Visual Attention Data for Image Quality Assessment Databases

Xiongkuo Min, Guangtao Zhai, Zhongpai Gao and Ke Gu
Institute of Image Communication and Information Processing, Shanghai Jiao Tong University, Shanghai, China
Shanghai Key Laboratory of Digital Media Processing and Transmissions
Email: {minxiongkuo, zhaiguangtao, gaozhongpai, gukesjtuee}@sjtu.edu.cn

Abstract— **Images usually contain areas that particularly attract people's attention and visual attention is an important feature of human visual system (HVS). Visual attention had been shown to be effective in improving performance of existing image quality assessment (IQA) metrics. However, with the quick advancement of IQA research, the booming of open IQA databases calls for associated comprehensive and accurate visual attention dataset. Despite of the large number of existing computational attention/saliency models, the most accurate measure of human attention is still human based. In this research, we first conduct extensive eye tracking experiments for all the pristine images from the seven widely used IQA databases (LIVE, TID2008, CSIQ, Toyama, LIVE Multiply Distortion, IVC and A57 databases). Then we propose a gaze-duration adaptive weighting approach to generate saliency maps from the eye tracking data. When applied on the IQA databases, experimental results suggest that accuracy of benchmark quality metrics, e.g. PSNR and SSIM can be systematically improved, outperforming existing saliency datasets. Both the eye tracking data and the saliency maps in this research will be made publicly available at `gvsp.sjtu.edu.cn`.**

I. Introduction

It is widely known that the most accurate approach to image quality assessment (IQA) is subjective viewing test. However, it is not always feasible to execute subjective tests in many circumstances. Therefore, much effort has been devoted to develop objective IQA metrics in the last decades. The ultimate goal of objective IQA is to provide computational metrics with results that are well correlated to human ratings. For the convenience of design and compare of different IQA metrics, many image databases were made publicly available [1]–[7]. These image databases consist of a number of unimpaired images and series of distorted images with different types of distortion at different levels. And the mean opinion scores (MOS) or Differential MOS (DMOS) of all these images are also provided in the databases. Since the HVS is the ultimate receiver and valuator of images, researchers tend to take properties of HVS into account. For example, SSIM [8] was designed with the assumption that HVS is good at extracting the structural information from images.

Since visual attention is an important feature of HVS, it is natural to investigate the effectiveness of applying subjective visual attention data into IQA [9]–[11]. Intuitively, those areas that attract more subjective attention should be heavily weighted during the pooling stage of IQA metrics. A. NINASSI *et al.*'s work [11] found that the performance gain was not clear when applying visual attention into IQA. In their test, eye tracking data was collected with the DSIS (Double Stimulus Impairment Scale) method. The visual attention data may not be very accurate because subjects had to view images of the same content frequently. And subjects also had to rate the quality of images during the experiments. On the other hand, more recent research of [9], [10] both verified the validity of visual attention weighting on several IQA databases (LIVE, Toyama and IVC). Liu and Heynderickx [9] conducted two experiments with and without quality rating task for the LIVE database and they found that task-free data tend to be more effective. But during the computation of saliency maps for eye tracking data, they ignored the duration of fixation points.

In this paper, we will provide visual attention data for all the widely used IQA databases including LIVE [1], TID [2], CSIQ [3], Toyama [4], LIVE multi [5], IVC [6] and A57 [7]. More specifically, we performed eye tracking experiments on all the unimpaired images from those databases in a task-free setting. And we employed a duration adaptive approach to generate saliency maps from eye tracking data. When applied on benchmark IQA approaches (PSRN and SSIM), our saliency maps bring higher performance improvement than existing datasets [9]–[11]. Both the eye tracking data and the saliency maps in this research will be made publicly available at `gvsp.sjtu.edu.cn`.

The rest of this paper is organized as follows. In Section II, the eye tracking experiment will be introduced in detail together with the approach to generate saliency maps from eye tracking data. Eye tracking data and the saliency maps will be verified and compared to existing datasets in Section III. Finally, conclusion is drawn in Section IV.

II. Eye Tracking Data and Saliency Maps

A. Eye Tracking Experiment

1) Apparatus: The eye tracking experiments were performed with Tobbi T120 Eye Tracker. Tobbi T120 has a sample frequency of 60 or 120 Hz. The Tobbi T120 Eye Tracker is integrated into a 17 inch TFT monitor, so as to make the user experience more natural. The resolution of the monitor is 1280×1024 pixels. Tobbi T120 typically has a spatial resolution of 0.3 degree and it's typical accuracy is 0.5 degree. It's head movement box (width $\times$ hight) is 30×22 cm at 70 cm, and the tracking distance is $50 \sim 80$ cm. A photo of a eye tracking test with Tobbi T120 is given in Fig.1.

