Foveation-Based Image Quality Assessment

Wen-Jiin Tsai* and Yi-Shih Liu

Department of Computer Science, National Chiao Tung University, Taiwan, R.O.C.

wjtsai@cs.nctu.edu.tw

liuys614@gmail.com

***Abstract*— Since human vision has much greater resolutions at the center of our visual field than elsewhere, different criteria of quality assessment should be applied on the image areas with different visual resolutions. This paper proposed a foveation-based image quality assessment method which adopted different sizes of windows in quality assessment for a single image. Visual salience models which estimate visual attention regions are used to determine the foveation center and foveation resolution models are used to guide the selection of window sizes for the areas over spatial extent of the image. Finally, the quality scores obtained from different window sizes are pooled together to get a single value for the image. The proposed method has been applied to IQA metrics, SSIM, PSNR, and UQI. The result shows that both Spearman and Kendall correlation coefficients can be improved significantly by our foveation-based method.**

***Index Terms*— Image quality assessment, foveation, human visual system, visual salience model.**

I. INTRODUCTION

Image/video quality assessment is important for some applications. PSNR and MSE, due to their simple computation, have become widely used metrics in image quality assessment. However, many research results show that neither PSNR nor MSE can always produce quality assessment results similar to those by human eyes. Images of the same PSNR values can have huge differences in perceptual quality. Therefore, a plenty of researches have focused on designing perception based image quality assessment (IQA) methods [1-4].

However, most of IQA methods process the whole image in the same way. This seems not in consistent with how human eyes judge the quality of an image. In general, people don't measure the quality of the whole image uniformly using the same criteria. Therefore, some of recent studies proposed to incorporate visual attention techniques into image/video quality metrics [5-8]. These methods either use eye-tracking systems or visual salience modeling to find out visual salience regions that attract our attention and then give these regions more importance and thus contribute more to quality score of the whole image. Different methods adopt different visual salience modeling and different weight assignment methods. Namely, except to different salience regions and weights used, they assess quality in the same way for the whole image.

It was well known that density of light-receptor cells decreases progressively from the center of the retina, known as the fovea, to the edges of the retina. As a result, when our eyes focus on visual salience regions, we can resolve pixels with very high resolutions in that area; in contrast, for the area outside or even far from the salience regions, since they are at the edges of our visual field, our eyes will resolve pixels with low resolutions. Hence, regions located in different vision resolutions should adopt different ways to measure their qualities, rather than using the same ways with different weights only. In this paper, a foveation-based image quality assessment method is proposed, which takes both visual attention and vision resolutions into considerations.

The paper is organized as follows. Section 2 presents the proposed method from three aspects: visual saliency models, quality assessment criteria selection, window size selection and quality score pooling. Section 3 evaluates the performance of applying the proposed method on existing IQA metrics: PSNR, SSIM, and UQI. Finally, conclusions are given in Section 4.

II. THE PROPOSED METHOD

To exploit the foveation of human visual system (HVS), we propose a foveated IQA method which is based on the idea that image quality should be measured with different criteria according to foveation of human eyes. Since our vision has much greater spatial resolutions at the center of our visual field than elsewhere, in our method, areas with different visual resolutions will apply different quality assessment criteria. The areas with high visual resolutions will adopt high(strict) quality assessment criteria; while the areas with low visual resolutions will use low(loose) criteria.

As for how to use a single IQA metric to assess an image with different criteria, we propose to use different window sizes for quality assessment. Since image statistical features are usually space-variant, most image quality assessment methods apply quality index locally rather than globally. In UQI[1], local statistics such as luminance distortion, contrast distortion, and loss of correlation are computed within a local 8x8 window, which moves pixel-by-pixel over the entire image. In SSIM[2], they modified the estimates of local statistics in UQI by using an 11x11 circular-symmetric Gaussian weighting function. In PSNRHVS[9] and PSNRHVSM[10], images are divided into non-overlapping 8x8 blocks. Every 8x8 block is processed independently and contributes to image quality metric equally.

