Adaptive Screen Content Image Enhancement Strategy using Layer-based Segmentation

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

[†]*Insti. of Image Commu. and Infor. Proce., Shanghai Jiao Tong University, China*
[‡]*School of Computer Science and Engineering, Nanyang Technological University, Singapore*
[§]*Luman Université, Université de Nantes, IRCCyN UMR CNRS 6597, Polytech Nantes, France*
[*]*Harbin Institute of Technology, China*

***Abstract*—The ubiquitous screen content images (SCIs) play a significant role in various scenarios currently. However, most SCIs captured by consumer devices are frequently corrupted with distortions, especially contrast distortion. Unlike the natural images, SCIs are composed of text, graphics and natural scene pictures so that traditional image enhancement methods are not suitable for these compound images. Therefore, we innovatively proposed an adaptive strategy for enhancing SCIs in this paper. Firstly, we devised a segmentation method to divide SCI into text and pictorial regions. Next, the famous guided image filter (GIF) with big and small kernel sizes served as unsharpness masking for processing different regions adaptively. For verifying performance, the proposed method was tested on recently prevalent SCI datasets including SIQAD, and Webpage Dataset. Experimental results indicate that the proposed approach outperforms state-of-the-art methods in most SCIs with flat background.**

***Index Terms*—Screen content image, image segmentation, guided filter, unsharpness masking, autoregressive model**

I. Introduction

Screen content image (SCI) is a crucial medium for various applications since it's informative and convenient. Unfortunately, most SCIs are captured by normal devices like cellphone and tablet which corrupt the SCIs with distortions. To the best of our knowledge, image enhancement strategy dedicated for SCI has not been systematically investigated before. Besides, traditional methods [1-4] are not appropriate for SCIs since SCIs always contain steep edges in text region, chaotic textural detail in pictorial region, and significant boundaries between pictorial region and background. Therefore, we attempted to design a robust strategy under the motivation of improving diverse SCIs' perceptual quality [17, 18].

The accurate and robust SCI segmentation method is necessary for SCI related researches, because any tiny segmentation fault will produce severe artifacts so as to destroy integral quality. In general, mainstream SCI segmentation methods can be classified as two categories, i.e. block-based methods [5,6] and layer-based methods[7,8]. However, these methods paid attention to extracting texture details from the SCI. And for SCI enhancement task, these methods will excessively enhance the image texture and boundaries so as to produce halo artifacts. Hence, firstly, we came up with an accurate and robust segmentation method suitable for various SCIs. Then we proposed an adaptive strategy based on GIF [4] to process different regions respectively. The proposed flowchart is shown in Fig.1 for viewers' conveniences.

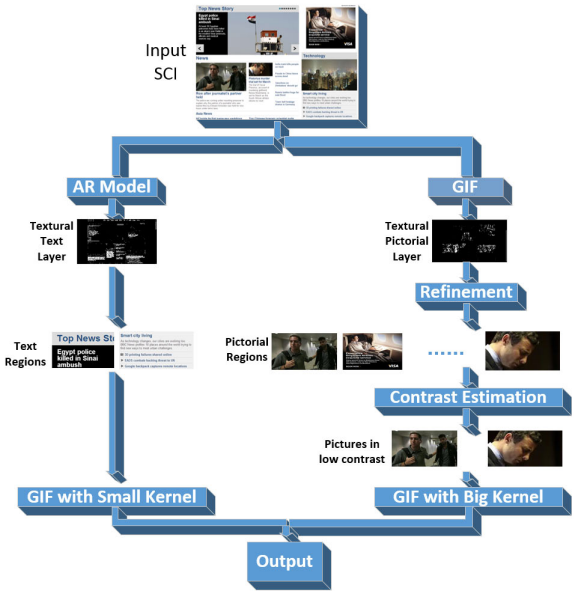


Fig. 1. Flowchart of the proposed method.

This paper is organized as follows. We first introduce the proposed layer-based segmentation strategy in section 2. Then we further introduce the enhancement method. In section 3, we compare the proposed method with other methods in terms of classification accuracy and SSIM.

