$ by equation 3.

$$\begin{cases} \overline{\boldsymbol{TTL}} = 1 - \boldsymbol{N}(\mathrm{SSIM}(\boldsymbol{I_{ar}}, \boldsymbol{I_{input}})) \\ \overline{\boldsymbol{TPL}} = 1 - \boldsymbol{N}(\mathrm{SSIM}(\boldsymbol{I_{gif}}, \boldsymbol{I_{input}})) \end{cases} \tag{3}$$

For SSIM($\boldsymbol{I_{ar}}, \boldsymbol{I_{input}}$), the higher values mean that I_{ar} has the similar values with the original image I_{input} in the corresponding positions. Videlicet, the lower values represent the pixels with severe distortions, *i.e.* coarse textural text layer $\overline{TTL}$. Therefore, the $\overline{TTL}$ can be obtained by equation 3, where $\boldsymbol{N}$ is a normalization function to make sure that the SSIM($\boldsymbol{I_{ar}}, \boldsymbol{I_{input}}$) is from 0 to 1. Analogously, we obtained the $\overline{TPL}$. It's worth noting that $\overline{TTL}$ also contains some sharp pictorial textural details which are similar to steep edges of text region, and vise versa. For refining coarse $\overline{TTL}$, we emphasized $\overline{TTL}$, while suppressed $\overline{TPL}$ by equation 4, and the same applies to $\overline{TPL}$.

$$\begin{cases} \boldsymbol{TTL} = \mathrm{binary}(\max(\overline{\boldsymbol{TTL}} - \boldsymbol{w} \times \overline{\boldsymbol{TPL}}, \boldsymbol{0})) \\ \boldsymbol{TPL} = \mathrm{binary}(\max(\overline{\boldsymbol{TPL}} - \boldsymbol{w} \times \overline{\boldsymbol{TTL}}, \boldsymbol{0})) \end{cases} \tag{4}$$

where weighting coefficient $\boldsymbol{w}$ is set equal to 2 based on a lot of experimental data. In addition, experimental data shows that most common SCIs, such as webpages and slides, have smooth backgrounds in a few base colors. Therefore, we found out the most frequent base colors accounting for at least 20% of all pixels, so that we could extract SBL in base colors. Heretofore, we obtained TTL, TPL, and SBL. However, remaining SPL and STL are difficult to differentiate since they have similar small variances. Consequently, we extracted a binary index map made up of SPL, TPL, and STL by $1 - SBL - TTL$ which is shown in Fig.2 (c).

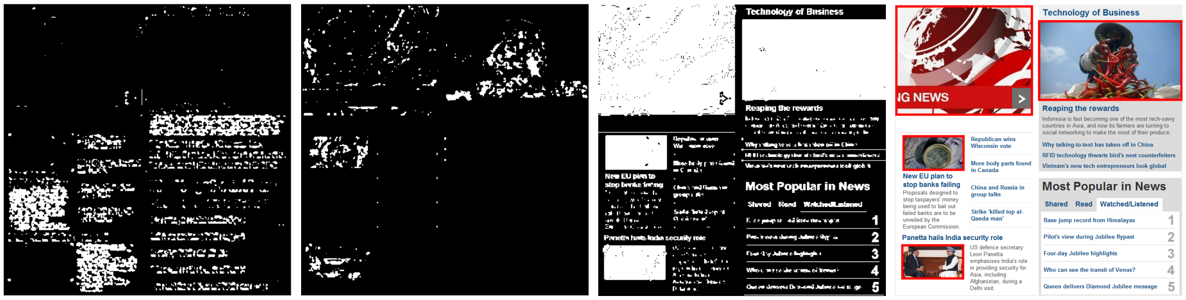
2.2. Quality Metric for Pictorial Layer

Driven by saving transmitting cost, we refined a feature vector $\boldsymbol{V_r}$ from reference SCI as the reduced-reference information, rather than utilize every pixel of the reference SCI like [3, 4, 5, 6]. Taking a departure from the binary index map as Fig.2 (c), we adopted Matlab function *bwareaopen.m* to eliminate tiny STL of the index map, then extracted information of remaining connected regions by function *bwconncomp.m*. Specifically, the information includes the number $\boldsymbol{N_r}$ of connected regions (i.e. the number of natural scene pictures located at the reference SCI), and corresponding location information. Furthermore, for each picture I_i (i=1,· · ·,$\boldsymbol{N_r}$), we extracted the coordinate values [$X_{i,u}$, $Y_{i,u}$], [$X_{i,b}$, $Y_{i,b}$],[$X_{i,l}$, $Y_{i,l}$] and [$X_{i,r}$, $Y_{i,r}$] representing upper, bottom, left, and right corners of I_i respectively. Consequently, we easily calculated width $\boldsymbol{W_i}$ and height $\boldsymbol{H_i}$ of I_i. Notably, regardless of picture content, we defined the inclination angle $\boldsymbol{A_i}$ of I_i as equation 5.

