Saliency Change Based Reduced Reference Image Quality Assessment

Jinjian Wu[1], Yongxu Liu[1], Guangming Shi[1], IEEE Senior Member, and Weisi Lin[2], Fellow, IEEE

***Abstract*—The image quality assessment (IQA) technique, which aims to perform coherently with subjective perception, is useful in quality-orientated image processing systems. In this paper, we suggest to take the saliency change into account for reduced reference (RR) IQA model. Generally, a saliency region will attract more attention, and our human vision is more sensitive to quality degradation on such region. Inspired by this, saliency values are firstly used to highlight these sensitive regions, and a local saliency weighted histogram (LSWH) based on visual orientation pattern is generated for visual feature extraction. Next, strong distortion may change the saliency from the reference to the distorted images. Thus, the saliency of each visual orientation pattern is measured, and a global saliency based histogram (GSBH) is created. Finally, by combining the LSWH and GSBH, a novel IQA model for reduced reference is introduced. Experimental results on five publicly available databases demonstrate that the proposed model uses only several values (9 values) as reference information, and performs consistently with subjective perception.**

***Index Terms*—Reduced-Reference (RR), Image Quality Assessment (IQA), Visual Saliency, Visual Content Extraction**

I. Introduction

With a tremendous growth of digital image/video technology and services, a reliable objective image quality assessment (IQA) method is greatly demanded for many quality-oriented image acquisition/processing systems. During the past decades, a lot of IQA methods have been proposed [1–4]. And based on the available of the reference image, these IQA methods can be classified into three categories [5]: Full-Reference (FR), No-Reference (NR) and Reduced-Reference (RR). Though FR IQA models always perform in accordance with subjective perception, such method do not work when reference images are not available. Meanwhile, the NR IQA can work without any reference information, but it is extremely difficult to design a reliable and efficient NR IQA. As a compromise between the two categories, the RR IQA technique requires only a small amount of the reference information and achieves a high performance. In this work, we focus on designing a reliable RR IQA method.

A large number of RR IQA models have been introduced in the past decades. Since the visual contents of natural scenes follow a certain type of statistical distribution (e.g., generalized Gaussian distribution), and distortions will degrade such a distribution characteristic. According to this natural scene statistic (NSS) assumption, Wang et al. [6] measure the quality as the NSS changes in the wavelet domain, and proposed the Wavelet-domain Natural Image Statistic Metric (WNISM). Following this assumption, Li et al. [7] measure the quality degradation with NSS on divisive normalization based wavelet. Moreover, Ma et al. [8] proposed to measure the NSS based quality degradation within the DCT domain, and Gao et al. [3] within the curvelets and contourlets coefficient. Moreover, Wu et al. [9] proposed to represent the image contents with several visual patterns for RR IQA. Soundararajan et al. [10] measured the extracted visual information with the scaled entropy on wavelet coefficient for RR IQA. Though these existing methods have greatly improved the performance of the RR IQA, there is still a remarkable gap between the RR IQA model and subjective perception.

In order to design a better RR IQA which performs more consistently with subjective perception, we investigate into the visual processing mechanism of the human visual system (HVS). The HVS processes an attention mechanism for visual signal processing [11], with which the HVS focuses only on a part of regions for detailed perception. Therefore, key regions play more important roles than the other regions for visual perception, and the degradation on these key regions have larger influence for quality assessment.

Inspired by this, we introduce a saliency change based RR IQA model. Firstly, saliency values are adopted to highlight these attention regions , and a local saliency weighted histogram (LSWH) based on visual orientation pattern [9] is generated for image content extraction. Next, considering that strong distortion may change the saliency regions between the reference and the distorted images, the saliency of each visual orientation pattern is used to create a global saliency based histogram (GSBH) for comparison. Finally, by combining the LSWH and GSBH, a novel RR IQA model is proposed.

II. RR IQA Modeling

In this section, the LSWH based image content is firstly extracted with the local saliency and the local visual pattern. Next, the influence of distortion on the global saliency is calculated with GSBH. Finally, considering the degradation from both LSWH and GSBH, the visual quality degradation is measured and a novel RR IQA model is created. The flowchart of the proposed RR IQA is shown in Fig. 1.

[1]Jinjian Wu (jinjian.wu@mail.xidian.edu.cn), Yongxu Liu, and Guangming Shi are with School of Electronic Engineering, Xidian University, Xi'an, Shannxi, China

[2]Weisi Lin is with School of Computer Engineering, Nanyang Technological University, Singapore

*This research was partially supported by the Natural Science Foundation of Shaanxi Provincial Department of Education (No. 2017JM6055) and the Rapid-Rich Object Search (ROSE) Lab at the Nanyang Technological University, Singapore. The ROSE Lab is supported by the Infocomm Media Development Authority, Singapore.