II. Proposed Enhancement Algorithm

A. Layer-based Segmentation Strategy

We innovatively proposed a robust layer-based SCI segmentation method suitable for diverse SCI enhancement tasks in this section. The proposed method consists of two steps, coarse segmentation step and refinement step.

Considering that the block-based methods [5,6] introduce serious blockness distortions in SCI enhancement tasks, we designed a layer-based method to divide SCI into five layers. Specifically, observation shows that most SCIs consist

978-1-5386-4881-0/18/$31.00 ©2018 IEEE

of pictorial layer, text layer, and smooth background layer. Furthermore, pictorial layer consists of smooth pictorial layer (SPL) and textural pictorial layer (TPL). Similarly, text layer also contains smooth text layer (STL) and textural text layer (TTL). Consequently, SPL, TPL, STL, TTL and smooth background layer (SBL) make up of a complete SCI without overlap. In coarse segmentation, we extracted TPL, TTL, and SBL by spatial filtering technique as follows.

A consensus has been reached that text layer contains a lot of short steep edges, while pictorial layer contains many chaotic texture details and long boundaries. Considering this heuristic information, we adopted the autoregressive (AR) model [9] and GIF 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

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

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 $\boldsymbol{Y}(\boldsymbol{i},:) = \boldsymbol{\gamma}^k(y_i)$ 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 $\hat{\boldsymbol{\alpha}} = (\boldsymbol{Y^T Y})^{-1}\boldsymbol{Y^T y}$. The AR model can protect pictorial details well, but it performs poorly on steep edges of text. This property is elaborated in zoom-in trapped by red rectangle in Fig.2 (c). The GIF can generate output according to the guide image. As Fig.2(d), the GIF behaves as an efficient edge-preserving smoothing operator when the guide image is identical to the original image.

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

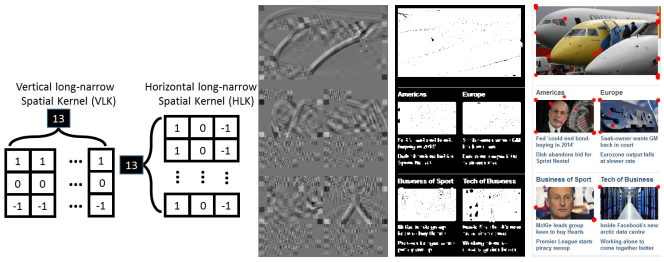
Fig. 2. (a)original SCI: red rectangle is detail of pictorial region, and green rectangle is detail of text region.(b)zoom-in of regions trapped by red and green rectangles from (a).(c)zoom-in of image processed by AR model.(d)zoom-in of image processed by GIF.

We calculated the similarity map S_1 between the AR model filtering result and the original image by famous SSIM metric[12]. In similarity map S_1, higher values mean that the AR model filtering result has the similar values with original image in these positions. Videlicet, the lower values represent regions with severe distortions, $i.e.$ coarse textural text layer ($\overline{TTL}$). Therefore, the $\overline{TTL}$ can be obtained by $1 - N(S_1)$, where N is a normalization function to make sure the S_1 is from 0 to 1. Analogously, the $\overline{TPL}$ can be generated by $1 - N(S_2)$, where S_2 is the similarity map between the GIF filtering result and the original image. Notably, $\overline{TTL}$ also contains some sharp pictorial textural details which are similar to steep edges of text region, and vise versa. For obtaining pure TTL, we emphasized $\overline{TTL}$, while suppressed $\overline{TPL}$ by equation 3. And the same applies to TPL.

$$\begin{cases} TTL = Binary(Max(\overline{TTL} - w \times \overline{TPL}, 0)), \\ TPL = Binary(Max(\overline{TPL} - w \times \overline{TTL}, 0)), \end{cases} \tag{3}$$

where weighting coefficient w is set as 2 based on experimental data. Moreover, experimental data shows that most SCIs, such as webpages and slides, have smooth backgrounds in a few base colors. Hence, we found out the most frequent base colors accounting for at least 20% of all pixels, so that we could extract SBL whose pixels are in base colors. Heretofore, we obtained TTL, TPL, and SBL coarsely.