Using small windows for local statistics can capture space variances in detail and hence serve as strict quality criteria;

978-1-4799-6139-9/14/$31.00 ©2014 IEEE

while using large windows can only capture them roughly and hence serve as loose criteria. As an example in Fig.1, the block in red edges will get a high SSIM score, 0.8025, when it is measured with a large window size 64x64; however, it will only get a low SSIM score 0.6346 when it is measured with a small window size 16x16. Therefore, in our approach, the image areas with high visual resolutions will adopt small windows (strict criteria) for quality assessment; while the areas with low visual resolutions will use large windows (loose criteria). We use visual salience modeling techniques to estimate visual attention regions, which are then regarded as the areas of high vision resolutions in the images and are used to guide window size selection. Fig. 2 shows the flowchart of the proposed method. In the following, we present our method mainly from three aspects: visual saliency modeling, quality assessment criteria selection, and quality score pooling.

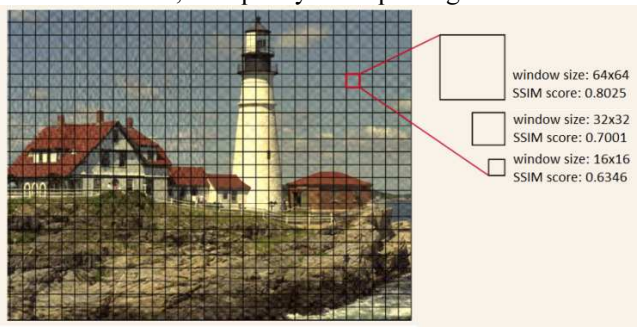

Fig. 1 Quality assessment with different window sizes.

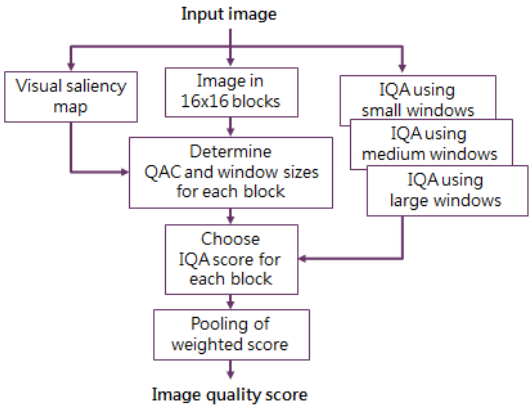

Fig. 2 Flow chart of the proposed method.

A. Visual Saliency Modeling

In Fig.2, for each image, we first calculate its visual saliency region which is the image areas that are likely to catch the attention of a human observer. There are already many different kinds of algorithms proposed for visual saliency computation in the literature. For a recent comprehensive survey of this field, readers can refer to [11, 12]. Actually, the proposed foveated IQA metric is not limited to any specific visual saliency model. Any model which can, in an image, correctly identify the areas of human visual attention can be adopted in our method. But, of course, the more correct the visual saliency model is, the better results the proposed method will obtain. The visual saliency model that we adopted in the experiments is based on Cheng's approach [13] which has been proved to be simple, effective, and can generate satisfactory results in most cases.

B. Quality Assessment Criteria Selection

We use the image's visual saliency map to guide foveated IQA criteria selection. First, the image is divided into non-overlapped blocks. Then, for each block, the mean visual saliency, S_{mean}, is computed as follows:

$$S_{mean}(B_n) = \frac{1}{h \times w} \sum_{y}^{W} \sum_{x}^{H} s(x, y) \quad (1)$$

where W and H are the width and the height of the block B_n, and $s(x, y)$ is the salience value of the pixel located at position (x, y) in block B_n. In our approach, a block will be assessed with *high*(strict) quality criteria if its S_{mean} is higher than 170 (the top 1/3 values in the range of 0-255). It will be assessed with *medium* quality criteria if it is located in the area surrounding the saliency region, with 64-pixel width along the boundary of saliency region. Otherwise, it will be assessed with *low*(loose) criteria.

Since blocks on the boundary of salient objects may obtain a low mean visual saliency, a 2-pass algorithm shown in Fig.3 is proposed to refine the criteria selection for these blocks. In the algorithm, the operation is performed for every block in the image in a raster scan order, tending to dilate the areas of high criteria and medium criteria. Fig. 4 shows the quality criteria determined for each block in an image from TID2008 database.

```
// 2-pass algorithm
for( each 16x16 block in the image, in a raster scan order) {
    if (any of its eight neighbors adopts high criteria)
        then current block adopts high criteria
    else if (any of its eight neighbors adopts medium criteria)
        then current block adopts medium criteria
    else current block adopts low criteria.
}
```

Fig. 3 The algorithm to improve criteria selection for salient object boundary.