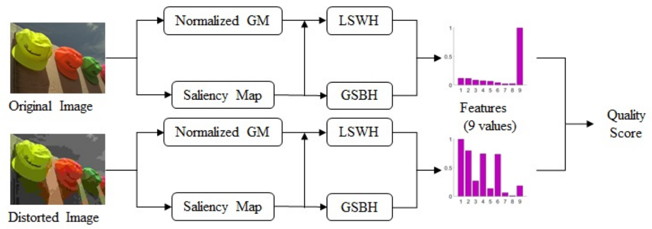

Fig. 1: The flowchart of the proposed RR IQA

.

A. LSWH based Visual Content Extraction

The saliency regions of an image normally refer to object regions with informative contents, which play a more important role for image perception. Thus, we try to weight the local visual pattern of each pixel with its saliency value, and extract the visual content of the whole image with LSWH.

Researches on cognitive science indicated that an orientation selectivity mechanism (OSM) is possessed by the HVS for visual pattern extraction. By analyzing the relationship between a pixel x and its neighborhood $\{x_i|i=1,...n\}$, the local visual pattern of a pixel is created in [9],

$$\mathcal{P}(x)=\mathcal{A}(\mathcal{R}(x|x_1,\cdots,x_n))=\{\mathcal{R}(x|x_1),\cdots,\mathcal{R}(x|x_n)\}, \tag{1}$$

where n is the number of the neighbor pixels (which is set as 8 in this work (same as that in [9])), and $\mathcal{R}(x|x_i)$ is the relationship between two pixels,

$$\mathcal{R}(x|x_i)=\begin{cases}1 & \text{if } |\mathcal{G}_d(x)-\mathcal{G}_d(x_i)|<\mathcal{T}\\ 0 & \text{otherwise}\end{cases}, \tag{2}$$

where $\mathcal{G}_d$ is the gradient direction for each pixel.

Next, the gradient magnitude (GM) of each pixel $\hat{\mathcal{M}}(x)$ is calculated as

$$\hat{\mathcal{M}}(x)=\sqrt{(\hat{\mathcal{M}}_v(x))^2+(\hat{\mathcal{M}}_h(x))^2}. \tag{3}$$

where $\hat{\mathcal{M}}_v$ and $\hat{\mathcal{M}}_h$ are the gradient changes along vertical and horizontal directions, respectively. To reduce the uncertainties caused by illumination changes and the dependency on local image content, a joint adaptive normalization (JAN) procedure is adopted as that in [4] to obtain the normalized gradient magnitude, which is represented as $\mathcal{M}(x)$. Thus, the local pattern for pixel x can be acquired, i.e., $\{\mathcal{P}(x),\mathcal{M}(x)\}$.

Then, we estimate the importance of each pixel with its saliency value. In the past decades, a large amount of saliency estimation models have been proposed. In this work, the graph-based saliency model [12] is adopted, and the saliency value for each pixel is represented as $\mathcal{S}(x)$.

By weighting the local visual pattern of each pixel with its corresponding the local saliency value, the local visual content of an image is represented as a LSWH,

$$\text{LSWH}_c(k)=\sum_{m=1}^{M}S(x_m)\cdot\mathcal{M}(x_m)\cdot\delta(\mathcal{P}(x_m),\mathcal{P}_k) \tag{4}$$

$$\delta(\mathcal{P}(x_m),\mathcal{P}_k)=\begin{cases}1 & \text{if } \mathcal{P}(x_m)=\mathcal{P}_k\\ 0 & \text{otherwise}\end{cases}, \tag{5}$$

where $\mathcal{LSWH}_c(k)$ represents the energy value of the k-th bin, M is total number of pixels in the image, and $\mathcal{P}_k$ means the k-th local pattern.

B. GSBH based Saliency Extraction

The distortion will degrade the structures of objects in an image, which may not only affect the saliency value of a region, but also change the saliency regions between the reference and distorted images. For example, with strong JPEG compression noise, the blockiness of the smooth region becomes obvious, and such region will attract our attention (which is not salient in the reference image). Thus, we suggest to consider the saliency change as a new feature for quality assessment.

Since the RR IQA aims to use less value and accurately measure the image quality, we try to represent the saliency characteristic of an image with several values. Here, the saliency image is mapping in to a histogram based on the pattern distribution, and the GSBH is acquired as,

$$\text{GSBH}_c(k)=\sum_{m=1}^{M}S(x_m)\cdot\delta(\mathcal{P}(x_m),\mathcal{P}_k) \tag{6}$$

and thus, the saliency image is mapped into a vector with K features.

