HOW TO APPLY SPATIAL SALIENCY INTO OBJECTIVE METRICS FOR JPEG COMPRESSED IMAGES?

Judith Redi, Hantao Liu[1], Paolo Gastaldo, Rodolfo Zunino, and Ingrid Heynderickx[1,2]

University of Genoa, DIBE, Via Opera Pia 11a - 16145 Genova – Italy
[1]Delft University of Technology, Mekelweg 4 - 2628 CD Delft – The Netherlands
[2]Philips Research Laboratories, Prof. Holstlaan 4 - 5656 AA Eindhoven – The Netherlands

ABSTRACT

This paper investigates how saliency obtained from eye-tracking data can be integrated into objective metrics for JPEG compressed images. The objective metrics used in this paper are both based on features, locally extracted from the images and serving as input to a neural network for the overall quality prediction. We compare various weighting functions to combine saliency with these objective metrics, taking into account the possible distraction due to artifacts that might affect the quality judgment. Experimental results indicate that including saliency into objective metrics in an appropriate way can further enhance their performance.

Index Terms— Image quality assessment, objective metric, visual attention, neural networks

1. INTRODUCTION

Modern research on image quality metrics [1,2] seems to favor the integration of saliency in distortion quantification models. When observing an image, the human eye traverses it to gather visual information efficiently, neglecting poorly informative regions (typically, background areas). Hence, one would expect intuitively that visual attention plays a significant role in distortion visibility, by enhancing or reducing the actual visibility depending on saliency.

Attention data can be either collected with subjective experiments using an eye tracker, or can be modeled. As a starting point, the use of eye-tracking data seems to be more appropriate, making the results independent of the reliability of the existing models. It has been shown in the past that eye-tracking data collected when evaluating image quality differ from those collected during free looking, especially in the case of JPEG compressed images [3,4]. These results indicate that artifacts may distract attention away from the natural scene saliency, and as such may affect the observer's quality judgment. It should also be noticed that artifacts can degrade some regions more heavily than others, e.g. due to luminance or texture masking. In the case of JPEG compression this can result in a higher annoyance for artifacts in the background than in the foreground (Fig. 2). This distraction away from natural scene saliency seems to depend on the specific kind of distortion [3,5].

The core problem in defining an effective strategy for the integration of saliency into objective metrics is finding coherence with human perception. The typical integration strategy [2, 6], consisting of the multiplication of each local objective metric value with the corresponding measured saliency (Fig.1a), implies that artifacts in neglected regions are assumed to be less annoying than those in attention areas. This strategy showed limitations when dealing with distortions such as JPEG compression [6]. This might be due to underestimating the distracting power of background artifacts, and their impact on the quality judgment [3,5].

In [7], a less trivial integration strategy (PF-SSIM) is proposed, taking into account both the saliency and the annoyance level of artifacts. In a first step, areas corresponding to fixations are weighted higher; afterwards, objective metric values are amplified for regions poor in quality (independent of their saliency). In this way, neglecting heavily distorted background regions is partially avoided. The effectiveness of this approach proves that more sophisticated integration strategies are worth to be investigated. However, PF-SSIM seems to perform differently for different distortions, obtaining less prominent results for JPEG compressed images.

This paper investigates various saliency integration strategies to predict perceived quality especially for JPEG images. Instead of simply using the saliency data, the refined scheme weights the local distortion with a specific function of the saliency data (see Fig. 1b). Artificial Neural