978-1-4799-3432-4/14/$31.00 ©2014 IEEE

Fig. 1. A Test with Tobbi T120 Eye Tracker.

2) Stimuli and Procedures: Five subjects participated in the eye tracking experiment, all of which were college students. The reason why we only include five subjects in the experiment would be discussed in the third paragraph of section III-B. During the tests, all subjects were asked to look freely at the monitor. As mentioned, the pristine images from those seven IQA databases were used as test stimuli. Statistics of the test images are listed in Table I. Note that some repeated images in those databases were excluded from our test.

TABLE I
TEST IMAGES USED IN THE EYE TRACKING EXPERIMENTS

Database name	Image size (W $\times$ H)	Image Number
LIVE	W:480$\sim$768 H:438$\sim$720	29
TID2008	512 $\times$ 384	25
CSIQ	512 $\times$ 512	30
Toyama	768 $\times$ 512	5
LIVE Multiply	1280 $\times$ 720	15
IVC	512 $\times$ 512	10
A57	512 $\times$ 512	3

To prevent from prolonged viewing test and improve test accuracy, we divided all these test images into four groups according to the source databases and image size. Each group consisted of 30 images or less. To reduce fatigue, there was a break for five minutes between the viewing of two test groups. Before the test, a five point calibration was performed for each participant. In each group, images were presented in a random order. Every image was presented on the display for ten seconds, and between two images, there was a five seconds gray screen.

We adopted a single-stimulus (SS) free-looking approach in our eye tracking experiment. The experiment was conducted in a normal lighting condition with a viewing distance of about 65 cm. A comparison of the specifications of those eye tracking experiments is listed in Table II. In this table, 'Ours' is this work, 'Engelke' indicates Engelke *et al.*'s [10], and 'Liu' denotes Liu and Heynderickx's experiments [9].

TABLE II
COMPARISON OF SETTINGS OF EYE TRACKING EXPERIMENTS. SS: SINGLE STIMULUS

Experiment	Ours	Engelke [10]	Liu [9]
Sample rate	60 Hz	40$\sim$45 Hz	50 Hz
Resolution	1280$\times$1080	1280$\times$1080	1024$\times$768
Viewing distance	65 cm	60 cm	70 cm
Method	SS	SS	SS
Subjects	5	15	20 per test
Viewing time	10 sec	12 sec	10 sec
Interval	5 sec	3 sec	3 sec

B. Processing of Eye Tracking Data

1) Eye Tracking Data: The eye tracking data derived from the experiments were those fixation points. A fixation point is represented by a location (coordinates) together with a gaze duration. We analyzed those duration data and found the mean value is: $\mu = 409.8029$ ms. A histogram of all fixation durations from our experiments is illustrated in Fig.2. A parametric fitting was conducted and we found the following Gamma distribution depicts the results well.

$$f(x) = c \cdot \frac{(kx)^{(\alpha-1)}\lambda^{\alpha}e^{(-\lambda kx)}}{\Gamma(\alpha)}, x > 0 \qquad (1)$$

with $\alpha = 2.681, \beta = 4.036, c = 0.001747, k = 0.001821$. The fitted curve is also illustrated in Fig.2 and we can see that the fixation duration obeys the gamma distribution approximatively.

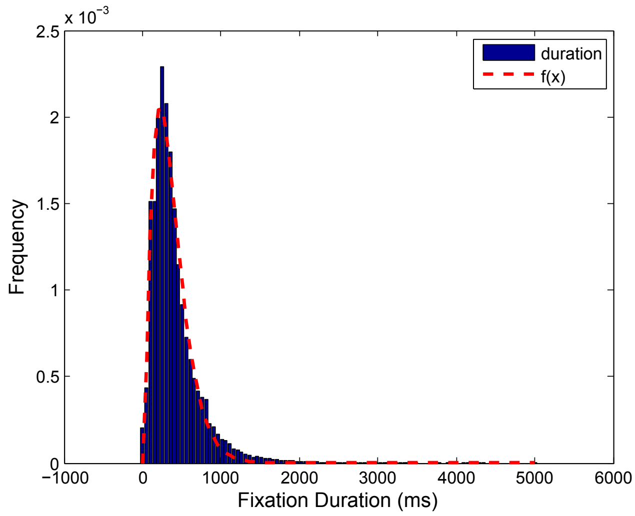

Fig. 2. Histogram of fixation duration with parametric fitting.

