TOWARDS AN EFFICIENT MODEL OF VISUAL SALIENCY FOR OBJECTIVE IMAGE QUALITY ASSESSMENT

Hantao Liu [1] *and Ingrid Heynderickx* [1, 2]

[1] Department of Mediamatics, Delft University of Technology, Delft, The Netherlands
[2] Group Visual Experiences, Philips Research Laboratories, Eindhoven, The Netherlands

ABSTRACT

Based on "ground truth" eye-tracking data, earlier research [1] shows that adding natural scene saliency (NSS) can improve an objective metric's performance in predicting perceived image quality. To include NSS in a real-world implementation of an objective metric, a computational model instead of eye-tracking data is needed. Existing models of visual saliency are generally designed for a specific domain, and so, not applicable to image quality prediction. In this paper, we propose an efficient model for NSS, inspired by findings from our eye-tracking studies. Experimental results show that the proposed model sufficiently captures the saliency of the eye-tracking data, and applying the model to objective image quality metrics enhances their performance in the same manner as when including eye-tracking data. [1]

Index Terms— Visual attention, eye-tracking, image quality assessment, objective metric, human visual system

1. INTRODUCTION

Over the last several decades, we have witnessed remarkable progress in the development of objective metrics for the automatic prediction of perceived image quality. Novel research tends to further improve the reliability of objective metrics by taking into account visual attention of the human visual system (HVS). Modeling this aspect in an objective metric is not a trivial task mainly due to the fact that the mechanism of human attention when assessing image quality is not fully understood yet.

To understand the basic added value of including visual attention in the design of objective metrics, "ground truth" data obtained from eye-tracking experiments are used [1]-[3]. It is demonstrated in [1] that adding natural scene saliency (NSS) is beneficial to image quality prediction in general terms. However, for a real-world implementation the eye-tracking data need to be substituted by a computational model of visual saliency.

A variety of models of visual saliency are available in literature, most of which are based on a bottom-up (scene-dependent) framework (e.g. [4]-[7]). The most well-known model is the one proposed in [4], in which multi-scale image features, including intensity, color and orientation are combined into a single saliency map. The model in [5] is derived from the analysis of the statistics of image features at observers' gaze, the model in [6] relies on the theory of maximizing information in a visual scene, and the model in [7] is based on combining current models of the HVS behavior. Directly applying these saliency models in the design of objective image quality metrics, however, is of a practical concern. First, the existing models are generally designed for a specific domain, and therefore, not necessarily applicable to image quality prediction. Second, they are intrinsically computationally expensive, which limits their use in real-time. Third, the accuracy of these saliency models in improving image quality prediction is not yet completely proved. Hence, it is highly desirable to develop a model for visual saliency, which is computationally efficient for real-time application, but still sufficiently reliable for image quality assessment.

To develop such a saliency model, we rely on the multi-scale approach taken in [4], but extend the idea by first refining the calculation of local contrast to achieve a more meaningful "object contrast" in the visual scene. In addition, our proposed model embeds two significant findings from our eye-tracking studies to further improve its efficiency and reliability.

2. PROPOSED SALIENCY MODEL

2.1. Findings from eye-tracking data

Since most of the existing objective metrics are based on the luminance signal of an image only, we investigate whether also saliency can be modeled only with the luminance component without significantly compromising its accuracy. Results reported in [8] show that human visual attention for quality assessment is indeed insensitive to color. Hence, this

[1] The implementation of the model of visual saliency is available on the web-site: http://mmi.tudelft.nl/iqlab/index.html

observation can be reasonably used to simplify the modeling of saliency, and the resulting saliency model can be directly implemented in objective metrics based on luminance only. Another important finding is that by integrating eye-tracking data of NSS to state-of-the-art objective metrics [1], the performance gain is non-existing for images without convergent salient features (i.e. a clear region of interest). This implies that the overall efficiency can be further improved by adaptively using the saliency model only for images with a clear region of interest. Both aspects are explicitly taken into account in the design of our model.