$$\boldsymbol{A_i} = \arctan\left(\frac{|\boldsymbol{Y_{i,u}} - \boldsymbol{Y_{i,l}}|}{|\boldsymbol{X_{i,u}} - \boldsymbol{X_{i,l}}|}\right) \tag{5}$$

Furthermore, for picture I_i, we adopted FEDM metric to calculate corresponding quality score $\boldsymbol{FE_{i,r}}$ based on free energy principle [14, 15]. Concretely speaking, suppose that the internal generative model $\boldsymbol{g}$ of human brain is parametric for visual perception, and the perceived scene can be explained by adjusting the parameter vector $\boldsymbol{\phi}$. Given the input visual signal $\boldsymbol{s}$, its surprise (measured by entropy) can be attained by integrating the joint distribution $\boldsymbol{p(s, \phi|g)}$ over the space of model parameter $\boldsymbol{\phi}$. The free energy is defined as equation 6.

$$\boldsymbol{f(\phi)} = -\int \boldsymbol{q(\phi|s)} \log \frac{\boldsymbol{p(s,\phi)}}{\boldsymbol{q(\phi|s)}} \mathrm{d}\boldsymbol{\phi} \tag{6}$$


Technology of Business
Reaping the rewards
Most Popular in News
Shared Read Watched/Listened
IG NEWS
New EU plan to stop banks failing
Panetta hails India security role

(a) (b) (c) (d)

Fig. 2. (a) Textural text layer (TTL); (b) Textural pictorial layer (TPL); (c) The index map including textural pictorial layer (TPL), smooth pictorial layer (SPL) and smooth text layer (STL); (d) Segmentation result. Notably, there are still some "impurities" in these coarse results.

Considering the computational and operational aspects of free energy, we adopted AR model to simulate human brain generative model $\boldsymbol{g}$, so that the quantitative measurement of FEDM is defined as entropy of error map $I_{i,\Delta}$ between input image I_i and its AR model filtering result $I_{i,ar}$ ($I_{i,\Delta} = I_i - I_{i,ar}$).

$$\boldsymbol{FE_{i,r}} = -\sum_{k} \boldsymbol{p_k(I_{i,\Delta})}\log \boldsymbol{p_k(I_{i,\Delta})} \tag{7}$$

Heretofore, we obtained the parameter vector $\boldsymbol{V_r}$ as

$$\begin{cases} \boldsymbol{V_r = \{V_{i,r} | i \in \{1,2,\cdots,N_r\}\}} \\ \boldsymbol{V_{i,r} = [X_{i,u}, Y_{i,u}, W_i, H_i, A_i, FE_{i,r}]} \end{cases} \tag{8}$$

As suggested by research about webpage saliency [9], i.e. human visual fixations usually fall in the top-left region when viewing the SCIs, we proposed the top-left bias pooling strategy to emphasis the impact of pictures' locations on ultimate quality score $\boldsymbol{score_p}$ of pictorial layer.

$$\begin{cases} \boldsymbol{score_p} = \sum_{i=1}^{N_r} \boldsymbol{\mu_i |FE_{i,r} - FE_{i,d}|} \\ \boldsymbol{\mu_i} = \frac{\boldsymbol{D([X_{i,c},Y_{i,c}],[1,1])^{-1}}}{\sum_{j=1}^{N_r} \boldsymbol{D([X_{j,c},Y_{j,c}],[1,1])^{-1}}} \end{cases} \tag{9}$$

Where $\boldsymbol{FE_{i,d}}$ is the free energy quality index of the $\boldsymbol{i}$th picture $I_{i,d}$ located at the distorted SCI (we can easily find out the location of $I_{i,d}$ using $\boldsymbol{V_r}$). Specifically, the pooling coefficient $\boldsymbol{\mu_i}$ is in inverse proportion to Euclidean distance (represented by $\boldsymbol{D}$) between centroid point $[\boldsymbol{X_{i,c}, Y_{i,c}}]$ of picture $I_{i,d}$ and upper-left corner $[\boldsymbol{1,1}]$ of the distorted SCI. The centroid point $[\boldsymbol{X_{i,c}, Y_{i,c}}]$ can be calculated as follows.

