INCORPORATING VISUAL ATTENTION MODELS INTO IMAGE QUALITY METRICS

Welington Y.L. Akamine and Mylène C.Q. Farias, Member, IEEE

Departament of Computer Science
University of Brasília (UnB),
Brasília, DF, 70910-900, Brazil

I. INTRODUCTION AND MOTIVATION

Objective visual quality metrics can be classified as *data metrics*, which measure the fidelity of the signal without considering its content, or *picture metrics*, which estimate quality considering the visual information contained in the data. Customarily, quality measurements in the area of image processing have been largely limited to a few data metrics, such as mean absolute error (MAE), mean square error (MSE), and peak signal-to-noise ratio (PSNR), supplemented by limited subjective evaluation. Although over the years data metrics have been widely criticized for not correlating well with perceived quality measurements, it has been shown that such metrics can predict subjective ratings with reasonable accuracy as long as the comparisons are made with the same content, the same technique, or the same type of distortions.

One of the major reasons why these simple metrics do not generally perform as desired is because they do not incorporate any human visual system (HVS) features in their computation. It has been discovered that, in the primary visual cortex of mammals, an image is not represented in the pixel domain, but in a rather different manner. Unfortunately, the measurements produced by metrics like MSE or PSNR are simply based on a pixel to pixel comparison of the data, without considering what is the content and the relationships among pixels in an image (or frames). In the past few years, a big effort in the scientific community has been devoted to the development of better image and video quality metrics that incorporate HVS features (i.e. picture metrics) and, therefore, correlate better with the human perception of quality [1].

A recent development in the area of image quality consists of trying to incorporate the perceptual importance of the different areas of the scene in the design of the metric. It has been hypothesized that visual distortions appearing in less salient areas might be less visible and therefore less annoying. As a consequence, researchers have started lately to incorporate saliency information in objective metrics, for example through visual importance pooling [2], [3]. In this paper, our goal is to investigate how computational visual attention models can be incorporated into image quality metrics.

II. VISUAL ATTENTION AND IMAGE QUALITY

When observing a scene, the human eye typically filters the large amount of visual information available on the scene and attend to selected areas [4]. The selection of areas is controlled through oculo-motor mechanisms that allow the gaze of attention to either hold on a particular location (fixation) or to shift to another location when sufficient information has already been collected (saccades). The selection of *fixations* is controlled by the "bottom-up" attention, which is based on the visual properties of the scene. Priority is given to areas with a high concentration of information, minimizing the amount of data to be processed by the brain while maximizing the quality of the information collected.

One of the most popular computational models for visual attention was proposed by Itti [4]. Itti's model analyzes the images looking for the most relevant visual properties: color, intensity, and orientation. The algorithm creates maps corresponding to these three properties and combines these maps to predict the saliency of the regions. Using a *winner-take-all* approach, predicted fixations are selected as the most salient points in the image.

In order to combine the attention information with the objective quality metrics a saliency map is needed, instead of the individual (ordered) fixation points given by Itti's algorithm. To obtain such a map, we filter the set of predicted fixation points with a Gaussian filter. In Figs. 1 and 2 the original image 'Caps' and the corresponding saliency map created by this process are depicted. Higher (lighter) luminance values correspond to a higher saliency, while lower (darker) values correspond to points where a low attention was predicted.

In the last few years, many works have incorporated saliency maps into image quality metrics. Nevertheless, most works use saliency maps obtained experimentally with the help of eye-trackers [5] [6]. In this work, our goal is to investigate the benefit of incorporating the information of

This work was supported in part by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) Brazil, in part by Fundação de Empreendimentos Científicos e Tecnológicos (Finatec), and in part by University of Brasília (UnB) ProIC/DPP.

Fig. 1. Images 'Caps'.

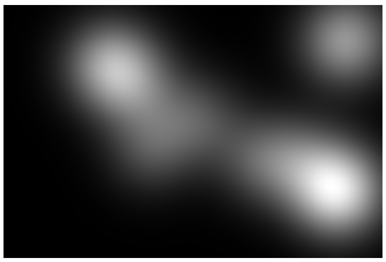

Fig. 2. Saliency map of image 'Caps' in Fig. 1.

saliency maps obtained using **computational** models into objective quality metric models.