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

Fig. 4 Window size selection (a) original image (b) visual salience map using Cheng's method (c) saliency region (d) regions of three different quality criteria (e) refined quality criteria for salient object boundary using the 2-pass algorithm.

C. Window Size Determination

Window size selection is based on formula (2), the foveation resolution model proposed in [15], where d_x is the distance between a pixel x on an image and the position of foveation point; and f_m is the cutoff frequency which means that any higher frequency component beyond it is invisible. The cutoff frequency plays an important role in determining the resolution ability of the human eyes.

$$f_m(d_x) = \min\left(\frac{e_2 \ln(1/CT_0)}{\alpha[e_2 + tan^{-1}(d_x)/Nv)]}, \frac{\pi Nv}{360}\right) \quad (2)$$

where e is retinal eccentricity (degrees), e2 is half-resolution eccentricity constant, CT_0 is the minimal contrast threshold, α is spatial frequency decay constant, N is image wide (in pixels), and v is the viewing distance (in image width). The best fitting parameter values given in [15] are α=0.106, e2=2.3, CT0=1/64. We used the same parameter values and assumed N=512, v=10 in our experiments. Let W_{high}, W_{medium}, and W_{low} be the window sizes for the regions of high, medium and low quality criteria, respectively. We select W_{high} to be the same to the window sizes used in the original methods (e.g., 11 for SSIM[2], 8 for UQI[1], and 2 for PSNR) and select W_{medium} and W_{low} to be inversely proportional to cutoff frequency using formula (2). That is,

$$\frac{W_{medium}}{W_{high}} = \frac{f_m(D_{high})}{f_m(D_{medium})} \quad (3)$$

$$\frac{W_{low}}{W_{high}} = \frac{f_m(D_{high})}{f_m(D_{low})} \quad (4)$$

where $D_{high} = 0$ because we assume the whole salience region is at the center of foveation; D_{medium} and D_{low} are the average distance to the boundary of salience region from the regions of medium and low quality criteria, respectively. Note that D_{medium} is 32 because the region of medium quality criteria is 64-pixel width surrounding saliency region. As for D_{low}, since it costs too high to measure the distance of every pixel in the region of low quality criteria to the boundary of saliency region, we simplify it as $D_{low} = (D_{left}+D_{right})/2$, where D_{left} and D_{right} are illustrated in Fig.5 and defined as follows

$$D_{left} = (x_{leftmost} - 0 - 64)/2 + 64 \quad (5)$$
$$D_{right} = (N - x_{rightmost} - 64)/2 + 64 \quad (6)$$

where $x_{rightmost}$ and $x_{leftmost}$ are the rightmost and leftmost pixels in saliency region, respectively. N is image width (in pixels).

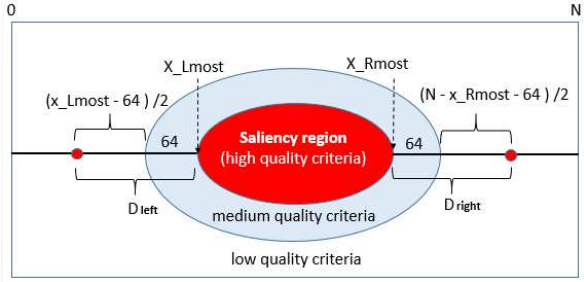

Fig. 5 Illustration of D_{left} and D_{right}.