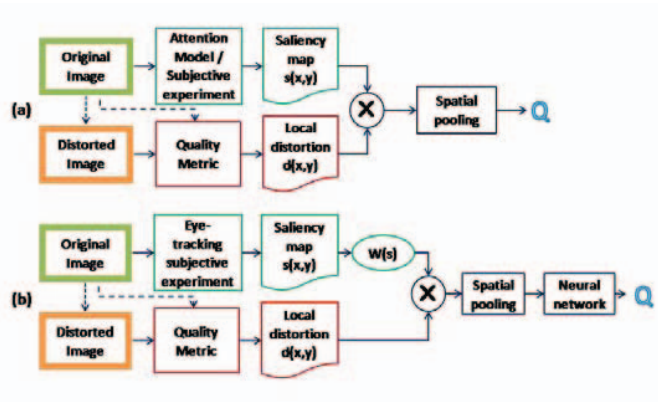

Figure 1 - A common approach for attention information integration into the quality assessment process (a), and our proposed approach (b)

(a) *(b)* *(c)*

Figure 2. Effects of JPEG compression *(b)* on a high quality sample *(a)*. By simple weighting artifact visibility according to salience *(c)* the annoyance of the background is neglected.

Networks are exploited to perform the non-linear mapping between the locally weighted distortion metrics and the overall quality estimation. The investigation is applied to two objective metrics targeted to JPEG compressed images. In Section 2 these two objective metrics are explained in more detail. Section 3 describes the integration strategy. Section 4 presents the use of a neural network for modeling the nonlinearity. The experimental validation of the approach is reported in Section 5.

2. OBJECTIVE METRICS TO QUANTIFY PERCEIVED BLOCKINESS

The two objective metrics considered in this paper are a blockiness-specific metric (NPBM) [8] and a Color Distribution-based [9] (CDM) metric. Both metrics aim at predicting quality degradation due to JPEG compression, but differ in several aspects: first, NPBM is a no-reference metric, whereas CDM is a reduced-reference one; secondly, NPBM quantifies the annoyance due to blockiness, while CDM encompasses the quality degradation due to all compression artifacts; finally, the two metrics are computed on a different local basis (see Fig. 3).

Given an image, I_o, and its distorted version, I_d, NPBM first identifies the grid of blocking artifacts in I_d, and then assigns to each pixel on the grid a value proportional to the blockiness visibility. CDM estimates the annoyance of artifacts by analyzing the color distribution across sub-regions of the image; the procedure yields a pair of vectors (one based on I_o and the other on I_d), holding as many values as the number of blocks in which the image is split. In both cases the local values first need to be combined with local saliency, and then serve as input for the neural network to estimate an overall quality score.

3. INTEGRATING SALIENCY INTO THE METRICS

The saliency integrated in the objective metrics is the natural scene saliency (NSS), obtained from eye-tracking experiments on original, uncompressed images. As mentioned above, the overall quality assessment might be affected by distraction of attention to the artifacts, but this is included in a later step by suitably weighting the natural scene saliency data before integration. Hence, when using the distracted saliency of distorted images, the artifact annoyance information might be double and overestimate the importance of artifacts in the overall quality estimation.

NPBM	CDM
Input: Image I_d, size W_I x H_I	**Input:** Images I_d, I_o, size W_I x H_I
1. Detect the grid of blocking artifacts BG(x,y) in I_d	1. Split I_d, I_o into n_b square, adjacent non overlapping regions
2. For each pixel(x,y) in the grid $G_d[x,y] = BM(I_d[x,y])$	2. For each region b_i in I_d, I_o $BV_o[i] = CDM(b_{i,o})$ $BV_d[i] = CDM(b_{i,d})$
Output: distortion grid G_d, dimension $W_I/8 \times H_I/8$	**Output:** BV_o, BV_d, both of dimension 1 x n_b

Figure 3. Pseudocode for objective blockiness metrics, assuming the size of blocking artifacts is 8x8.

3.1. Estimating local saliency

Eye tracking records the observer's pupil movements in terms of fixation points and saccades. From these data, a saliency map is constructed by applying to each fixation point a gray-scale patch, having a Gaussian intensity distribution with variance, σ, that approximates the size of the fovea (about 2° visual angle). The saliency value, $NSS_i(k,l)$, at location (k,l) of the saliency map for image I_i (having W_I x H_I pixels) is defined as:

$$NSS_i(k,l) = \sum_{j=1}^{T} \exp\left[-\frac{(x_j - k)^2 + (y_j - l)^2}{\sigma^2}\right] \quad (1)$$

where $k \in [1, W_I]$, $l \in [1, H_I]$; (x_j, y_j) are the spatial coordinates of the jth fixation ($j=1 \ldots T$) recorded for all subjects.