$$\begin{cases} \boldsymbol{X_{i,c} = X_{i,u}} + \sqrt{(\frac{\boldsymbol{W_i}}{\boldsymbol{2}})^2 + (\frac{\boldsymbol{H_i}}{\boldsymbol{2}})^2}\sin(\boldsymbol{A_i} + \arctan(\frac{\boldsymbol{H_i}}{\boldsymbol{W_i}})) \\ \boldsymbol{Y_{i,c} = Y_{i,u}} + \sqrt{(\frac{\boldsymbol{W_i}}{\boldsymbol{2}})^2 + (\frac{\boldsymbol{H_i}}{\boldsymbol{2}})^2}\cos(\boldsymbol{A_i} + \arctan(\frac{\boldsymbol{H_i}}{\boldsymbol{W_i}})) \end{cases} \tag{10}$$

2.3. Quality Metric for Text Layer

Experimental analysis about subjective scores of SCI database [3] pointed out that blur and contrast change were dominative distortions for text layer, because these distortion types impacted human perception severely when reading text. Therefore, in this section, we proposed two novel quality features for measuring contrast change and blurness of text layer.

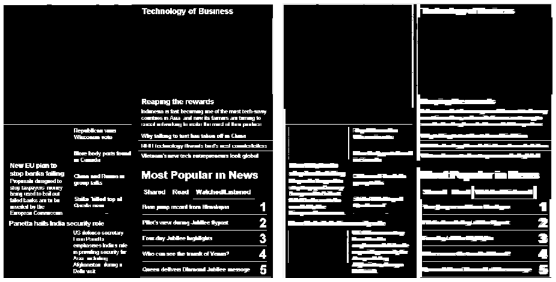
Based on the parameters of $\boldsymbol{V_r}$ obtained in section 2.2, for each distorted SCI, we can easily extract an index map $\boldsymbol{M_t}$ made up of SBL, STL, and TTL. Subsequently, for $\boldsymbol{M_t}$, we extracted the gray values of background and text by counting the frequence of each gray value. In other word, most text layer usually contains pure background and uniform text color, so that background and text gray values have the two largest frequences. Therefore, we obtained four parameters $\boldsymbol{B_r}$, $\boldsymbol{T_r}$, $\boldsymbol{B_d}$ and $\boldsymbol{T_d}$ by counting the top-two frequences of the reference and distorted SCIs. And $\boldsymbol{B_r}$, $\boldsymbol{T_r}$, $\boldsymbol{B_d}$, $\boldsymbol{T_d}$ represent reference SCI's background, text, distorted SCI's background, text separately. Above all, the first feature is :

$$\boldsymbol{f_1} = \frac{1}{255}\frac{|\boldsymbol{B_r - B_d}|}{|\boldsymbol{T_d - B_d}| + C1} \tag{11}$$

where C_1 is a positive constant (set as 1) used to avoid instability when denominator is close to zero. The weighting coefficient $\frac{1}{255}$ guarantees that the $\boldsymbol{f_1}$ is from 0 to 1. Obviously, higher $|\boldsymbol{B_r - B_d}|$ means severe contrast change distortion, while lower $|\boldsymbol{T_d - B_d}|$ means that the text and background of distorted SCI is in low contrast. The higher $\boldsymbol{f_1}$ means that it's difficult for human eyes to distinguish between text and background, i.e. the lower quality score. The second feature $\boldsymbol{f_2}$ is designed for measuring blurness of text layer. Notably, we obtained the text layer $\boldsymbol{TL}$ ($\boldsymbol{TL}$=binary($TTL + STL$)) shown in Fig.3 by eliminating background pixels whose gray values are $\boldsymbol{B_d}$ from $\boldsymbol{M_t}$. We firstly adopted Matlab function $bwareaopen.m$ to eliminate the tiny connected regions (noise) from $\boldsymbol{TL}$, then we utilized $bwconncomp.m$ to find out the number $\boldsymbol{N_{t,d}}$ of remaining connected regions of $\boldsymbol{TL}$. The $\boldsymbol{f_2}$ is calculated as

$$\boldsymbol{f_2} = \frac{|\boldsymbol{N_{t,r} - N_{t,d}}|}{\boldsymbol{N_{t,r}}} \tag{12}$$

where $\boldsymbol{N_{t,r}}$ is the number of connected regions of the reference SCI's text layer. Notably, we refined a brief vector



(a) (b)

Fig. 3. The text layer of (a) reference SCI ($\boldsymbol{N_{t,r}}$=211), (b) distorted SCI corrupted by motion blur ($\boldsymbol{N_{t,d}}$=34).