For avoiding severe artifacts in SCI enhancement tasks, we were supposed to find out the explicit boundaries between pictures and background, especially for SCIs containing multiple pictures. However, coarse segmentation only provides textural details of different regions, which are insufficient to determine the exact locations of each pictures. Besides, remaining STL and SPL are difficult to differentiate because they have similar color and variation information. Consequently, we proposed a refinement strategy to accurately determine every pictures' positions one by one.



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

Fig. 3. (a) Vertical and horizontal long-narrow spatial kernels. (b) Right angle corner map generated by (a). (c) Binary map including STL, SPL and TPL of the original SCI. (d) Red points represent right angle corners extracted by the proposed method.

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

Fig. 4. (a) A picture patch of Fig.3 (d). (b) Zoom-in of area trapped by red rectangle in (a). (c) Filtering result of (b) by VLK. (d) Filtering result of (c) by HLK.

We devised a right angle corner detection method using two long-narrow spatial kernels shown in Fig.3(a). Firstly, we adopted 3×13 VLK as gradient filter to process original SCI. Next, the filtering result of VLK was processed by 13×3 HLK so as to generate Fig.3(b). Obviously, the filtering operation by VLK and HLK exactly produced four smooth square blocks in the four vertices of each picture. We analyzed the reason for this situation as follows. For most SCIs, there are sharp boundaries between pictures and smooth background, especially for pictures having strong contrast with

978-1-5386-4881-0/18/$31.00 ©2018 IEEE

background, such as dark picture and white background shown as Fig.4(a). Therefore, VLK produced high gradient values in horizontal boundaries of each picture, as shown in Fig.4(c). Then we adopted HLK to process Fig.4(c) and obtained a smooth square block as Fig.4(d). Furthermore, based on abundant experimental data, we empirically drew three important properties from these square blocks as equations 4,

$$\begin{cases} |\boldsymbol{Mean(block)}| > 2 \times \boldsymbol{Mean(I)}, \\ \boldsymbol{Std(block)} < \frac{1}{4} \times \boldsymbol{Std(I)}, \\ \boldsymbol{Symbol(block)} = 1, \end{cases} \tag{4}$$

where $\boldsymbol{block}$ represents the square block, $\boldsymbol{I}$ is the original SCI, $\boldsymbol{Mean}$ represents mean value, $\boldsymbol{Std}$ is standard deviation, and $\boldsymbol{Symbol} = 1$ means that all pixels of the block have the same signs, i.e. homogeneity. And the block size is 13×13. We extracted right angle corners as Fig.3(d). Specifically, Fig.3(c) is the binary map containing STL, SPL and TPL of Fig.3(d). After obtaining TPL, TTL, and SBL in coarse segmentation step, we can get Fig.3(c) easily by $1-TTL-SBL$. We found that each picture area of Fig.3(c) approximates to an inerratic connected region. Heretofore, we designed a region growing method with right angle corner points serving as growing seeds. Concretely speaking, we adopted a 20×20 window, and slid this window starting from growing seed on the Fig.3(c). The break condition is $P(window) < 95\%$, where $P(window)$ represents the percentage of positive pixels of the window. Eventually, we found out the coordinate values of upper-left and bottom-right corners of every picture precisely, as shown in Fig.5.

B. Adaptive Enhancement Strategy

We devised an adaptive enhancement strategy dedicated for various SCIs. Notably, most SCIs contain either single picture or multi-pictures like Fig.5. Moreover, the multi-pictures always contain different levels of contrast distortions. Taking Fig.6 (a) for instance, the bottom-right picture is in fine contrast, while the others are in low contrast. Firstly, we simply estimated the histogram of each picture using [2], and found out the severely distorted images in low contrast, so as to avoid over-enhancing pictures in fine contrast.