3.2. Integration strategy

To integrate human saliency in the objective quality prediction, the principle of locally weighting distortion metrics is still applied. The integration process instead is modified, including the use of specifically designed functions of the saliency. Hence, the following steps need to be performed (see also Fig.1.b):

1. A predefined function of the saliency data $W(S)$ is computed;
2. The local distortion values are weighted with the local modified saliency values;
3. A spatial pooling strategy assembles a global descriptor of the image.

Let $NSS(x,y)$ be the value of natural scene saliency measured experimentally; this research considers three weighting functions for the saliency data:

$$WNSS_1(x,y) = NSS(x,y) + 1$$

$$WNSS_2(x,y) = \begin{cases} 1 - NSS(x,y); & if\ NSS(x,y) < 0.5 \\ NSS(x,y); & otherwise \end{cases}$$

$$WNSS_3(x,y) = 1 - 2\exp(-NSS(x,y)/\sigma)$$

$WNSS_1$ (as also used in [6]) privileges foreground areas higher, but avoids the inconvenience of nullifying all background areas. $WNSS_2$ highlights both foreground and background regions, attenuating the transition (e.g., edges) areas. Finally, $WNSS_3$ is a smoothed version of $WNSS_2$, being a reverse Gaussian window.

The procedure that integrates saliency in the objective metrics operates on a local basis, and therefore complies with the local computation strategy involved in the specific quality metrics. The procedure for the NPBM metric operates on the distortion grid, whereas the procedure associated with CDM proceeds on a region-by-region basis. The associated pseudo-codes are outlined as follows:

NPBM

Inputs: a distortion grid G_d, a saliency map NSS,a weighting function WNSS

1. Extract from NNS(x,y) pixels included in the G_d and store them in a salience grid

$$SG[x,y] = \{NSS(x,y) \mid (x,y) \in G_d\}$$

2. For each pixel in the grid, compute the weighting coefficient

$$WSG[x,y] = WNSS(SG[x,y])$$

3. For each pixel (x,y) in the grid G_d *compute the weighted metric WSM*

$$WSM\_NPBM[x,y] = G_d[x,y]*WSG[x,y]$$

Output: *WSM_NPBM*, of dimension $W_I/8$ x $H_I/8$

CDM

Inputs: two distortion vectors BV_o, BV_d, a saliency map NSS, a weighting function WNSS.

1. Divide NSS(x,y) in n_b sub-regions b_i of n_p pixels, corresponding to those employed for the CDM computation, and for each b_i compute the average salience value:

$$SB[i] = \frac{1}{n_p} \sum_{(x,y) \in b_i} NNS(x,y)$$

2. For each block in the vector, compute the weighting coefficient

$$WSB[i] = WNSS(SB[i])$$

3. For each element i in BV_d compute the weighted metric WSM

$$WSM\_CDM_o[i] = BV_o[i]*WSB[i]$$
$$WSM\_CDM_d[i] = BV_d[i]*WSB[i]$$

Output: WSM_CDM_o,WSM_CDM_d, both of dimension $1xn_b$

The third step involves spatial pooling to aggregate the locally weighted distortion information into a single global descriptor. A statistical approach is employed for that purpose, and uses a percentile-based representation of the distribution of the weighted metric values over the image. The resulting global descriptor feeds the neural network.