D. Quality Score Pooling

The final stage of the proposed method is to combine the quality scores of blocks (sc_i) over the spatial extent of the image into a single value (called SC). Without weighting, the pooling simply takes mean scores of the blocks as the final score. With weighting, three weighting factors, wt_H, wt_M, and wt_L indicating the relative importance of quality assessment criteria are used to combine the scores as

$$SC = wt_H \sum_{i=0}^{N_H} \frac{sc_i}{N_H} + wt_M \sum_{i=0}^{N_M} \frac{sc_i}{N_M} + wt_L \sum_{i=0}^{N_L} \frac{sc_i}{N_L} \quad (6)$$

where N_H, N_M, and N_L denote the number of blocks that are assessed with high, medium, and low criteria, respectively. We select the three weighting factors to be proportional to cutoff frequency as below, subject to $wt_H + wt_M + wt_L = 1$.

$$\frac{wt_M}{wt_H} = \frac{f_m(D_{medium})}{f_m(D_{high})} \quad (7)$$

$$\frac{wt_L}{wt_H} = \frac{f_m(D_{low})}{f_m(D_{high})} \quad (8)$$

III. Experimental Results

To evaluate the effects of applying the proposed foveation-based approach on image quality assessment, three IQA metrics are adopted: SSIM, PSNR, and UQI, two of which are subjective methods and the other is objective. There are 9 variations of IQA methods in comparison. The names with prefix f- denote the proposed Foveated methods without weightings, e.g., f-SSIM, f-PSNR, and f-UQI; while the names with prefix wf- denote the proposed Foveated methods with weightings, e.g., wf-SSIM, wf-PSNR, and wf-UQI.

To evaluate the performance of various IQA methods, our experiments adopted the images in TID2008 [14] and TID2013 [16] databases. TID2013 is a database extended from TID2008. Both of them are intended for full-reference image quality verification and are available for free downloads. The TID2008 contains 1700 images (25 reference images x 17 distortion types x 4 distortion levels) which are grouped all together in a *full* subset, or into different subsets including *Noise, Noise2, Safe, Hard, Simple, Exotic, Exotic2* with different distortions. Each subset may contain more than one distortion type. Compared to TID2008, the TID2013 database includes one more level and seven new types of distortions which are considered to be important from theoretical and practical viewpoints [16]. It contains 3000 images, grouped all together in a *full* subset, or into different subsets including *Noise, Actual, Simple, Exotic, New,* and *Color*. The subjective experiments in both TID2008 and TID2013 were calculated as MOS (Mean Opinion Score), ranging from 0 to 9. The higher the MOS is, the higher the visual quality of the images is.

A. Experiment Results

Spearman correlation (SCRR) and Kendall correlation (KCRR) coefficients are two indexes used in image quality assessment to compute the correlation of objective measures with human perception. To compare the accuracy of the proposed image quality metric, we compute SCRR and KCRR

coefficients of each method under comparison. The results are listed in Tables 1 and 2, where Δ(%) is the enhancement performance of wf-SSIM, wf-PSNR and wf-UQI compared to their original methods, SSIM, PSNR and UQI, respectively.

Compared to original SSIM, it is observed that f-SSIM performed better in terms of SCRR and KCRR coefficients for all the image subgroups in TID2008, especially for *Exotic*. The performance is enhanced more obviously in TID2013 than in TID2008. Compared to original PSNR, the proposed f-PSNR performed much better, especially for TID2008, even though it is slightly degraded on Exotic2. For f-UQI, it performed better than UQI for both TID2008 and TID2013. The results of f-SSIM, f-PSNR, and f-UQI show that the IQA performance can be enhanced by applying different quality assessment criteria on different regions of the image according to foveation.

Compared with f-SSIM, f-PSNR and f-UQI, the wf-SSIM, wf-PSNR, and wf-UQI show even better performance for both TID2008 and TID2013, no matter in terms of SCRR or KCRR coefficients. Both wf-SSIM and wf-UQI show significant improvement on *Exotic* in both TID2008 and TID2013; while wf-PSNR on *Exotic* in TID2008 and *Color* in TID2013. Compared to original SSIM, wf-SSIM improves SCRR up to 83.7% and KCRR up to 88.8%; compared to original PSNR, wf-PSNR improves SCRR up to 75% and KCRR up to 58.4%; while compared to original UQI, wf-UQI improves SCRR up to 92.81% and KCRR up to 99.49%. The results show that, with appropriate weightings, the Foveated IQA can further improve the performance.

IV. Conclusion

In many IQA methods, images are usually divided into windows for quality assessment in order to capture pace variances. Using small windows for local statistics can capture space variances in detail; while using large windows can only capture them roughly. The proposed foveated method adopts different window sizes in quality assessment to simulate the different visual resolutions when observers watch an image. The results show that the performance of a window-based IQA metric can be improved significantly by taking human foveation into considerations.