Secondly, we adopted the GIF serving as unsharpness masking for improving different regions. Concretely speaking, the original image is processed by spatial low-pass filter to generate the smoothed one, then the detail version is obtained by point-by-point subtraction of original image and its associated smoothed version. Furthermore, the detail image is amplified and added back to the original image to generate the final enhanced image. Specifically, GIF behaves as a quick edge-preserving smoothing operator when the guide image is identical to the original image. Hence, the GIF has the edge-preserving ability compared to the Gaussian filter. Moreover, in comparison to the bilateral filter, the GIF performs better near edges. The GIF also is of low computation complexity. Therefore, we improved the GIF to be an adaptive method for satisfying real-time SCI related enhancement tasks.

Although the unsharpness masking is a powerful tool for image enhancement, it produces some perceptually harmful artifacts such as halo effect around text and boundaries like Fig.6(d). The halo effect always happens in steep edges where light changes to dark or vise versa. Especially for small-scale edge details, when we use the big low-pass filter kernel (radius r), the surrounding pixels of tiny edge details will be mistakenly extracted as high frequency component, and be added back to the original image, so that the surrounding region around tiny edge details will be over enhanced. Besides, for SCIs containing coloured background, the halo effect looks more obvious. The same distortion appears on the boundaries between pictorial region and background. Conversely, small kernel will amplify noise of pictorial regions as Fig.6 (e). Hence, we improved the GIF as follows :

$$a_k = \frac{\frac{1}{|w|}\sum_{i \in w_k} I_i p_i - \mu_k \overline{p}_k}{\sigma_k^2 + \epsilon}, b_k = \overline{p}_k - a_k \mu_k, \tag{5}$$

$$\begin{cases} w = 16 \times 16, \epsilon = 0.1^2 & I_i, p_i \in picture, \\ w = 2 \times 2, \epsilon = 0.05^2 & I_i, p_i \in text, \end{cases} \tag{6}$$

$$q_i = \overline{a}_i I_i + \overline{b}_i, o = (I - q) \times \alpha + q, \tag{7}$$

where p is the original SCI, $I = p$ is the guide image, $|w|$ is the number of pixels in window w_k, and $r = |w|^{\frac{1}{2}}$ is the kernel radius. μ_k and σ_k^2 are mean and variance of I in window $|w_k|$. ϵ is a regularization parameter. q is the filtering result of the GIF. α is the amplification coefficient for enhancing high frequency component. o is the final enhanced result. The proposed adaptive strategy guarantees the satisfying enhancement performance under the premise of no artifacts.

III. Experimental Results

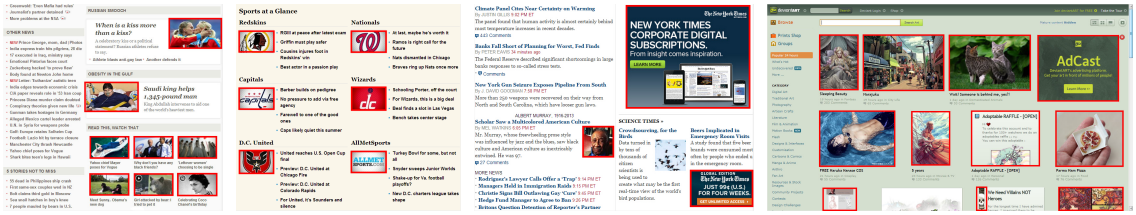
We adopted the prevalent SIQAD [10], and Webpage Dataset [11] as test datasets. SIQAD consists of 20 source SCIs and 140 distorted images with contrast change, and Webpage Dataset includes 149 SCIs. The segmentation results are shown in Fig.6. Most segmentation results are accurate, and we obtained the specific location of each picture from the same SCI. But one picture of Fig.5(c) has not been extracted, because there is little textual detail of this picture, and its base color is similar to the background. We extracted the pictorial regions and text regions of test datasets manually as ground truth, and calculated the average precision and recall of SPEC [5] and the proposed segmentation method as Table 1.