4. NEURAL NETWORKS FOR OBJECTIVE METRICS ENHANCEMENT

Machine learning tools can bring a significant contribution to quality assessment systems [9]. In fact, previous research proved that the use of computational intelligence methods to map a numerical image description into a quality score can render the process of human perception quite effectively. To that end, an extension of the Multi Layer Perceptron paradigm, namely, the "Circular Back Propagation" (CBP) neural network (NN) [10] is proposed. A CBP NN includes an additional input, computed as the sum of the squared values of all input elements. This addition allows the NN either to adopt the standard sigmoidal behavior, or a bell-shaped radial function, depending on the data. As a result, the NN does not need any a-priori assumption to formulate the model, yet allowing the use of conventional back-propagation algorithms for training.

The task of the CBP NN is to approximate the mapping function between the percentile-based global descriptor of the objective metric and the actual subjective quality rating. In case of the NPBM metric a single CBP NN was employed. When dealing with the CDM color based metric, the ensemble strategy described in [8] was adopted.

5. EXPERIMENTAL RESULTS

5.1. Visual attention data collection

To obtain the visual attention data, eye movements were recorded with an infrared video-based tracking system (iView X RED, SensoMotoric Instruments), having a sampling rate of 50 Hz, a spatial resolution of 0.1°, and a gaze position accuracy of 0.5°-1.0°. A chin rest was employed to reduce head movements and to ensure a constant viewing distance of 70 cm. The 29 source images of the LIVE database [11] were displayed on a 19-inch CRT monitor with a resolution of 1024x768 pixels and an active screen size 365x275mm. The images were shown to twenty inexperienced participants, who were requested to freely look to the images during 10s. Each session was preceded by a 3x3 point grid calibration. The intensity of the resulting saliency map was linearly normalized to the range [0, 1].

5.2. Experimental validation

For the experimental validation only the JPEG dataset of LIVE was employed, being a well established benchmark for quality metrics' tests. The dataset was divided into two groups: a training set including 161 stimuli, and a second group for testing the generalization ability of the NN, including 72 stimuli. Image content included in the test set was not part of the training set, to ensure a robust mapping, independent of the particular content of the samples.

NPBM was implemented as described in [8], and CDM as reported in [9]. Eleven percentiles of the distribution of the blockiness visibility computed with NPBM were taken for the global descriptor, while for CDM 6 percentiles of the distribution of metric values for both the original and the distorted image were merged in a single global descriptor. All NN employed in the experiment were equipped with 3 hidden neurons.

For both metrics the performance was validated:

1. without any saliency information;
2. with integration of saliency using proportional weighting ($WNSS_0$ in the following);
3. with integration of saliency using the three weighting functions $WNSS_1$, $WNSS_2$, $WNNS_3$.

To evaluate the metrics' performance, the correlation coefficients and the RMSE were chosen as significant indicators [12] and reported in Table I for both metrics and every studied weighting function. Table I also gives the difference in performance of every weighted metric with respect to the original one (i.e., without integrating salience information). A first, clear outcome is that the proportional weighting penalizes the metrics' performance, as also observed in [6]. Conversely, when enhancing the metrics with a more specific weighting function of the saliency data, an improvement in performance can be observed, consistently for both metrics. In particular, $WNSS_1$ provides the highest gain in performance, indicating that indeed regions of interest should be weighted higher, but background regions should not be nullified. $WNSS_3$ yields a slightly worse performance with respect to $WNSS_2$: this may be explained by the fact that the $WNSS_3$ weights mid-saliency regions to zero, causing again the loss of possible important information for quality evaluation.

Compared to PF-SSIM [7], our proposed approach proves to be more effective, in terms of improvement of performance with respect to the original metric, independent of the weighting function used. Evaluating the accuracy of each metric, it should be taken in account that in [7], no subset of images was used to test the regressed model; therefore its robustness is not proved. Conversely, the proposed model is proved to be reliable and independent of the specific image content.

Table 1 - Performance indicators (Pearson Correlation Coefficient and RMSE) for NPBM and CDM, for every studied weighting function. Performance of the PF-SSIM metric is also given as a term of comparison.