TABLE I
RESULTS FOR TID2008 DATABASE

SCRR	SSIM	f-SSIM	wf-SSIM	Δ(%)	PSNR	f-PSNR	wf-PSNR	Δ(%)	UQI	f-UQI	wf-UQI	Δ(%)
Noise	0.81	0.82	0.83	2.2%	0.70	0.83	0.84	19.5%	0.53	0.53	0.55	5.32%
Noise2	0.85	0.85	0.86	1.8%	0.61	0.76	0.78	28.1%	0.60	0.61	0.62	3.51%
Safe	0.86	0.86	0.87	1.6%	0.69	0.81	0.83	21.0%	0.64	0.65	0.66	3.92%
Hard	0.87	0.87	0.88	1.0%	0.70	0.81	0.80	15.2%	0.76	0.76	0.77	0.92%
Simple	0.90	0.91	0.92	1.9%	0.80	0.87	0.89	11.0%	0.78	0.80	0.82	4.08%
Exotic	0.36	0.37	0.65	83.7%	0.25	0.23	0.43	75.0%	0.29	0.21	0.56	92.81%
Exotic2	0.60	0.60	0.76	26.5%	0.31	0.26	0.30	-1.3%	0.55	0.51	0.69	27.11%
Full	0.77	0.78	0.83	7.3%	0.53	0.57	0.57	8.8%	0.60	0.59	0.65	8.83%
KCRR	SSIM	f-SSIM	wf-SSIM	Δ(%)	PSNR	f-PSNR	wf-PSNR	Δ(%)	UQI	f-UQI	wf-UQI	Δ(%)
Noise	0.60	0.62	0.63	4.0%	0.50	0.62	0.64	28.5%	0.36	0.37	0.39	6.06%
Noise2	0.64	0.66	0.67	3.4%	0.42	0.56	0.58	37.5%	0.42	0.43	0.44	4.29%
Safe	0.66	0.67	0.68	3.3%	0.49	0.61	0.64	30.7%	0.45	0.46	0.48	4.63%
Hard	0.67	0.68	0.68	1.8%	0.52	0.62	0.62	19.2%	0.57	0.57	0.57	1.42%
Simple	0.72	0.73	0.75	3.5%	0.60	0.68	0.71	18.2%	0.59	0.60	0.62	5.79%
Exotic	0.24	0.25	0.46	88.8%	0.18	0.17	0.28	58.4%	0.20	0.15	0.39	99.49%
Exotic2	0.43	0.43	0.56	30.8%	0.23	0.20	0.20	-10.7%	0.39	0.36	0.51	29.82%
Full	0.57	0.58	0.63	9.6%	0.37	0.42	0.42	13.0%	0.44	0.43	0.48	9.20%

TABLE II
RESULTS FOR TID2013 DATABASE

SCRR	SSIM	f-SSIM	wf-SSIM	Δ(%)	PSNR	f-PSNR	wf-PSNR	Δ(%)	UQI	f-UQI	wf-UQI	Δ(%)
Noise	0.76	0.87	0.87	15.5%	0.82	0.85	0.86	4.7%	0.65	0.67	0.68	4.4%
Actual	0.79	0.88	0.89	12.9%	0.83	0.88	0.90	8.5%	0.69	0.71	0.72	3.3%
Simple	0.84	0.91	0.92	10.2%	0.91	0.91	0.92	0.5%	0.77	0.77	0.79	2.5%
Exotic	0.63	0.74	0.82	29.7%	0.60	0.55	0.56	-5.7%	0.53	0.55	0.65	22.7%
New	0.58	0.62	0.62	6.9%	0.62	0.81	0.80	28.8%	0.50	0.51	0.52	4.2%
Color	0.51	0.55	0.55	8.9%	0.54	0.77	0.76	41.9%	0.46	0.46	0.47	2.4%
Full	0.64	0.74	0.77	21.5%	0.64	0.71	0.70	10.0%	0.55	0.56	0.61	10.4%
KCRR	SSIM	f-SSIM	wf-SSIM	Δ(%)	PSNR	f-PSNR	wf-PSNR	Δ(%)	UQI	f-UQI	wf-UQI	Δ(%)
Noise	0.55	0.67	0.68	23.4%	0.62	0.65	0.67	7.4%	0.46	0.48	0.49	5.2%
Actual	0.58	0.69	0.70	21.7%	0.62	0.70	0.71	13.9%	0.50	0.51	0.52	4.0%
Simple	0.63	0.73	0.75	18.6%	0.75	0.74	0.75	1.1%	0.56	0.56	0.58	3.8%
Exotic	0.46	0.55	0.62	37.1%	0.43	0.39	0.38	-9.6%	0.38	0.39	0.47	24.3%
New	0.42	0.47	0.48	14.1%	0.47	0.61	0.61	31.2%	0.35	0.36	0.37	4.8%
Color	0.38	0.43	0.43	13.2%	0.41	0.57	0.57	38.7%	0.32	0.32	0.33	3.4%
Full	0.46	0.56	0.59	27.9%	0.47	0.53	0.53	11.7%	0.40	0.41	0.44	10.6%