2.2. Generation of the saliency map

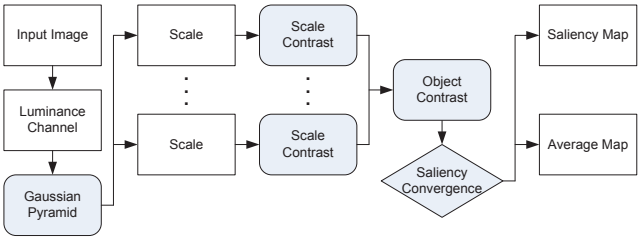


Fig. 1. Schematic overview of the proposed model of visual saliency for objective image quality assessment.

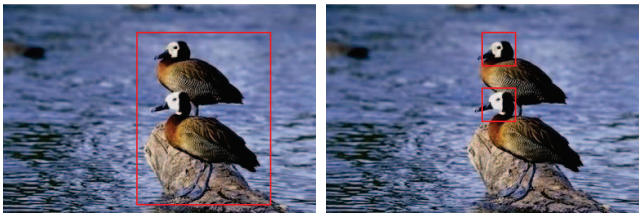


Fig. 2. Illustration of the principle of modeling saliency as object contrast: (a) the object contrast perceived as saliency in a global way, (b) the object contrast perceived as saliency in a local way.

The schematic overview of our proposed model is given in Figure 1. It only uses the luminance component of the image material as input. The model is based on the principle that contrast is a dominant factor in human visual perception; an object that makes itself distinct (i.e. with high contrast) from its vicinity is most likely to attract human attention. Such "object contrast" can be perceived both on a global and local scale. As illustrated in Figure 2, the two birds and the rock as a whole stand out from the surrounding water; locally the white face of each bird pops out from its neighborhood. To simulate this perception, a multi-scale approach is employed for estimating saliency. Unlike the approach in [4], which generates saliency by calculating the pixel-by-pixel center-surround differences between scales, we construct a saliency map by extracting contrast at each individual scale (hereafter referred to as "scale contrast"). As such, each scale contrast map represents the image features standing out at a specific scale, i.e. detailed features stand out at a fine scale, while macro features pop out at a coarse scale. Combining the scale contrast maps over all scales yields a perceptually more meaningful "object contrast" map, indicating the saliency of the visual scene. The spatial distribution of saliency is then examined, and only when the map contains convergent salient regions, it is retained.

2.2.1. Scale contrast map

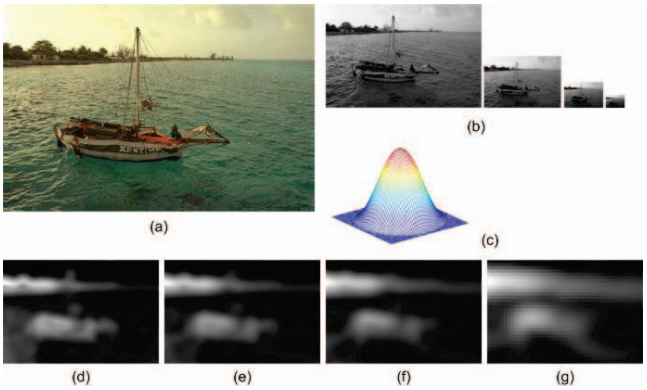


Fig. 3. An example of calculating the scale contrast (SC) in the multi-scale image space: (a) original image, (b) Gaussian pyramids with scale levels σ=1, 2, 3, 4, (c) illustration of the circular raised cosine weighting function, (d)-(g) the scale contrast maps calculated at the four scales of (b), respectively [(e), (f) and (g) are resized by interpolation to the scale level σ=1].