		Pearson CC	Diff. wrt No Salience	RMSE	Diff. wrt No Salience
NBPM	No Salience	0.9119	-	11.1536	-
	WSSN0	0.8188	- 0.0931	16.2731	5.1195
	WSSN1	0.9301	0.0182	9.9035	- 1.2501
	WSSN2	0.9221	0.0102	10.3373	- 0.8164
	WSSN3	0.9216	0.0097	10.3243	- 0.8293
CDM	No Salience	0.9005	-	6.1895	-
	WSSN0	0.6596	- 0.2409	11.5149	5.3254
	WSSN1	0.9109	0.0104	5.8793	- 0.3102
	WSSN2	0.9103	0.0098	5.9984	- 0.1911
	WSSN3	0.9056	0.0051	6.1151	- 0.0745
Moorth y and Bovik	SSIM	0.9695	-	6.4217	-
	PF-SSIM	0.9737	0.0042	6.4385	0.0168
	MS SSIM	0.9635	-	6.4982	-
	MS PF-SSIM	0.9659	0.0024	6.2847	- 0.2135

Overall results indicate that the proposed integration strategy is promising, although in a preliminary stage. Designing appropriate saliency weighting functions, related to the particular artifact observed, could indeed bring some added value to objective quality metrics, provided those functions contribute to consistently model the attention distraction caused by artifacts.

5. CONCLUSIONS

A new method for integrating visual attention into quality assessment systems for JPEG compressed images is proposed. Instead of directly weighting the local distortion metric values with the actual saliency values, appropriate functions are used to weight the attention data, before combining them with the distortion metric values. Experimental results seem to validate the proposed approach, confirming an enhancement in performance when appropriately weighted human saliency is embedded in the metric.

6. REFERENCES

[1] R. Barland and A. Saadane, *Blind Quality Metric using a Perceptual Importance Map for JPEG-2000 Compressed Images*, in Proc. IEEE Int. Conf. ICIP 2006,

[2] N. Sadaka, L. Karam, R. Ferzli, and G. Abousleman, *A no-reference perceptual image sharpness metric based on saliency-weighted foveal pooling*, in Proc. IEEE Int. Conf. ICIP 2008

[3] A. Ninassi, O. Le Meur, P. Le Callet, D. Barba, A. Tirel, *Task impact on the visual attention in subjective image quality assessment*, in Proc. EUSIPCO-06 (2006)

[4] C. Vu, E. C. Larson, and D. Chandler, *Visual Fixation Patterns when Judging Image Quality: Effects of Distortion Type, Amount, and Subject Experience,* in Proc. IEEE SSIAI 2008

[5] Z.Wang and X.Shang, *Spatial pooling strategies for perceptual image quality assessment*, in Proc. IEEE Int. Conf. ICIP 2006

[6] A. Ninassi, O. Le Meur, P. Le Callet, and D.Barba, *Does where you gaze your attention on an image affect your perception of quality? Applying Visual Attention to image quality metrics*, in Proc. IEEE Int. Conf ICIP 2007

[7] A. Moorthy and A. Bovik, "Perceptually significant spatial spooling techniques for image quality assessment," in Proc. Electronic Imaging, 2009

[8] H. Liu and I. Heynderickx, *A No-Reference Perceptual Blockiness Metric*, in Proc. IEEE Int. Conf. ICASSP, pp. 865-868,

[9] J. Redi, P. Gastaldo, R. Zunino and I. Heynderickx: *Reduced-reference assessment of perceived quality by exploiting color information*. In Proc. VPQM '09 (2009)

[10] S. Ridella, S. Rovetta, R. and Zunino,: *Circular back-propagation networks for classification*. IEEE Trans. Neural Networks 8 (1997) 84-97

[11] R. Sheikh, Z. Wang, L. Cormack, and A. Bovik: LIVE Image Quality Assessment Database http://live.ece.utexas.edu/research/quality

[12] VQEG, "Final report from the video quality experts group on the validation of objective models of video quality assessment", March 2000, http://www.vqeg.org/.