TABLE I
THE ACCURACY AND RECALL OF SEGMENTATION METHODS SPEC (BLOCK-BASED) *vs* THE PROPOSED METHOD (LAYER-BASED)

Method	Test Dataset	Precision	Recall
SPEC [5]	SIQAD	0.64	0.7962
Proposed method	**SIQAD**	**0.81**	**0.8938**
SPEC [5]	WebpageDataset	0.58	0.6645
Proposed method	**WebpageDataset**	**0.69**	**0.7245**

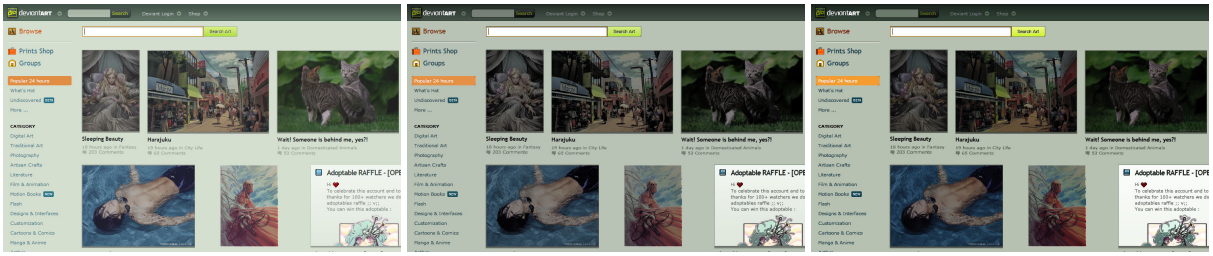
Subjective performances of different methods are shown in Fig.6. Notably, Fig.6(a) contains six pictures, and the bottom-right one is in fine contrast, while the others are underexposed. Obviously, the HMF [15] and GHMF [16] methods destroy the integral quality severely, because they process the image from a global perspective, so that Fig.6(b-c) are polarized. In Fig.6(d), there are severe halo artifacts around text, and

978-1-5386-4881-0/18/$31.00 ©2018 IEEE



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

Fig. 5. Segmentation results for various SCIs.



(a) SCI with contrast distortions (b) HMF (c) GHMF

(d) GIF using big kernel (e) GIF using small kernel (f) The proposed method

Fig. 6. Enhancement results of different methods for SCI in Webpage Dataset.

TABLE II
THE QUALITY SCORES OF ENHANCEMENT METHODS

Method	Test Dataset	SSIM	RWQMS	QMC
HMF	SIQAD	0.7337	0.8831	0.1845
GHMF	SIQAD	0.6973	0.8825	0.2272
GIF(big r)	SIQAD	0.8836	0.8675	0.0222
GIF(small r)	SIQAD	0.8189	0.9037	0.2835
Proposed	**SIQAD**	**0.9363**	**0.9278**	**0.0009**
HMF	WebpageDataset	0.8189	0.8722	0.2064
GHMF	WebpageDataset	0.7919	0.8548	0.2672
GIF(big r)	WebpageDataset	0.7702	0.7364	0.0237
GIF(small r)	WebpageDataset	0.7483	0.9016	0.0515
Proposed	**WebpageDataset**	**0.8959**	**0.9370**	**0.0004**

boundaries between pictures and background (refer to the long orange block and the boundaries of pictures in Fig.6(d)). Moreover, the text regions are corrupted with colour distortions, and the bottom-right picture is over-enhanced. Besides, GIF with small kernel amplifies noise in pictorial regions as Fig.6(e). The proposed method improves the contrast of underexposed pictures, and introduces no artifacts. In addition, the text regions are protected well. We also used objective full-reference image quality metric SSIM [12], the reduced-reference SCI quality metric RWQMS [13], and quality metric of contrast QMC [14] to evaluate algorithm performance quantitatively. Notably, both SIQAD and Webpage Dataset provide reference images without distortion, so that we can calculate the SSIM easily. RWQMS is an ultramodern specifically designed quality metric for SCIs, and is qualified to estimate SCI perceptual quality credibly. Considering that the contrast change is the most obvious distortion here, we also calculated the QMC dedicated for contrast distortion. Specifically, big SSIM and RWQMS scores mean good quality, while small QMC score means good quality. The quality scores are listed in Table 2.

