An Image Quality Assessment Using Mean-Centered Weber Ratio and Saliency Map

Soyoung CHUNG†, Nonmember and Min Gyo CHUNG†a, Member

SUMMARY Chen proposed an image quality assessment method to evaluate image quality at a ratio of noise in an image. However, Chen’s method had some drawbacks that unnoticeable noise is reflected in the evaluation or noise position is not accurately detected. Therefore, in this paper, we propose a new image quality measurement scheme using the mean-centered WLNI (Weber’s Law Noise Identifier) and the saliency map. The experimental results show that the proposed method outperforms Chen’s and agrees more consistently with human visual judgment.

key words: image quality assessment, Weber’s law, saliency map, image enhancement, human visual perception

1. Introduction

A metric or process which evaluates the quality of an image is commonly called IQM (Image Quality Measure). Entropy, AMBE (Absolute Mean Brightness Error), PSNR (Peak Signal-to-Noise Ratio), and MSSIM (Multi-scale Structural Similarity) are some frequently used IQMs. Chen developed an IQM to incorporate the characteristics of the human visual perception for luminance and texture, and measure the image quality at a ratio of noise in the image [1]. Compared to existing methods, Chen’s method showed an improved result, but had the weakness that unnoticeable noise is reflected in the evaluation or noise position is not accurately detected. Therefore, in this paper, we propose a new image quality measurement scheme, which uses the mean-centered WLNI (Weber’s Law Noise Identifier) and the saliency map to remove the Chen’s shortcomings. According to the experimental results, the proposed IQM detects the noise artifacts more accurately than Chen’s and agrees more consistently with human visual judgment.

2. The Proposed IQM

This section gives the proposed IQM, which is an extension of Chen’s. A block diagram of the proposed IQM is shown in Fig. 1. Two major modifications made to Chen’s IQM are: a) Chen observed that most annoying noise artifacts appear in the form of edges, and used the Sobel edge operator to detect edge points with noise artifacts. However, inspired by WLNI [2], we use the mean-centered WLNI, a variant of WLNI designed for the purpose of image quality assessment; and b) The saliency map is a 2D map to encode the local conspicuity of objects in the visual environment [3]. We use the saliency map to find the attended locations in the image and give a higher penalty to the noise points that occur in the region of visual attention.

1. A color image is converted to the gray-scale image $I(x,y)$, where $I(x,y)$ represents the intensity value of a pixel at $(x,y)$. Consider a set of neighboring pixels around a pixel $(x_c, y_c)$ in the image $I$. Typically the set is a block $B$ of size $n \times n$ centered at the pixel $(x_c, y_c)$. Then, the mean-centered WLNI at the pixel $(x_c, y_c)$, called $M(x_c, y_c)$, is defined as follows.

$$M(x_c, y_c) = \begin{cases} \frac{\sum_{y=0}^{h-1} \sum_{x=0}^{w-1} [I(x,y)-H(x,y)]}{\mu-I(x_c,y_c)}, & \text{if } \mu \neq I(x_c, y_c) \\ \text{otherwise} \end{cases}$$

where $\mu$ is the average intensity of the image $I$, i.e., $\mu = (\sum_{i=0}^{w-1} \sum_{j=0}^{h-1} I(x,y))/wh$ ($w$ and $h$ are the $I$’s width and height, respectively). Actually, $M(x_c, y_c)$ denotes the edge magnitude at the pixel $(x_c, y_c)$. Compute $M_s(x,y)$ and $M_t(x,y)$ for all the pixels $(x,y)$ in the original image and the test image, respectively.

2. Let $I(x,y)$ be the subimage of size $m \times m$ centered at the pixel $(x_c, y_c)$ and $P_{I(x_c,y)}(k)$ be the probability density function for the subimage. Compute the local entropy $H[P_{I(x_c,y)}]$ or $H(x,y)$ for all the subimages in the original image. The entropy is computed to measure the texture strength in the subimage.

3. Calculate the noise map $I_N(x,y)$ by using Eq. (1).

$$I_N(x,y) = \begin{cases} 1, & \text{if } (H(x,y) < T_1) \&\& (M_s(x,y) < T_2) \&\& (M_t(x,y) > T_3) \\ 0, & \text{otherwise} \end{cases}$$
where $T_1 = 3.5, T_2 = 0.2,$ and $T_3 = 0.4$ are empirically estimated thresholds. Equation (1) tells that a pixel is classified as a noise point if it is in the smooth area, edge is not detected in the original image, but detected in the test image.