2) Saliency Map: Using location and duration data of the fixations, we can generate the saliency maps. Some attempts have been made to compute saliency maps from the fixation data. Liu and Heynderickx [9] overlaid 2D Gaussian masks of the same shape centred at each fixation location. We also

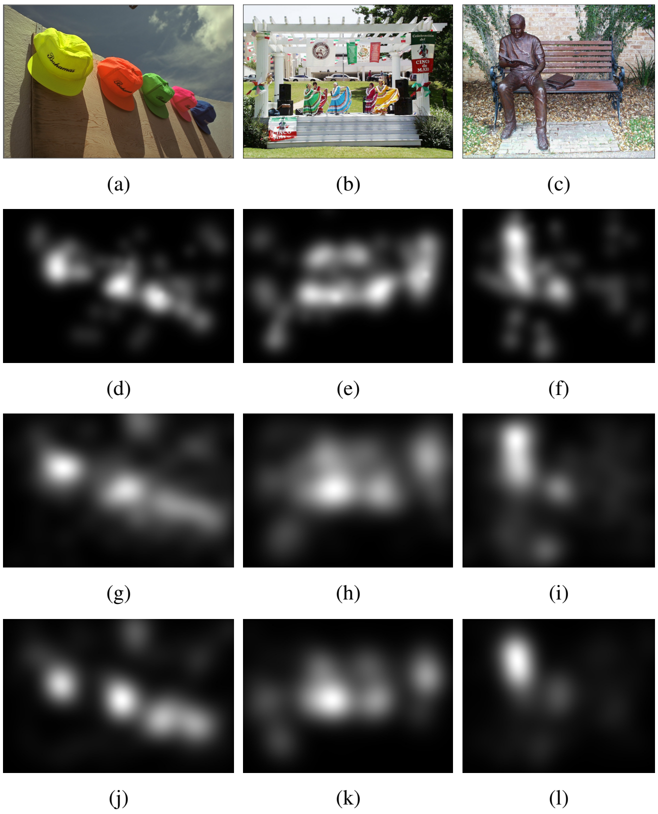

Fig. 3. Illustration of some saliency maps. (a)(b)(c)source images from LIVE database, (d)(e)(f)saliency maps from this work, (g)(h)(i)saliency maps from Liu and Heynderickx [9], (j)(k)(l)saliency maps from Engelke *et al.* [10].

used 2D Gaussian masks, but with the consideration of fixation duration. Longer fixation duration generates a wider mask. The final saliency map is computed as:

$$SM_i(k,l) = \sum_{j=1}^{T} \exp\left[-\frac{(x_i-k)^2+(y_i-l)^2}{[c\cdot\log(d_j)]^2}\right] \quad (2)$$

where (k,l) is the coordinates of the saliency map for image I_i. $k \in [1,M]$; $l \in [1,N]$ and $M \times N$ is the size of stimuli image I_i. $SM_i(k,l)$ is the value of the saliency map at (k,l). T is the number of all fixations from all subjects. The position of the jth fixation point is determined by (x_j,y_j). In Eq.(2), standard deviation of the Gaussian mask is computed as a logarithm function of the fixation duration:

$$\sigma_j = c\cdot\log(d_j) \quad (3)$$

In Eq.(3), d_j is the fixation duration of jth fixation point, and c is a contrast number. It is easy to imagine that we use a greater gaussian patch for a fixation that had a longer fixation duration. In this paper, we set c as 6 for all images except for those from the LIVE Multiply Distortion database [5]. Considering the fact that images from LIVE Multiply Distortion database generally have greater image size, we set $c = 10$. In Liu and Heynderickx's work [9], the same Gaussian mask was used despite of the different fixation duration. A. NINASSI *et al.*'s work [11] took fixation duration into account, but as affected by the DSIS testing approach, performance gain was not substantial. An illustration of saliency maps from this work and those in [9] and [10] is given in Fig.3.

