

SUBJECTIVE ASSESSMENT OF IMAGE QUALITY INDUCED SALIENCY VARIATION

Lucie Lévêque¹, Wei Zhang², and Hantao Liu³

¹Department of Computer Science and Software Engineering, Xi'an Jiaotong-Liverpool University, Suzhou, China

²School of Telecommunications Engineering, Xidian University, Xi'an, China

³School of Computer Science and Informatics, Cardiff University, Cardiff, United Kingdom

ABSTRACT

Our previous study has shown that image distortions cause saliency distraction, and that visual saliency of a distorted image differs from that of its distortion-free reference. Being able to measure such distortion-induced saliency variation (DSV) significantly benefits algorithms for automated image quality assessment. Methods of quantifying DSV, however, remain unexplored due to the lack of a benchmark. In this paper, we build a benchmark for the measurement of DSV through a subjective study. Sixteen experts in computer vision were asked to compare saliency maps of distorted images to the corresponding saliency maps of the original images. All saliency maps were rendered from ground truth human fixations. A statistical analysis is performed to reveal the behaviours and properties of human assessment of the saliency variation. The benchmark is made publicly available to the research community.

Index Terms— Image quality, distortion, eye-tracking, saliency, visual attention

1. INTRODUCTION

Visual saliency represents an important mechanism in the human visual system (HVS) that allows effective selection of the most relevant information in a visual scene [1]. Saliency has been widely studied in relation to image quality assessment [2]-[4]. Knowing where people look in images is significantly beneficial for improving the reliability of image quality assessment (IQA) algorithms as well as their associated applications [5]-[7]. However, fundamental challenges to fully simulating saliency in IQA remain. This is mainly due to the fact that our knowledge about how saliency plays a role in image quality assessment and how to express that perception in an efficient mathematical model is rather limited.

In our previous research [7], we carried out an eye-tracking study in order to better understand how image distortions and saliency concurrently affect the human visual system. The study revealed that given a distortion-free image and its “ground truth” saliency obtained from eye-tracking,

the occurrence of distortions in the image alters viewers’ gaze behaviour, yielding changes to the original saliency. As an observation also revealed from this study, the saliency variation is strongly related to the change in image quality. As illustrated in Fig. 1, if an image is of high quality, then most viewers will concentrate their gaze around the salient objects (see Fig. 1(a)), whereas if the image quality is degraded, viewers’ gaze will deviate from the objects of interest (see Fig. 1(b) and (c)). Hence, the saliency maps of the distorted images tend to have a different pattern than that of their original images, e.g., Fig.1 (a) shows a concentrated saliency map, while Fig.1(b) and (c) present more dispersed saliency maps. Also, it can be seen from Fig. 1 that the lower the image quality, viewers’ gaze is more likely to be distracted by distortions, and the wider the variation in the saliency map of the distorted image relative to that of its corresponding higher quality content. For example, the saliency map of Fig. 1(c) is more spread throughout the scene than the saliency map of Fig. 1(b). Further to this research, a follow-up study [8] has been undertaken to investigate the relationship between the distortion-induced saliency variation (DSV) and the perceived image quality. In that study, DSV was simply measured (i.e., by use of off-the-shelf metrics) and used to approximate image quality. The result demonstrated that DSV can potentially form a good basis for image quality prediction. However, challenges to reliably measuring DSV remain.

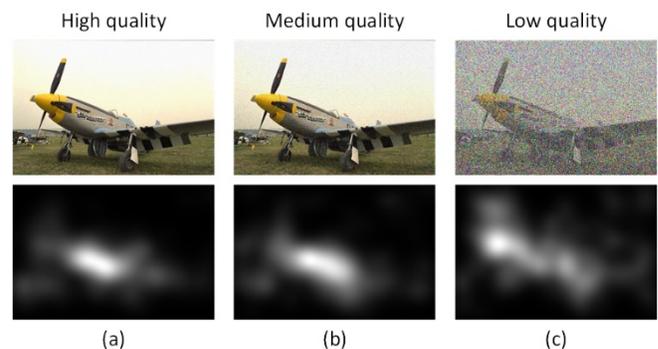


Fig. 1. Illustration of the distortion induced saliency variation.