4. As shown in [3], compute the saliency map for the original color image. In short, the computation process consists of the following steps.

- Assuming that $r$ (red), $g$ (green), and $b$ (blue) represent 3 component images of the input color image, extract some visual features related to intensity, colors, and orientations: a) intensity image $I = (r + g + b)/3$; b) four normalized color images ($R = r - (g + b)/2$, $G = g - (r + b)/2, B = b - (r + g)/2, Y = (r + g)/2 - |r - g|/2 - b$); c) four orientation maps $O(\theta)$ for orientations $\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$, which are obtained from $I$ using Gabor filters.

- Using the above visual features, create dyadic Gaussian pyramids $I(\zeta), R(\zeta), G(\zeta), B(\zeta), Y(\zeta)$, and $O(\zeta, \theta)$, where $\zeta \in [0.8]$ is the scale.

- Perform center-surround difference operation (denoted $\ominus$) on the Gaussian pyramids to produce 42 feature maps: 6 for intensity, 12 for color, and 24 for orientation.

$$I(c, s) = |[I(c) \ominus I(s)|,$$

$$RG(c, s) = |[R(c) - G(c)) \ominus (G(s) - R(s))|,$$

$$BY(c, s) = |[B(c) - Y(c)) \ominus (Y(s) - B(s))|,$$

$$O(c, s, \theta) = |O(c, \theta) \ominus O(s, \theta)|,$$

where $c \in \{2, 3, 4\}$ and $s = c + \delta, \delta \in [3, 4]$.

- 42 feature maps are combined through across-scale addition operation (denoted $\oplus$) into 3 conspicuity maps ($\bar{T}$ for intensity, $\bar{C}$ for color, and $\bar{O}$ for orientation) at the scale $\sigma = 4$.

$$\bar{T} = \oplus_{c=2}^{4} \oplus_{s=c+\delta}^{s+c+3} N(I(c, s)),$$

$$\bar{C} = \oplus_{c=2}^{4} \oplus_{s=c+\delta}^{s+c+3} N(RG(c, s) + N(BY(c, s)),$$

$$\bar{O} = \sum_{\theta \in \Theta} N(\oplus_{c=2}^{4} \oplus_{s=c+\delta}^{s+c+3} N(O(c, s, \theta)),$$

where $N(\cdot)$ is a map normalization function and $\Theta = \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$.

- Lastly, the saliency map $S$ is obtained as follows:

$$S = \frac{1}{3} \left( N(\bar{T}) + N(\bar{C}) + N(\bar{O}) \right).$$

Now convert $S(x, y)$ into the binary saliency map $\hat{S}(x, y)$ (with threshold value of 0.2), where $\hat{S}(x, y) = 1$ means that the pixel $(x, y)$ belongs to the region of visual attention. Modify $I_N(x, y)$ using $\hat{S}(x, y)$.

$$I_N(x, y) = \begin{cases} \alpha I_N(x, y), & \text{if } \hat{S}(x, y) = 1 \\ I_N(x, y), & \text{otherwise,} \end{cases}$$

where $\alpha$, a penalty given to the noise points within the visually attended region, is currently set to 2.

5. Finally, compute the rating score $R$.

$$R = \frac{\sum_{x=0}^{N-1} \sum_{y=0}^{N-1} I_N(x, y)}{wh}.$$

3. Experimental Results

For the experiment, 15 images were chosen from the Kodak and LIVE image databases [4], [5]. Upon these image, we executed 6 histogram equalization-based image enhancement methods (HE, BBHE, BHEPL, DSIHE, RMSHE, and RSIHE) [1] and acquired a total of 90 test images.

25 people participated in the experiment and produced human Mean Opinion Scores (MOS) for the test images. The respondents were asked to rate the test image by comparing it with the original image. The raw values $v_{ij}$ for the $i$th respondent and the $j$th test image were converted into the $Z$ values $Z_{ij}$: $Z_{ij} = (v_{ij} - \mu_j)/\sigma_j$, where $\mu_j$ and $\sigma_j$ are the mean and standard deviation of all the scores rated by the $i$th respondent, respectively. The MOS for the $j$th image, $MOS_j$, was then defined as the average of $Z_{ij}(i = 1, 2, \cdots, 25)$.