III. The Validation of Eye Tracking Data and Saliency Maps

A. *Saliency-weighted Image Quality Metrics*

To verify the effectiveness of the eye tracking data, we applied saliency maps into two widely used full reference (FR) image quality assessment metrics: peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) index [8]. We generated the distortion maps first, then pooled the distortions with weights given by the saliency maps. The saliency map weighted quality metric is defined as:

$$SM\text{-}Q = \frac{\sum_{x=1}^{M}\sum_{y=1}^{N} SM_i(x,y)\cdot DM_i(x,y)}{\sum_{x=1}^{M}\sum_{y=1}^{N} SM_i(x,y)} \quad (4)$$

where $SM\text{-}Q$ is the saliency map weighted quality value, M × N is the size of the image. SM_i is the saliency map that gives the weight, given by Eq.(2). DM_i is the distortion map between the distorted image D_i and the source image I_i. (x,y) are the spatial coordinates. And in Eq.(4), different DM_i will give different $SM\text{-}Q$. If DM_i is the distortion map calculated by the SSIM metric, then $SM\text{-}Q$ is saliency based SSIM (SM-SSIM). If DM_i is given by square error, then $SM\text{-}Q$ is saliency based mean square error (SM-MSE), and we can calculate saliency based PSNR (SM-PSNR) from SM-MSE easily. An illustration of saliency based distortion maps is given in Fig.4.

B. *Performance of Saliency-weighted Quality Metrics*

We conducted the test of saliency weighted image quality metrics on seven IQA image databases [1]–[7]. A widely used performance metric: Spearman Rank-Order Correlation Coefficient (SROCC) was employed. Performance of four metrics (PSNR, SM-PSNR, SSIM, SM-SSIM) in terms of SROCC was listed in Table III. In this table, 'SM-P' is short for SM-PSNR, 'SM-S' denotes SM-SSIM, and 'LIVEm' indicates LIVE Multiply Distortion. From Table III, we can see that a substantial performance gain can be achieved when applying the visual attention data into IQA metrics.

Liu and Heynderickx provided saliency maps for the LIVE database, and Engelke *et al.* provided saliency maps for LIVE, Toyama and IVC database. A comparison of IQA metrics' performance using different saliency maps is also given in Table IV. In Table IV, 'w/o' means without using saliency maps and meanings of other fields are self-evident. From Table IV, we can see that the saliency maps in this work lead to higher performance gain as compared to existing data.

In this paper, the eye movement data was the average of only 5 participants. During our test, it was noticed that superimposing too many subjects' fixation data always result in oversmoothed saliency maps and therefore, lead to very limited performance gain for the IQA metrics. Since there are

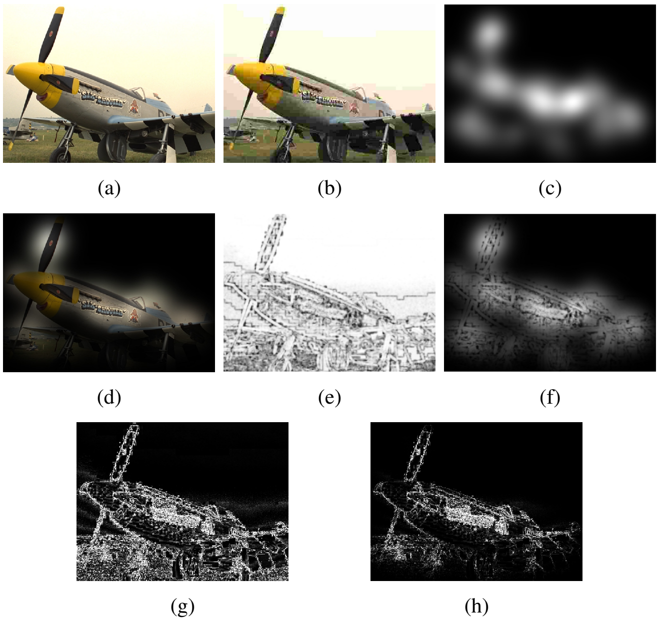

Fig. 4. Illustration of saliency based distortion maps. (a)source image, (b)JPEG compressed image, (c)saliency map, (d)saliency weighted image, (e)distortion map of SSIM for (b)and(a), (f)saliency weighted SSIM distortion map, (g)distortion map of squared error for (b)and(a), (h)saliency weighted squared error distortion map.