References

[1] Z. Wang, A. Bovik, "A universal image quality index", IEEE Signal Processing Letters, vol. 9, pp. 81-84 (2002)

[2] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[3] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006.

[4] D. M. Chandler and S. S. Hemami, "VSNR: A wavelet-based visual signal-to-noise ratio for natural images," *IEEE Trans. Image Process.*, vol. 16, no. 9, pp. 2284–2298, Sep. 2007.

[5] Y. Tong, F. Alaya Cheikh, A. Tremeau and H. Konick, "Full Reference Image Quality Assessment Based on Saliency Map Analysis," Accepted for publication in the International Journal of Imaging Systems and Technology, in 2010.

[6] A. Ninassi , O. L. Meur , P. L. Callet and D. Barbba "Does where you gaze on an image affect your preception of quality? applying visual attention to image quality metric", *Proc. IEEE Int. Conf. Image Process.*, vol. 2, pp.169 -172 2007.

[7] H. Liu and I. Heynderickx, "Visual attention in objective image quality assessment: based on eye-tracking data," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 7, pp. 971–982, 2011.

[8] A. Guo, D. Zhao, L. Shaohui, X. Fan, and W. Gao, "Visual attention based image quality assessment," in *Proceedings of the IEEE International Conference on Image Processing (ICIP '11)*, pp. 3297–3300, September 2011.

[9] Egiazarian K., Astola J., Ponomarenko N., Lukin V., Battisti F., Carli M. "New full-reference quality metrics based on HVS", CD-ROM Proceedings of the Second Int. Workshop on Video Proc. and Quality Metrics, Scottsdale, USA, 2006, 4 p.

[10] Ponomarenko N., Silvestri F., Egiazarian K., Carli M., Astola J., Lukin V. "On between-coefficient contrast masking of DCT basis functions", CD-ROM Proc. of the Third Int. Workshop on Video Proc. and Quality Metrics, USA, 2007, 4 p.

[11] A. Toet, "Computational versus psychophysical bottom-up image saliency: a comparative evaluation study," *IEEE Trans. PAMI*, vol. 33, pp. 2131-2146, 2011.

[12] A. Borji and L. Itti, "State-of-the-art in visual attention modeling," *IEEE Trans. PAMI*, vol. 35, pp. 185-207, 2013.

[13] M.-M. Cheng, G.-X. Zhang, N. J. Mitra, X. Huang, and S.-M. Hu, "Glogal contrast based salient region detection," In CVPR, pp.409-416, 2011

[14] Ponomarenko N., Carli M., Lukin V., Egiazarian K., Astola J., Battisti F. "Color Image Database for Evaluation of Image Quality Metrics", Proc. of Intern. Workshop on Multimedia Signal Processing, Australia, Oct. 2008, pp. 403-408.

[15] Z. Wang and A. C. Bovik. "Embedded Foveation Image Coding", *IEEE Trans. Image Processing,* Vol. 10, No. 10, pp. 1397-1410, Oct., 2001.

[16] N. Ponomarenko "Color image database TID2013: Peculiarities and preliminary results", *Proc. 4th Eur. Workshop Visual Inf. Process.*, pp.1 -6 2013.