To make better use of saliency in IQA algorithms, a sophisticated DSV measure is needed. However, there is a paucity of literature regarding DSV, in particular there is no appropriate benchmark that has been universally adopted. In this paper, we approach the measurement of DSV via a subjective study, where sixteen experts in computer vision evaluated saliency maps of distorted images in terms of how they were different from the saliency maps of the original images, and to what extent. We also present the statistics and properties of the benchmark.

2. MEASUREMENT OF DISTORTION-INDUCED SALIENCY VARIATION

2.1. Stimuli

The study used the SIQ288 database [9], which contains 288 images of varying quality (including 18 distortion-free originals distorted with different types of artifacts at various levels of degradation), and their corresponding saliency maps obtained from eye-tracking. This database involved a large number of participants (i.e., 160 observers) in order to collect unbiased eye-tracking data so that each viewer did not have to view multiple variations (i.e., distorted images) of the same scene. The distorted images were systematically selected from the LIVE database [10]-[11], which consists of five distortion types including white noise (WN), JPEG compression (JPEG), Gaussian blur (GBLUR), JPEG2000 compression (JP2K), and simulated fast-fading in wireless channels (FF). For each distortion type, three distorted versions per original scene were selected, reflecting three distinct levels of perceived quality, i.e., “High” (i.e., with perceptible but not annoying artifacts), “Medium” (i.e., with noticeable and annoying artifacts) and “Low” (i.e., with very annoying artifacts). Fig. 2 illustrates an original image and its saliency map contained in the SIQ288 database. Fig. 3 illustrates the saliency maps of all distorted images that are originated from the same scene as shown in Fig. 2.



Fig. 2. Illustration of an original image and its saliency map contained in the SIQ288 database.

2.2. Experimental procedure

A controlled experiment was designed to measure the distortion-induced saliency variation (DSV). Subjects were requested to compare and score the similarity between the saliency map of a distorted image and that of the corresponding original scene. As the rating interface shown

in Fig. 4, two stimuli, i.e., the reference saliency map (on the basis of the original) at the left-hand side and the test saliency map (on the basis of the distorted image) at the right-hand side were simultaneously displayed on the same screen. Note that the saliency maps of the original scenes were also evaluated in the same experimental session as the test stimuli as a hidden reference. The scoring scale ranged from 0 to 100 and included additional semantic labels (i.e., “Bad”, “Poor”, “Fair”, “Good” and “Excellent”) at intermediate points. Subjects were asked, by moving the slider on the scoring scale, to express their opinions on to what extent the test saliency map is similar to the reference saliency map. The stimuli were displayed on a Dell 19” liquid-crystal display with a native resolution of 1920×1080 pixels. The experiment was conducted in a standard office environment [12] and the viewing distance was approximately 60cm.