TABLE III
PERFORMANCE OF IQA METRICS IN TERMS OF SROCC

Metric	PSNR	SM-P	Gain	SSIM	SM-S	Gain
LIVE	0.8755	0.8946	1.91%	0.9100	0.9352	2.52%
TID	0.5529	0.5601	0.72%	0.6256	0.6974	7.18%
CSIQ	0.8057	0.8192	1.35%	0.8368	0.8762	3.94%
Toyama	0.6130	0.7182	10.52%	0.7865	0.8585	7.20%
LIVEm	0.6771	0.7583	8.12%	0.6455	0.7598	11.43%
IVC	0.6887	0.7304	4.17%	0.7785	0.8030	2.45%
A57	0.6189	0.6297	1.08%	0.4072	0.4979	9.07%

inevitable noisy measures in the eye tracking data, averaging over a too large number of fixation data tend to mix the 'accurate' measures with those 'noisy' measures and therefore reduce the overall credibility. As a consequence, although dozens of participants were involved in the test, we chose to average from only five of the personnels we deemed the most responsible. And as shown in this section, the final eye-tracking data revealed in this work seems to be more valid than previous data collected from larger number of viewers.

IV. Conclusion

It is a trend for image quality research to take into account of visual attention data. Despite of the large number of image quality databases, there is no comprehensive subjective visual attention data available. To fill this void, in this paper we

TABLE IV
PERFORMANCE OF IQA METRICS OF DIFFERENT SALIENCY MAPS IN TERMS OF SROCC

Saliency Maps		w/o	ours	Engelke [10]	Liu [9]
LIVE	PSNR	0.8755	**0.8946**	0.8893	0.8911
	SSIM	0.9100	**0.9352**	0.9328	0.9299
Toyama	PSNR	0.6130	**0.7182**	0.6960	-
	SSIM	0.7865	**0.8585**	0.8270	-
IVC	PSNR	0.6887	**0.7304**	0.7291	-
	SSIM	0.7785	**0.8030**	0.7756	-

conducted eye tracking experiments for the reference images from seven widely used IQA databases (LIVE, TID2008, CSIQ, Toyama, LIVE Multiply Distortion, IVC, A57). We also proposed a fixation duration adaptive approach to generate saliency map from the eye tracking data. Extensive experimental results were provided to verify the effectiveness of the proposed data and approach. Both the eye tracking data and the saliency maps will be made available to the community at `gvsp.sjtu.edu.cn`.

Acknowledgment

This work was supported in part by NSFC (61221001,61371146).

References

[1] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, "Live image quality assessment database release 2," [Online], available: http://live.ece.utexas.edu/research/quality.
[2] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti, "Tid2008-a database for evaluation of full-reference visual quality assessment metrics," *Advances of Modern Radioelectronics*, vol. 10, no. 4, pp. 30–45, 2009.
[3] E. C. Larson and D. M. Chandler, "Categorical image quality (csiq) database 2009," [Online], available: http://vision.okstate.edu/csiq.
[4] Y. Horita, K. Shibata, Y. Kawayoke, and Z. M. P. Sazzad, "Mict image quality evaluation database 2000," [Online], available: http://mict.eng.u-toyama.ac.jp/mict/index2.html.
[5] D. Jayaraman, A. Mittal, A. K. Moorthy, and A. C. Bovik, "Objective quality assessment of multiply distorted images," in *Signals, Systems and Computers (ASILOMAR), 2012 Conference Record of the Forty Sixth Asilomar Conference on*. IEEE, 2012, pp. 1693–1697.
[6] A. Ninassi, P. L. Callet, and F. Autrusseau, "Subjective quality assessment-ivc database 2005," [Online], available: http://www2.irccyn.ecnantes.fr/ivcdb.
[7] D. M. Chandler and S. S. Hemami, "A57 database 2007," [Online], available: http://foulard.ece.cornell.edu/dmc27/vsnr/vsnr.html.
[8] W. Zhou, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *Image Processing, IEEE Transactions on*, vol. 13, no. 4, pp. 600–612, 2004.
[9] H. Liu and I. Heynderickx, "Visual attention in objective image quality assessment: based on eye-tracking data," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 21, no. 7, pp. 971–982, 2011.
[10] U. Engelke, A. Maeder, and H.-J. Zepernick, "Visual attention modelling for subjective image quality databases," in *Multimedia Signal Processing, 2009. MMSP'09. IEEE International Workshop on*. IEEE, 2009, pp. 1–6.
[11] A. Ninassi, O. Le Meur, P. Le Callet, and D. Barbba, "Does where you gaze on an image affect your perception of quality? applying visual attention to image quality metric," in *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, vol. 2. IEEE, 2007, pp. II–169.