$\boldsymbol{V_{r,t}}$=$[\boldsymbol{B_r}, \boldsymbol{N_{t,r}}]$ containing only two parameters to represent text layer of the reference SCI. Heretofore, we obtained the quality score of text layer as $\boldsymbol{score_t} = \frac{1}{2}\boldsymbol{f_1} + \frac{1}{2}\boldsymbol{f_2}$. Eventually, the final quality score is defined as

$$\boldsymbol{score} = \theta \boldsymbol{score_p} + (1-\theta)\boldsymbol{score_t} \tag{13}$$

where the weighting coefficient θ is the area ration between pictorial layer and the whole SCI.

3. EXPERIMENTAL RESULT

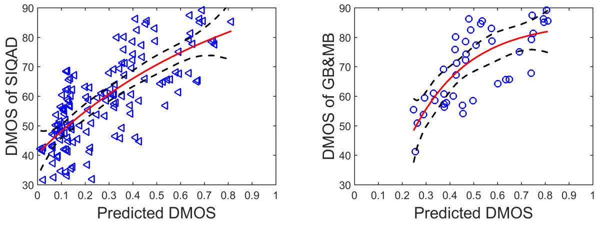
We adopted SIQAD [3] as test database to validate the performance. SIQAD is a large-scale SCI quality assessment database consisting of 20 source and 980 distorted SCIs. The distortion types of SIQAD include Gaussian Noise (GN), Gaussian Blur (GB), Motion Blur (MB), Contrast Change (CC), JPEG, JPEG2000, and Layer Segmentation Based Coding (LSC). We compared the proposed method with state-of-the-art full-reference and reduced-reference quality metrics. Suggested by video quality experts group (VQEG) [17], we first used a logistic regression function $q(\boldsymbol{score}) = \beta_1(\frac{1}{2} - \frac{1}{1+exp(\beta_2(\boldsymbol{score}-\beta_3))}) + \beta_4\boldsymbol{score} + \beta_5$ to generate the mapped score, where $\beta_j(j = 1,2,3,4,5)$ are free parameters to be determined during the curve fitting process. Then we calculated the frequently used performance evaluations $PLCC$, $SROCC$ and $RMSE$ to validate the prediction accuracy.

Table 1. Performance over all distortion types

IQA Metrics	PLCC	SROCC	RMSE
SSIM [4]	0.7445	0.7433	9.4713
PSNR	0.5788	0.5539	11.5691
VIF [18]	0.8026	0.7857	8.4642
ADD-GSIM [19]	0.6844	0.6842	10.436
PSIM [20]	0.7144	0.7056	10.016
Q_s [5]	0.8573	0.8456	7.3030
SQMS [6]	**0.8872**	**0.8803**	**6.6039**
SPQA [3]	0.8631	0.8579	7.2297
Proposed	**0.8126**	**0.7962**	**8.2633**

Table 2. Performance over Gaussian Blur and Motion Blur

IQA Metrics	PLCC	SROCC	RMSE
SSIM [4]	0.8537	0.8481	7.1334
Q_s[5]	**0.8972**	**0.8856**	**6.7335**
SIQM [6]	0.8785	0.8750	6.9241
SPQA [3]	0.8687	0.8636	6.8262
Proposed	**0.8907**	**0.8846**	**6.7638**



(a) (b)

Fig. 4. Scatter plots of the proposed metric over (a) all distortion types of SIQAD, (b) GB and MB distortions of SIQAD.

The performances in SIQAD of competitive metrics have been reported in Table 1. Experimental results imply that the proposed method outperforms most general quality metrics, i.e. SSIM, VIF, VSI, PSNR, FSIM. Moreover, as a reduced-reference metric, the proposed method is highly competitive compared with full-reference metrics SPQA, SIQM, and Q_s dedicated for SCI. Notably, the performances shown in Table 2 indicate that we bold the top metric with Q_s in Gaussian Blur and Motion Blur distortions of SIQAD since the feature $\boldsymbol{f_2}$ improves the performance for blur significantly.

The scatter plots of DMOS versus the proposed metric on all distortion types and GB&MB distortions of SIQAD are presented in Fig.4, where the red lines are curves fitted with the logistic regression function [17] and the blue dash lines are the 95% confidence intervals of the fitting.

4. CONCLUSION

We firstly designed a novel layer-based segmentation method to divide SCI into text and pictorial layers. Subsequently, we proposed a reduced-reference SCI quality metric considering the perceptual characters of different regions. Validation experiments show encouraging performances, especially for blur distortions. The development of this metric for border application scenarios such as evaluating SCIs corrupted by realistic distortions is worth addressing in the future.