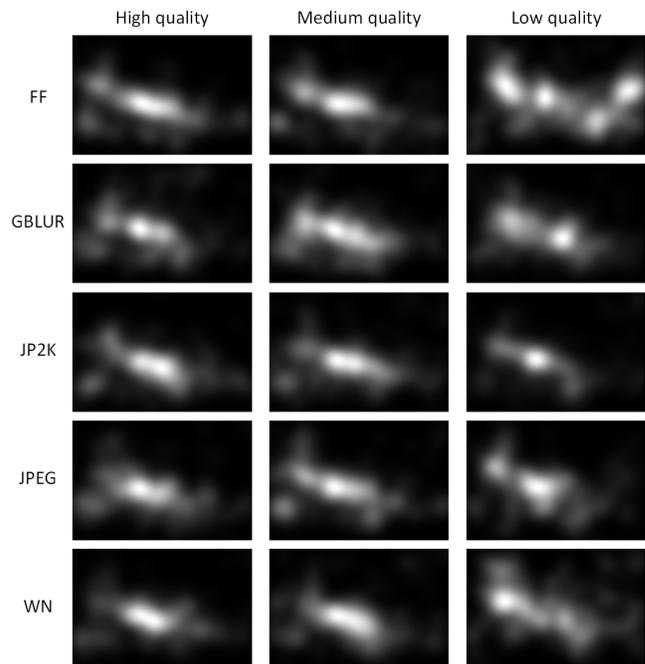


Fig. 3. Illustration of the saliency maps of all distorted images that are originated from the same scene (see Fig. 2) contained in the SIQ288 database.

The participants of the study were recruited from the Visual Computing Research Group at the School of Computer Science and Informatics, Cardiff University. The sixteen subjects, being nine senior PhD students and seven staff members, were considered experts in computer vision. Each participant was briefed on the objective and procedure of the experiment and provided with a training session to familiarise themselves with the images and saliency maps used and with how to use the scoring interface. After training, the test stimuli were shown in a different random order to each participant in a separate session. There was no time limit for the participants to make the assessment.

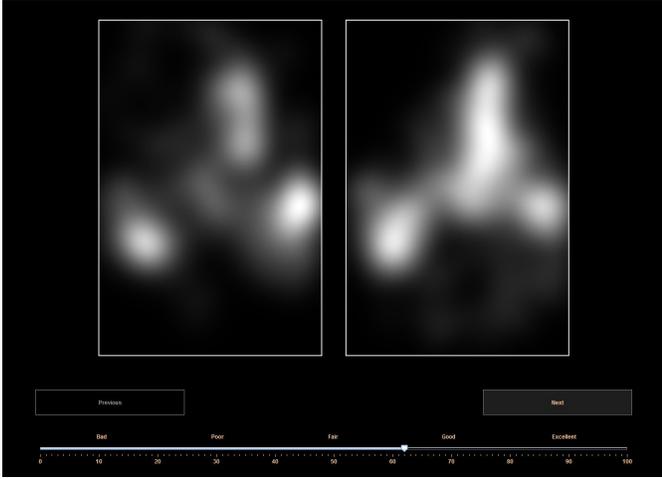


Fig. 4. Illustration of the scoring interface.

3. EXPERIMENTAL RESULTS

3.1. Processing of raw data

First, the subjective scores were transformed to difference scores in order to discount likely subject preferences for certain saliency patterns:

$$d_{ij} = s_{ij_ref} - s_{ij} \quad (1)$$

where s_{ij} denotes the raw score given by the i -th subject to the j -th test saliency map (from a distorted image) and s_{ij_ref} denotes the raw score given by the i -th subject to the saliency map of the hidden reference corresponding to the j -th saliency map. Then the difference scores were converted into z-scores:

$$z_{ij} = (d_{ij} - \mu_i) / \sigma_i \quad (2)$$

where μ_i is the mean of all difference scores for the subject i , and σ_i is the standard deviation. Note z-scores are calculated to account for the differences between subjects in the use of the scoring scale and calibrate them towards the same mean and standard deviation. The z-scores were linearly mapped to [0, 100]. Finally, the difference mean saliency variation score (DMSS) of each stimulus was computed as the mean of the rescaled z-scores (z'_{ij}) over all subjects:

$$DMSS_j = \frac{1}{s} \sum_{i=1}^s z'_{ij} \quad (3)$$

where s is the number of subjects. The resulting scores of DMSS serve as the benchmark for the measurement of distortion-induced saliency variation.