To calculate the scale contrast map, let's denote the luminance channel of an image of M×N (height × width) pixels as I (i, j) for i∈[1, M], j∈[1, N]. The luminance channel is then progressively low-pass filtered and sub-sampled using dyadic Gaussian pyramids, with σ=[0,…,8] yielding 9 spatial scales [9]. The resolution of the image at scale level σ is $1/2^{\sigma}$ times the original image resolution. To reduce the computational power, while sufficiently representing the fine and coarse scales, we use four scale levels (i.e. σ=[1, 2, 3, 4]) in our model. The scale contrast (SC) is defined within an image patch superposed on each pixel location (i, j) as:

$$SC(i,j) = \sqrt{\sum \omega_p \left(\frac{I_p - \bar{I}_p}{\bar{I}_p}\right)^2}, \quad p \in [1, K] \tag{1}$$

where $\bar{I}_p$ is the mean luminance of the local image patch

$$\bar{I}_p = \sum \omega_p I_p \tag{2}$$

I_p is the pixel intensity at location p, and K is the total number of pixels in the image patch. For the weighting factor ω_p a circular raised cosine weighting function [5] as illustrated in Figure 3(c) is adopted. Note that this function can be replaced by an alternative function without expected

change in performance. The size of the image patch is adapted to the scale size, taking into account the highlighted features in a specific scale (i.e. in our experiments, {1/5, 1/4, 1/3, 1/2}×min(1/2$^{\sigma}$×[M, N]) is used for the scale levels of σ=1, 2, 3, 4, respectively). The entire procedure is shown as an example in Figure 3, where the SC map clearly gives prominence to the outstanding features at the scale it is calculated for, e.g. the boom and ropes of the sailing boat are highlighted at the scale σ=1, while the body of the boat is highlighted at the scale σ=4

2.2.2. Selecting saliency

After calculating the SC maps at the different scales, they are linearly combined into a single conspicuity map, as illustrated in Figure 4(b). The combined map reveals a more comprehensive "object contrast" map, including features standing out in both coarse and fine scales. As discovered in [1], the spatial distribution of saliency affects its actual added value in image quality prediction. To include this aspect, we propose a simple method to automatically detect images with bit spread in saliency. To do so, the conspicuity map is divided into blocks of equal sizes (i.e. block size is 1/20 of the map size as used in our experiments). A block that contains a pixel with its intensity above a certain threshold (i.e. 40% of the maximal intensity is used in our experiments) is considered as "covered". If all blocks are "covered", the conspicuity map is considered not converged, and therefore, is not used in the subsequent image quality prediction. More complex algorithms may be designed, but are outside the scope of this paper. In case the conspicuity map is retained, it is added to a center bias model to generate the saliency map (as already suggested in [10]). The whole procedure is shown in Figure 4.

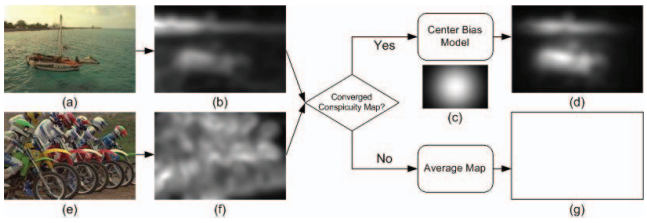


Fig. 4. An example of saliency generation: (a) original image, (b) conspicuity map of (a), (c) a model of center bias, (d) saliency map of (a), (e) original image without convergent salient features, (f) conspicuity map of (e), (g) average (i.e. no saliency) map of (e).

3. EXPERIMENTAL RESULTS

To validate the proposed model of visual saliency, we first evaluate how well the model can predict our NSS data of [1]. Then, we verify the added value of the model in image quality prediction. The data in [1] contain human saliency maps (HSM), obtained from 20 observers looking freely to 29 source images of the LIVE database [11]. The correlation coefficient (i.e. ρ in the range [-1, 1]) between the subjectively measured HSM and the computational saliency map (CSM) is calculated. To make a fair comparison, the analysis of saliency convergence is not included here (i.e. the conspicuity map is always considered converged). Figure 5 illustrates the ρ values of our proposed model for the 29 images (the content and ordering of the images can be found in [11]). The averaged ρ value is **0.6**. We visualize some examples of saliency maps in Figure 6; i.e. for the images with a ρ value around 0.6 and for the images with the highest and lowest ρ values. As can be seen for a ρ value of 0.6, the model captures most of the saliency measured with eye-tracking data. Of course, there is still room to further improve the correspondence between the modeled and measured saliency, but most probably at the expense of its computational cost. For the purpose we have in mind here, however, it is more important to evaluate whether a ρ value of 0.6 is sufficient to replace modeled saliency for measured saliency in objective metrics for quality prediction.