To evaluate the performance of an IQM, it is necessary to examine how the quality rating scores $\{R_j\}$ of the IQM are correlated with the MOS values $\{MOS_j\}$. As demonstrated in VQEG [6], the performance of the IQM can be evaluated from 3 perspectives: a) prediction accuracy: the IQM’s ability to predict the human’s MOS with low error, b) prediction monotonicity: the degree to which the IQM’s predictions agree with the relative magnitudes of MOS values, and c) prediction consistency: the degree to which the IQM maintains prediction accuracy over a variety of images. These attributes were evaluated through the following metrics: Pearson correlation coefficient (PCC) for prediction accuracy, Spearman rank order correlation coefficient (SROCC) for prediction monotonicity, and Outlier ratio (OR) for prediction consistency. OR is the ratio of outlier-scores to total scores. The $j$th point (test image) is an outlier if it satisfies $|MOS_j - R_j| > 2 \Delta_j$, where $\Delta_j$ is the standard deviation of $Z_{ij}(i = 1, 2, \cdots, 25)$.

Table 1 shows the performance evaluation results of existing IQMs and the proposed IQM. All the three metrics clearly indicate that the proposed IQM works best and yields the evaluation results to be more consistent with human visual judgment.

As mentioned earlier, Chen’s IQM had the drawbacks that noise position is not correctly detected and visually unnoticeable noise is reflected in the image quality evaluation. As demonstrated in Figs. 2 and 3, the proposed IQM

<table>
<thead>
<tr>
<th></th>
<th>PCC</th>
<th>SROCC</th>
<th>OR</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSNR</td>
<td>0.1607</td>
<td>0.1204</td>
<td>0.0755</td>
</tr>
<tr>
<td>SSIM</td>
<td>0.2560</td>
<td>0.2685</td>
<td>0.0643</td>
</tr>
<tr>
<td>UIQ</td>
<td>0.2321</td>
<td>0.2314</td>
<td>0.0760</td>
</tr>
<tr>
<td>Chen</td>
<td>0.0695</td>
<td>0.0700</td>
<td>0.0771</td>
</tr>
<tr>
<td>Proposed</td>
<td>0.5077</td>
<td>0.5237</td>
<td>0.0090</td>
</tr>
</tbody>
</table>
performs much better than Chen’s in the above 2 aspects: namely, the proposed IQM more accurately detects the positions of noise points that are visible to human eyes (Fig. 2), and effectively excludes the noise points that are imperceptible to human eyes from the image quality assessment (Fig. 3).

More specifically, to locate the true noise points of the subimage in Fig. 2 (c), we ask human specialists to manually trace out the contour along noisy edges. As a result, we can identify 2,533 true noise points on the contour. We then match this true contour with those in Figs. 2 and 3, and obtain 1,196 matched points (hit rate = 1,196/2,533 = 0.47) and 1,651 matched points (hit rate = 1,651/2,533 = 0.65) for Chen’s and the proposed IQMs, respectively. The ratio of the hit rates (γ1 = 0.65/0.47 = 1.38) implies that the proposed IQM is 1.38 times better than Chen’s in the performance of noise position detection. The box plot for the γ1 values of all 90 test images is shown in Fig. 4 (a).

Now consider detecting the imperceptible noise points of the subimage in Fig. 3 (a). In the similar manner above, we can find that the size of the subimage is 264 × 412, so the total number of true non-noise points in the subimage is 108,768 (= 264 × 412); however, the number of noise points (in white) in Fig. 3 (b) is 8,575 (false noise detection rate = 8,575/108,768 = 0.079); and the number of noise points in Fig. 3 (c) is 524 (false noise detection rate = 524/108,768 = 0.005). The ratio of the false noise detection rates (γ2 = 0.079/0.005 = 16.36) implies that the proposed IQM is 16.36 times better than Chen’s in the performance of imperceptible noise detection. Figure 4 (b) shows the box plot for the γ2 values of the entire test images. We believe that the combined effect of both the mean-centered WLNI and the saliency map yields the excellent performance of the proposed IQM, as shown in Fig. 4. Especially, the mean-centered WLNI uses Weber’s law to identify noise in a similar way that humans recognize noise.

4. Conclusion

In this paper, we extend Chen’s IQM and propose a new IQM to remove Chen’s drawbacks by using the mean-centered WLNI and the saliency map. According to the experimental results, the proposed IQM outperforms Chen’s and agrees more consistently with human visual judgment.

Acknowledgments

This work was supported by a research grant from Seoul Women’s University (2014).

References