3.2. Analysis of behaviours of human assessment

We conducted a statistical analysis to reveal the behaviours and properties of human assessment of the saliency variation. An ANOVA (Analysis of Variance) was performed based on the matrix of $\{z_{ij}'\}$, where $i=\{1,2,\dots,16\}$ and $j=\{1,2,\dots,270\}$. Each difference saliency variation measure z_{ij}' is associated with a set of labels {"reference image scene", "distortion type", "distortion level", "subject"}. In the ANOVA, z_{ij}' was selected as the dependent variable, the "image scene", "distortion type", and "distortion level" as fixed independent variables and the "subject" as random independent variable. The two-way interactions of "distortion type", and "distortion level" are included. The results of the ANOVA analysis are summarised in Table I, including the F-statistic (F-value), the degrees of freedom (df) and the significance (p-value). The results show that "image scene", "distortion type" and "distortion level" have a significant effect (i.e., $p < 0.05$) on the measure of saliency variation. Overall, there is no statistically significant difference between subjects in scoring the saliency variation (i.e., $p > 0.05$ for "subject"). This means the experts performed consistently in the experiments. The interaction between "distortion type" and "distortion level" is significant (i.e., $p < 0.05$), which implies that the difference in measurement between the three levels of distortion is not the same for the five distortion types.

TABLE I. RESULTS OF THE ANOVA ANALYSIS

Source	df	F	Sig. (p)
Image scene	17	58.298	7.2E-179
Distortion type	4	7.078	0.00001
Distortion level	2	212.382	1.2E-88
Subject	15	0.0001	1
Distortion type *	8	10.274	2.5E-14
Distortion level			

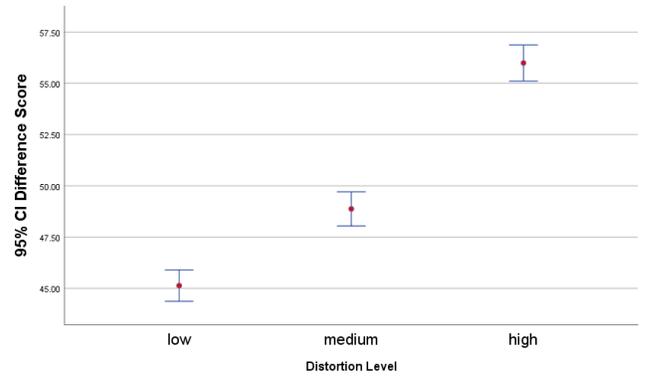


Fig. 5. Illustration of the difference saliency variation score averaged over relevant stimuli for the low, medium and high distortion levels. Error bars indicate a 95% confidence interval.

Now, we focus on some important properties of the statistics. Fig. 5 illustrates the difference saliency variation score averaged over all relevant stimuli for the low, medium and high distortion levels. Pairwise comparisons were performed using an independent samples t -test. The t -test results show that the difference between low and medium levels (i.e., $F=10.962$, $p<0.05$) is significant, and between medium and high levels (i.e., $F=4.475$, $p<0.05$) is also significant. This clearly shows that the wider the saliency variation relative to the reference, the higher the distortion in the image and the lower the image quality. This also implies that the distortion-induced saliency variation is reliably quantified by the subjective assessment.

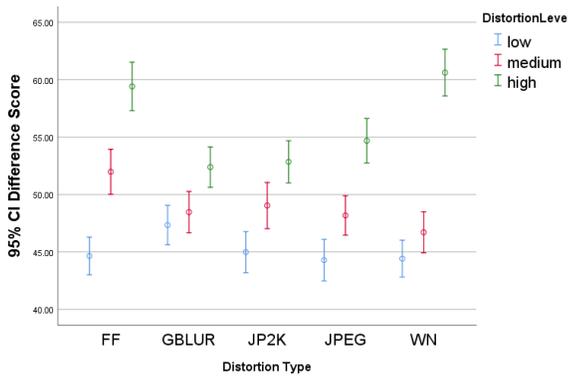


Fig. 6. Illustration of the difference saliency variation score averaged over relevant stimuli for the low, medium and high distortion levels of each distortion type. Error bars indicate a 95% confidence interval.