IV. CONCLUSION

We innovatively proposed an adaptive enhancement strategy for various frequently-used SCIs. Firstly, we designed a robust layer-based segmentation to divide SCIs into pictorial regions and text regions accurately. Next, the GIF with different kernel sizes serving as unsharpness masking were used to process different regions respectively. Experimental results show that the proposed method is qualified to process diverse SCIs, and is the best choice for SCI enhancement tasks.

V. ACKNOWLEDGEMENT

This work is supported by the Major State Basic Research Development Program of China (973 Program 2015CB351804), the National Science Foundation of China under Grants 61672193.

978-1-5386-4881-0/18/$31.00 ©2018 IEEE

REFERENCES

[1] T. Arici, S. Dikbas, Y. A histogram, "A histogram modification framework and its application for image contrast enhancement," *IEEE Transcation on Image Process.*, vol. 18, no. 9, pp. 1921-1935, Sept. 2009.

[2] K Gu, W Lin, G Zhai, X Yang, W Zhang and C.W. Chen, "No-reference quality metric of contrast-distorted images based on information maximization," *IEEE Transactions on Cybernetics, 2017, 47(12): 4559-4565.*.

[3] K Gu, G Zhai, W Lin, and M Liu, "The analysis of image contrast: From quality assessment to automatic enhancement". *Cybernetics, IEEE Transactions on*, 2016, 46(1): 284-297.

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

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

[6] Z Pan, H Shen, Y Lu, S Li, and N Yu, "A low-complexity screen compression scheme for interactive screen sharing." *Circuits and Systems for Video Technology, IEEE Transactions on*, 2013, 23(6): 949-960.

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

[8] S. Minaee, A. Abdolrashidi, and Y. Wang, "Screen content image segmentation using sparse-smooth decomposition." *2015 49th asilomar conference. IEEE* , 2015: 1202-1206.

[9] Zhai G, Wu X, Yang X, and W Zhang, "A psychovisual quality metric in free-energy principle." *IEEE Transactions on Image Processing,,* 2012, 21(1): 41-52.

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

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

[12] Wang Z, Bovik A C, H.R. Sheikh, and E.P. Simoncelli, "Image quality assessment: from error visibility to structural similarity." *Image Processing, IEEE Transactions on,* 2004, 13(4): 600-612.

[13] Wang S, Gu K, 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.*2016, 532 - 543

[14] K Gu, G Zhai, W Zhang and C.W. Chen, "Automatic contrast enhancement technology with saliency preservation." *IEEE Transactions on Circuits and Systems for Video Technology*, 2015, 25(9): 1480-1494.

[15] T. Arici, S. Dikbas, and Y. Altunbasak, "A histogram modification framework and its application for image contrast enhancement." *IEEE Transcation on Image Process.*,vol. 18, no. 9, pp. 1921-1935, Sept. 2009.

[16] K Gu, G Zhai, S Wang, M Liu, J Zhou, and W Lin, "A general histogram modification framework for efficient contrast enhancement." *Circuits and Systems (ISCAS), 2015 IEEE International Symposium on. IEEE,*,2015: 2816-2819.

[17] X Min, K Ma, K Gu, G Zhai, Z Wang, W Lin, "Unified Blind Quality Assessment of Compressed Natural, Graphic, and Screen Content Images." *IEEE Transactions on Image Processing*,2017, 26(11), 5462-5474.

[18] X Min, K Gu, G Zhai, M Hu, X Yang, "Saliency-Induced Reduced-Reference Quality Index for Natural Scene and Screen Content Images." *Signal Processing*,2018, vol. 145, 127-136.

978-1-5386-4881-0/18/$31.00 ©2018 IEEE