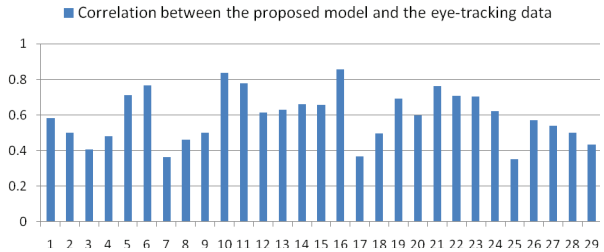


Fig. 5. Correlation coefficient (ρ) between the human saliency map (HSM) and the computational saliency map (CSM) over 29 images (the content and ordering of the images can be found in [11]).

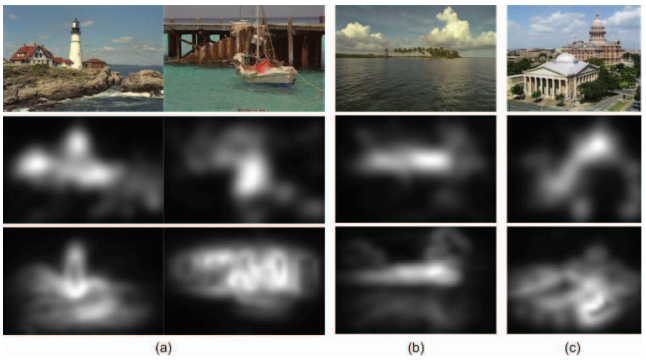
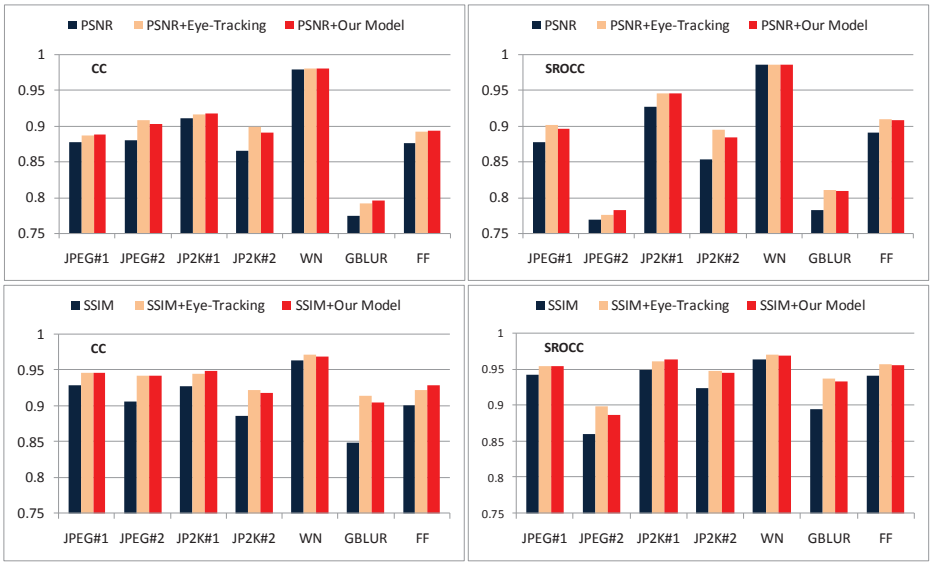


Fig. 6. Examples of images with representative ρ values achieved by our proposed model: (a) images with their ρ values around **0.6** (i.e. the averaged ρ value achieved by our model), (b) image with the highest ρ value, and (c) image with the lowest ρ value.