The impact of the distortion type is illustrated in Fig. 6. It can be seen from the figure that there is a linear relationship between the measure of saliency variation and the strength of distortion for FF, JP2K and JPEG. For GBLUR and WN, the increase of distortion from the low to medium level does not cause a significant saliency variation; and only high distortion leads to significantly noticeable changes to saliency relative to the reference. This tendency is also clearly reflected in Fig. 3 when visually comparing the saliency maps. The difference of the impact may be due to that GBLUR and WN distortions appear as an evenly distributed pattern (see effects of WN in Fig. 1), gaze is focused on salient objects when the distortion is mild (i.e., low to medium distortion), and gaze is significantly affected when distortion is high and begins to reduce the visibility of the salient objects. The FF, JP2K and JPEG are localised distortions so that gaze is more sensitive to the occurrence of locally pop-out distortions, e.g., a mild localised distortion in the background may distract gaze from the focus of salient objects, and consequently cause an obvious saliency variation.

4. CONCLUSIONS

This paper considered the development of a benchmark for the measurement of distortion-induced saliency variation, via a subjective study with expert assessors. The difference between the saliency map of a distorted image and the saliency map of its corresponding distortion-free reference has been reliably quantified. Going forward, we intend to develop algorithms that can automatically measure saliency variation; and use these algorithms to aid in image quality assessment and relevant applications.

5. REFERENCES

- [1] A. B. Watson, *Digital Images and Human Vision*. Cambridge, MA, USA: MIT Press, 1997.
- [2] A. Ninassi, O. Le Meur, P. Le Callet, and D. Barba, "Does where you gaze on an image affect your perception of quality? Applying visual attention to image quality metric", *IEEE International Conference on Image Processing*, pp. 169-172, 2007.
- [3] H. Liu and I. Heynderickx, "Visual Attention in Objective Image Quality Assessment: Based on Eye-Tracking Data," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 7, pp. 971-982, July 2011.
- [4] F. Röhrebein, P. Goddard, M. Schneider, G. James, and K. Guo, "How does image noise affect actual and predicted human gaze allocation in assessing image quality?" *Vision research*, vol. 112, pp. 11–25, 2015.
- [5] U. Engelke, H. Kaprykowsky, H. Zepernick, and P. Ndjiki-Nya, "Visual attention in quality assessment," *IEEE Signal Process. Mag.*, vol. 28, pp. 50–59, Nov. 2011.
- [6] L. Zhang, Y. Shen, and H. Li, "VSI: A visual saliency-induced index for perceptual image quality assessment", *IEEE Transactions on Image Processing*, vol. 23, pp. 4270-4281, 2014.
- [7] W. Zhang, A. Borji, Z. Wang, P. Le Callet and H. Liu, "The Application of Visual Saliency Models in Objective Image Quality Assessment: A Statistical Evaluation," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 6, pp. 1266-1278, June 2016.
- [8] W. Zhang and H. Liu, "Toward a Reliable Collection of Eye-Tracking Data for Image Quality Re-search: Challenges, Solutions, and Applications," in *IEEE Transactions on Image Processing*, vol. 26, no. 5, pp. 2424-2437, May 2017.
- [9] W. Zhang and H. Liu, "SIQ288: A saliency dataset for image quality research," *IEEE 18th International Workshop on Multimedia Signal Processing (MMSp)*, Montreal, QC, 2016, pp. 1-6.
- [10] H. R. Sheikh, Z. Wang, L. Cormack, and A. Bovik. *LIVE Image Quality Assessment Database Release 2*. [Online]. Available: <http://live.ece.utexas.edu/research/quality>

[11] H. Sheikh, M. Sabir, and A. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," IEEE transactions on Image Processing, vol.15, pp. 3440-3451, 2006.

[12] Methodology for the Subjective Assessment of the Quality of Television Pictures, Standard BT.500-11, International Telecommunication Union, Geneva, Switzerland, 2002, pp. 53-56.