5. ACKNOWLEDGEMENT

This work was supported by the National Science Foundation of China (61422112, 61371146, 61521062, 61527804), National High-tech R&D Program of China (2015AA015905), and Science and Technology Commission of Shanghai Municipality (15DZ0500200).

6. REFERENCES

[1] H. Shen, Y. Lu, F. Wu, and S. Li, “A high-performanance remote computing platform,” *in ICPCC*, pp. 1-6, Mar. 2009.

[2] C.-Y. Huang, C.-H. Hsu, Y.-C. Chang, and K.-T. Chen, “GamingAnywhere: An open cloud gaming system,” *in ACM MSC*, pp. 36-47, 2013.

[3] Yang H, Fang Y, Lin W. “Perceptual quality assessment of screen content images.” *IEEE Transactions on Image Processing*, 2015, 24(11): 4408-4421.

[4] Wang Z, Bovik A C, Sheikh H R, et al. “Image quality assessment: from error visibility to structural similarity.” *Image Processing, IEEE Transactions on,* 2004, 13(4): 600-612.

[5] S. Wang, K. Gu, K. Zeng, Z. Wang, and W. Lin, “Objective quality assessment and perceptual compression of screen content images,” *IEEE Computer Graphics and Applications*, 2017, to appear.

[6] K. Gu, S. Wang, H. Yang, W. Lin, G. Zhai, X. Yang, and W. Zhang, “Saliency-guided quality assessment of screen content images,” *IEEE Trans. Multimedia*, vol. 18, no. 6, pp. 1-13, Jun. 2016.

[7] K. Gu, D. Tao, J.-F. Qiao, and W. Lin, “Learning a no-reference quality assessment model of enhanced images with big data,” *IEEE Trans. Neural Netw. Learning Syst.*, 2017.

[8] K. Gu, J. Zhou, J.-F. Qiao, G. Zhai, W. Lin, and A. C. Bovik, “No-reference quality assessment of screen content pictures,” *IEEE Trans. Image Process.*, 2017.

[9] Shen C, Zhao Q. “Webpage saliency”. *European Conference on Computer Vision, Springer International Publishing*, 2014: 33-46.

[10] Lin T, Hao P. “Compound image compression for real-time computer screen image transmission.” *IEEE transactions on Image Processing*, 2005, 14(8): 993-1005.

[11] Pan Z, Shen H, Lu Y, et al. “A low-complexity screen compression scheme for interactive screen sharing.” *Circuits and Systems for Video Technology, IEEE Transactions on*, 2013, 23(6): 949-960.

[12] Minaee S, Wang Y. “Screen content image segmentation using least absolute deviation fitting.” *Image processing (ICIP)*, 2015 IEEE international conference on. IEEE, 2015: 3295-3299.

[13] Minaee S, Abdolrashidi A, Wang Y. “Screen content image segmentation using sparse-smooth decomposition.” *2015 49th asilomar conference on signals, systems and computers*. IEEE, 2015: 1202-1206.

[14] Zhai G, Wu X, Yang X, et al. “A psychovisual quality metric in free-energy principle.” *Image Processing, IEEE Transactions on,* 2012, 21(1): 41-52.

[15] K. Gu, G. Zhai, X. Yang, and W. Zhang, “Using free energy principle for blind image quality assessment,” *IEEE Trans. Multimedia*, vol. 17, no. 1, pp. 50-63, Jan. 2015.

[16] He K, Sun J, Tang X, “Guided image filtering.” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 2013, 35(6): 1397-1409.

[17] Video Quality Experts Group et al., “Final report from the video quality experts group on the validation of objective models of video quality assessment,” VQEG, Mar, 2000.

[18] Hamid R Sheikh and Alan C Bovik, “Image information and visual quality,” *Image Processing, IEEE Transactions on*, vol. 15, no. 2, pp. 430C444, 2006.

[19] K. Gu, S. Wang, G. Zhai, W. Lin, X. Yang, and W. Zhang, “Analysis of distortion distribution for pooling in image quality prediction,” *IEEE Trans. Broadcasting*, vol. 62, no. 2, pp. 446-456, Jun. 2016.

[20] K. Gu, L. Li, H. Lu, X. Min, and W. Lin, “A fast reliable image quality predictor by fusing micro- and macro-structures,” *IEEE Trans. Ind. Electron.*, vol. 64, no. 5, pp. 3903-3912, May 2017.

[21] Wang S, Zhang X, Liu X, et al. “Utility-driven adaptive preprocessing for screen content video compression.” *IEEE Transactions on Multimedia*, 2017, 19(3): 660-667.

[22] Shao F, Gao Y, Li F, et al. “Toward a Blind Quality Predictor for Screen Content Images.” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2017.