Thus, to investigate whether our model is sufficiently sound to serve as computational alternative for eye-tracking data, and consequently, can be used for real-time quality assessment, we repeat the experiment as described in [1]. As such, the saliency predicted by our model is added to the



(a)

	CC	SROCC
PSNR	0.88	0.87
PSNR + Eye-tracking	0.90	0.89
PSNR + Our Model	**0.90**	**0.89**
	CC	SROCC
SSIM	0.91	0.92
SSIM + Eye-tracking	0.94	0.95
SSIM + Our Model	**0.94**	**0.94**

(b)

Table 1. Correlation coefficients (i.e. CC, SROCC) of objective metrics calculated for the LIVE database [11].

PSNR and SSIM metrics to assess the image quality of the LIVE database [11]. In our experiments, four images (namely "bikes", "buildings", "paintedhouse", and "stream" in [11]) are detected as not having convergent salient features, and thus, the average saliency map as illustrated in Figure 4 is applied. The metrics' performance is quantified by the Pearson (CC) and Spearman (SROCC) correlation coefficients [1]. As illustrated in Table 1(a), adding our model to PSNR and SSIM consistently improves their performance for different image distortion types. The corresponding averaged correlation for the entire database is listed in Table 1(b). Experimental results show that the amount of gain in performance (i.e. the increase in correlation) achieved by adding our model to an objective metric is similar to what can be obtained by including eye-tracking data.

4. CONCLUSIONS

In this paper, an efficient model of visual saliency for the use in objective image quality assessment is designed. Inspired by findings from eye-tracking studies, we built the saliency model based on the luminance component only, and simply calculated the object contrast over multi-scales. Moreover, the spatial distribution of the resulting saliency was analyzed to decide whether or not to incorporate it in image quality prediction. The proposed model is sufficiently consistent with eye-tracking data, and applying it to state-of-the-art objective metrics indeed enhances their performance to the same extent as adding the "ground truth" saliency. As such, the proposed model is promising in terms of both computational efficiency and practical reliability for real-time image quality assessment.

5. REFERENCES

[1] H. Liu and I. Heynderickx, "Visual Attention in Objective Image Quality Assessment: based on Eye-Tracking Data," *IEEE Trans. CSVT*, vol. 21, pp. 971-982, July, 2011

[2] O. Le Meur, A. Ninassi, P. Le Callet, and D. Barba, "Overt visual attention for free-viewing and quality assessment tasks Impact of the regions of interest on a video quality metric," *Signal Processing: Image Communication*, pp. 547-558, 25(2010).

[3] C. T. Vu, E. C. Larson, and D. M. Chandler, "Visual fixation patterns when judging image quality: Effects of distortion type, amount, and subject experience," in *Proc. IEEE SSIAI*, pp. 73-76, Mar. 2008.

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

[5] U. Rajashekar, A. C. Bovik, and L. K. Cormack, "Gaffe: A gaze-attentive fixation finding engine," *IEEE Trans. IP*, vol. 17, no. 4, pp. 564–573, Apr. 2008.

[6] N. D. B. Bruce and J. K. Tsotsos, "Attention based on information maximization," *Journal of Vision*, 7(9):950a, 2007.

[7] O. Le Meur, P. Le Callet and D. Barba, "A coherent computational Approach to model the bottom-up visual attention," *IEEE Trans. PAMI*, vol. 28, no. 5, pp. 802-817, 2006.

[8] H. Liu and I. Heynderickx, "Visual Attention Modeled with Luminance Only: from Eye-Tracking Data to Computational Models," *VPQM*, January 2010.

[9] P. J. Burt and E. H. Adelson, "The Laplacian Pyramid as a compact image code," *IEEE Trans. on Communications*, vol. 31, no. 4, pp. 532-540, 1983.

[10] P. Tseng, R. Carmi, I. G. M. Camerson, D. P. Munoz and L. Itti, "Quantifying center bias of observers in free viewing of dynamic natural scenes," Journal of Vision, vol. 9, no. 7, 2009.

[11] 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