C. Visual Quality Degradation Measurement

The quality is finally measured as the changes from both LSWH and GSBH. With Eq. (4), the local visual contents of the reference $\mathcal{I}^r$ and distorted $\mathcal{I}^d$ images are extracted, i.e., LSWH^r and LSWH^d. Meanwhile, their global saliencies are extracted with Eq. (6), namely, GSBH^r and GSBH^d. Considering the degradation on both parts, the quality is as their similarity,

$$\mathcal{Q}(\mathcal{I}^d|\mathcal{I}^r)=\frac{1}{N}\sum_{k=1}^{N}\frac{2\cdot\text{F}_c^r(k)\cdot\text{F}_c^d(k)+c}{\text{F}_c^r(k)^2+\text{F}_c^d(k)^2+c} \tag{7}$$

where c is a constant to avoid the denominator being zero, and is set as 0.01 in this work, and F is,

$$\text{F}_c^j(k)=\text{LSWH}_c^j(k)^\alpha\cdot\text{GSBH}_c^j(k)^\beta. \tag{8}$$

where α and β are used to weight LSWH and GSBH, which are empirically set as 1.7 and 0.56 respectively in this paper.

III. Experimental Result Analysis

In this section, we firstly analyze the efficiency of the proposed model with a visualized example. Then, we compare the proposed model with these existing RR IQA methods to further illustrate the efficiency of our model.

(a) ORG

(b) GBN, MOS=4.474

(c) J2K, MOS=2.846

(d) JPG, MOS=2.487

Fig. 2: An example of efficiency analysis on the *Caps* image

.

A. *Efficiency Analysis*

By considering the visual saliency for quality prediction, the proposed model can highlight the object regions and suppress the background for quality assessment. Such procedure performs more consistently with the HVS. In order to demonstrate the effectiveness of the proposed model, a latest RR IQA model (i.e., the OSVP [9]) is firstly chosen for comparison. The comparison result is shown in Fig. 2. In this test, the reference *Hats* image is contaminated by three types of distortions, which are the Gaussian blur noise (GBN) in Fig. 2 (b), the JPEG2000 compression noise (J2K) in Fig. 2 (c), and the JPEG compression (JPG) in Fig. 2 (d). The MOS (mean opinion score, larger MOS represents better quality) for them are 4.474, 2.846, and 2.487, respectively, namely, Fig. 2 (b) has the best quality, and Fig. 2 (d) has the worst quality.

Without considering the saliency importance, the OSVP model equally processes the whole image. As a result, the prediction results from OSVP (on Fig. 2 (b)-(d)) are not consistent with subjective perception. The predicted qualities of OSVP (larger value represents better quality) for Fig. 2 (b)-(d) are 0.922, 0.944, and 0.923, respectively, which means Fig. 2 (c) has the best quality, and Fig. 2 (b) has the worst quality. Meanwhile, it is also conveyed from the prediction results that Fig. 2 (d) has a much similar score of visual quality with Fig. 2 (b).

By considering the saliency degradation, the proposed model can accurately predict the qualities of the three distorted images. The predicted qualities for them are 0.867, 0.767, and 0.655, respectively, which means Fig. 2 (b) is the one with best quality and Fig. 2 (d) is the worst one (performs consistently with subjective perception).

B. *IQA Performance*

For a comprehensive analysis, we further compare the proposed model with three latest RR IQA methods (i.e., WNISM [6], RRED [10], and OSVP [9]) and two classical FR IQA methods (PSNR and MS-SSIM [2]) on five databases (i.e., CSIQ, LIVE, TID2013, IVC, and TOYAMA). The number of values required for these RR IQA models (WNISM, RRED, OSVP, and the proposed) are 18, 9, 9, and 9, respectively. In order to measure the quality prediction accuracy from different models, three metrics are adopted, which are the Spearman rank order correlation coefficient

TABLE I: IQA Performance Comparison on Five Large Databases and Average

Dist.	Algo.	RR				FR	
	Dist.	Proposed	OSVP	RRED	WNISM	PSNR	MS-SSIM
CSIQ	PLCC	**0.900**	0.843	0.758	0.737	0.800	0.899
	SRCC	**0.901**	0.849	0.777	0.757	0.800	0.913
	RMSE	**0.115**	0.141	0.172	0.178	0.158	0.115
LIVE	PLCC	**0.907**	0.862	0.830	0.743	0.872	0.941
	SRCC	**0.912**	0.867	0.834	0.755	0.876	0.951
	RMSE	**11.53**	13.84	15.24	18.28	13.36	9.259
TID2013	PLCC	**0.779**	0.724	0.748	0.629	0.702	0.831
	SRCC	**0.755**	0.654	0.689	0.523	0.703	0.786
	RMSE	**0.778**	0.856	0.822	0.964	0.883	0.690
IVC	PLCC	**0.818**	0.757	0.651	0.534	0.720	0.911
	SRCC	**0.811**	0.747	0.625	0.446	0.688	0.898
	RMSE	**0.702**	0.795	0.925	1.030	0.846	0.503
TOYAMA	PLCC	**0.868**	0.827	0.745	0.779	0.731	0.915
	SRCC	**0.858**	0.825	0.742	0.774	0.722	0.907
	RMSE	**0.656**	0.743	0.880	0.828	0.901	0.532
Weighted Average	PLCC	**0.825**	0.771	0.759	0.668	0.747	0.866
	SRCC	**0.811**	0.731	0.726	0.606	0.747	0.842
	RMSE	**2.323**	2.738	2.951	3.509	2.691	1.906

(SRCC), The Pearson linear correlation coefficient (PLCC) and the root mean squared error (RMSE).