More specifically, in this paper we calculate the saliency map of colored images using a slightly modified version of Itti's attention model. Then, we incorporate the information from the saliency map into three different metrics: *Mean Square Error* (MSE), *Peak signal-to-noise ratio* (PSNR), and *Structural Similarity* (SSIM) [7]. While PSNR and MSE are two very simple fidelity metrics, SSIM is a more complex FR image quality metric that makes use of HVS properties to estimate quality. The general equation for SSIM is:

$$SSIM(I_o, I_t) = \frac{(2\mu_o\mu_t + C_1)(2\sigma_{ot} + C_2)}{(\mu_o^2 + \mu_t^2 + C_1)(\sigma_o^2 + \sigma_t^2 + C_2)}. \quad (1)$$

where μ the average intensity, σ is the standard deviation, and σ_{ot} is the covariance between the original image (I_o) and the test image (I_t). The variables C_1, C_2, C_3 are control constants used to avoid problems when the denominator reaches values close to zero.

In order to combine the saliency and quality models, we used an approach similar to the approach used by Liu and Heynderickx [6]. Since all the metrics considered generate 'distortion maps' that have the same size as the 'saliency maps', we use the points in the saliency maps as weights for the distortion maps. In the case of MSE, the integration model is obtained by multiplying each *pixel* in the MSE-distortion map by the corresponding pixel in the saliency map. In other words, the modified saliency-based MSE (SBMSE) is given by:

$$\text{SBMSE} = \frac{\sum_{x=1}^{M}\sum_{y=1}^{N} MSE(x,y)\cdot S(x,y))}{\sum_{x=1}^{M}\sum_{y=1}^{N} S(x,y)}, \quad (2)$$

where MSE(x, y) is the MSE-distortion map, $S(x, y)$ is the saliency map pixel, and x and y are the horizontal and vertical coordinates.

In the same way, the modified saliency-based PSNR (SBPSNR) is given by the following equation.

$$\text{SBPSNR} = 20\log\frac{MAX_i}{\sqrt{\text{SBMSE}}}. \quad (3)$$

Also, the modified saliency-based $SSIM$ (SBSSIM) formula is given by

$$\text{SBSSIM} = \frac{\sum_{x=1}^{M}\sum_{y=1}^{N}(SSIM_{map}(x,y)\cdot S(x,y))}{\sum_{x=1}^{M}\sum_{y=1}^{N} S(x,y)} \quad (4)$$

where $SSIM_{map}(x, y)$ is the SSIM-distortion map.

Preliminary Results show that the saliency information has improved the results of the two simple fidelity metrics tested. For the more complex metric, $SSIM$, the addition of saliency did not always represent an improvement in performance. These results are not surprising given that some of the more complex metrics already incorporate the visual properties considered by the attention models.

III. REFERENCES

[1] Zhou Wang and Alan C Bovik. Mean Squared Error : Love It or Leave It ? *IEEE Signal Processing Magazine*, (January):98–117, 2009.

[2] C Oprea, I Pirnog, C Paleologu, and M Udrea. Perceptual Video Quality Assessment Based on Salient Region Detection. In *Telecommunications, 2009. AICT '09. Fifth Advanced International Conference on*, pages 232–236, May 2009.

[3] J. Redi, Hantao Liu, P. Gastaldo, R. Zunino, and I. Heynderickx. How to apply spatial saliency into objective metrics for jpeg compressed images? In *Image Processing (ICIP), 2009 16th IEEE International Conference on*, pages 961 –964, nov. 2009.

[4] Laurent Itti and Christof Koch. Computational modelling of visual attention. *Nature Reviews Neuroscience*, 2(3):194–203, 2001.

[5] Hantao Liu and Ingrid Heynderickx. Studying the added value of visual attention in objective image quality metrics based on eye movement data.

[6] Hantao Liu and I. Heynderickx. Studying the added value of visual attention in objective image quality metrics based on eye movement data. In *Image Processing (ICIP), 2009 16th IEEE International Conference on*, pages 3097 –3100, nov. 2009.

[7] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE TRANSACTIONS ON IMAGE PROCESSING*, 13(4):600–612, 2004.