The performance of these IQA models are listed in Tab. I. By comparing with these existing RR IQA models, we can see that the proposed model performs the best on all of the five databases. As can be seen, the proposed model has achieved a remarkable improvement on CSIQ and LIVE database (for which both the PLCC and SRCC values are larger than 0.9). And the proposed model has also achieved an obvious improvement on the rest three databases (i.e., TID2013, IVC, and TOYAMA). The weighted average performance on the five databases are listed at the bottom right of Tab. I, which also confirms that the proposed model performs obviously better than the other RR IQA models.

Moreover, though the proposed model only uses 9 values for quality assessment, its performance is comparable with these classic FR IQA models. As shown in Tab. I, the proposed model outperforms the PSNR on all of the five databases. And by comparing with MS-SSIM, the proposed model performs similar with it on CSIQ, while worse on the other databases. In summary, the proposed RR IQA model achieves a remarkable improvement from these existing RR IQA models, and performs highly consistent with subjective perception.

IV. CONCLUSION

In this work, we have proposed a novel RR IQA model with saliency degradation. The visual processing of the HVS has been firstly investigated, and the visual attention has been adopted for IQA model designing. Since the saliency regions attract more attentions and play more important role for visual perception, we have tried to highlight these regions for IQA. By weighting the local visual orientation pattern of each pixel with its saliency value, the local visual feature was firstly extracted, and then mapped into a LSWH. Besides, strong distortions will always directly change the saliency regions (different saliency regions between the reference and distorted images). Thus, the saliency of each type of visual orientation pattern has been measured, and a GSBH has been created to represent the global saliency degradation. Finally, by combining the LSWH and GSBH, a novel model for RR IQA is created. Experimental results on five public databases demonstrate that our model uses only 9 values of the reference information and performs consistently with the HVS perception.

REFERENCES

[1] J. Wu, W. Lin, G. Shi, and A. Liu, "Perceptual quality metric with internal generative mechanism," *IEEE Transactions on Image Processing*, vol. 22, no. 1, pp. 43 –54, Jan. 2013.

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

[3] X. Gao, W. Lu, D. Tao, and X. Li, "Image quality assessment based on multiscale geometric analysis," *IEEE Transactions on Image Processing*, vol. 18, no. 7, pp. 1409–1423, Jul. 2009.

[4] W. Xue, X. Mou, L. Zhang, A. C. Bovik, and X. Feng, "Blind image quality assessment using joint statistics of gradient magnitude and laplacian features," *IEEE Transactions on Image Processing*, vol. 23, no. 11, pp. 4850–4862, 2014.

[5] J. Wu, W. Lin, G. Shi, and A. Liu, "Reduced-reference image quality assessment with visual information fidelity," *IEEE Transactions on Multimedia*, vol. 15, no. 7, pp. 1700–1705, Nov. 2013.

[6] Z. Wang and E. P. Simoncelli, "Reduced-reference image quality assessment using a wavelet-domain natural image statistic model," in *SPIE*, vol. 5666, 2005, pp. 149–159.

[7] Q. Li and Z. Wang, "Reduced-reference image quality assessment using divisive normalization-based image representation," *IEEE Journal of Selected Topics in Signal Processing*, vol. 3, no. 2, pp. 202–211, Apr. 2009.

[8] L. Ma, S. Li, F. Zhang, and K. N. Ngan, "Reduced-reference image quality assessment using reorganized dct-based image representation," *IEEE Transactions on Multimedia*, vol. 13, no. 4, pp. 824–829, Aug. 2011.

[9] J. Wu, W. Lin, G. Shi, L. Li, and Y. Fang, "Orientation selectivity based visual pattern for reduced-reference image quality assessment," *Information Sciences*, vol. 351, pp. 18–29, Jul. 2016.

[10] R. Soundararajan and A. Bovik, "RRED indices: Reduced reference entropic differencing for image quality assessment," *IEEE Transactions on Image Processing*, vol. 21, no. 2, pp. 517 –526, Feb. 2012.

[11] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, Nov. 1998.

[12] J. Harel, C. Koch, P. Perona *et al.*, "Graph-based visual saliency," in *NIPS*, vol. 1, no. 2, 2006, p. 5.