A VISUAL ATTENTION BASED REFERENCE FREE PERCEPTUAL QUALITY METRIC

Ali Shariq Imran, Fahad Fazal Elahi Guraya, Faouzi Alaya Cheikh
Gjøvik University College
P.O.Box-191, N-2802, Gjøvik , Norway
phone: + (47) 96822072, email: ali.imran@hig.no

ABSTRACT

In this paper we study image distortions and impairments that affect the perceived quality of blackboard lectures images. We also propose a novel reference free image quality evaluation metric that correlates well with the perceived image quality. The perceived quality of images of blackboard lecture contents is mostly affected by the presence of noise, blur and compression artifacts. Therefore, the importance of these impairments are estimated and used in the proposed quality metric. In this context there is no reference, distortion free, image; thus we propose to evaluate the image perceived quality based on the features extracted from its content. The proposed objective metric estimates the blockliness and blur artifacts in the salient regions of the lecture images. The use of a visual saliency models allows the metric to focus only on the distortions in perceptually important regions of the images; hence mimicking the human visual system in its perception of image quality. The experimental results show a very good correlation between the objective quality scores obtained by our metric and the mean opinion scores obtained via psychophysical experiments. The obtained objective scores are also compared to those of the PSNR.

Index Terms— Reference Free, Quality Evaluation, Perceptual Quality Metric, Lecture Images, Text Saliency

1. INTRODUCTION

Recent advances in e-learning technologies coupled with significant internet growth have led to the widespread use and availability of digital lecture videos [1]. Most of these images and lecture videos are still based on the use of traditional blackboard with handwritten text. Apart from the e-learning, these images and videos are widely used for the creation, storage and retrieval of multi-media learning objects [1], optical character recognition for handwritten text and for lecture video summarization and indexing. However the quality evaluation of blackboard images and lecture videos is still an untouched area.

Image quality can be measured objectively (by an algorithm) or subjectively (carried out by viewing sessions where viewers are asked for their opinion score on the image quality). Objective quality metrics can be divided into three categories namely full reference, reduced-reference and reference free metrics. In the full reference quality estimation, the distorted image is usually subtracted from the original image. Peak Signal to Noise Ratio (PSNR), Mean Absolute Error (MAE) and Mean Squared Error (MSE) are still the most widely used full reference quality metrics. In reduced-reference, features are extracted from both the reference and distorted images which are later used to quantify the degradation in the distorted image. Both of these categories of objective metrics make use of the original image as a reference. This however is not always possible, since in most cases original image is not present. Therefore, the reference free objective quality metrics are desirable. These metrics use virtual reference, i.e. they are application domain dependent and rely on finding the known coding artifacts.

In our application, excess of chalk dust can create noise in the board images during repetitive writing and erasing of text. The quality of the content is more severely degraded with the introduction of compression artifacts by the image or video encoders. Work has been done in recent years to develop systems for enhancing the visual quality of white board images [2] and documents acquired with portable digital cameras [3]. Most of these systems rely on applying various enhancement technique directly on the image to enhance the text [4, 5] without evaluating the quality of the media. Blocking and blurring are the prominent artifacts affecting the readability of the blackboard content that are introduced by compression. The initial work to estimate the blockliness was carried out by Wang and Alan [6, 7, 8]. Different techniques since then are adopted to estimate these artifacts based on edge sharpness level [9, 10], contrast similarity [11], geometric moments [12], and based on other spatial features [13, 14].

In this paper, we propose a non-reference quality metric for black-board images based on visual attention analysis. Visual attention always plays an important role in determining the perceived image quality. It is widely accepted that under normal viewing conditions, human eye tends to follow visually salient regions [15]. Here we have proposed a quality metric that gives higher weights to salient regions and lower to non-salient regions of a black-board image. Barland and
Perkis combined the attention model for saliency with perceptual quality in [16, 17].

The rest of this paper is organized as follows: Section 2 presents our quality metric that computes the quality score of a black-board image after computing the blurriness and blockliness in the salient regions of the image. Section 3 shows the experimental setup and psychophysical experiment results. In section 4, the results obtained from psychophysical experiments are compared with the existing full reference matrices. In the last section conclusions and future directions are explored.

2. PROPOSED MODEL

The simplified process flow chart for our proposed attention based reference free quality metric model is shown in Figure 1. The system takes the lecture images as an input. It then computes the text saliency, estimates the blocking and blurring artifacts on the salient regions, and gives an objective quality value. Psychophysical experiments are conducted to obtain mean opinion score. The same sets of images are used for the subjective experiments. Lastly, correlation is performed between the objective quality score obtained as a result of artifacts estimation and the mean opinion score.

2.1. Saliency Detection

In our proposed quality estimation metric, first of all we have to detect salient regions of the image for that we need a saliency detection model. Traditional bottom-up saliency detection models [18, 19] make use of the features such as color, intensity, orientation to compute the salient regions in an image. However these techniques fail when applied to blackboard images with handwritten text because of minor variation in color and intensity. So for detecting salient regions on blackboard images, that is the regions with text we have proposed a simple technique. At first we detect the horizontal and vertical edges in the image using Sobel edge detector followed by noise removal from the images using a median filter. The resultant image with edges is considered as the saliency map of the original image. The original image is shown in Figure 2(a) top left corner and the saliency map is shown in Figure 2(b) top right corner. Here we are assuming that the images contain only the inner part of the blackboard with text, so that only the blackboard text regions could be detected as salient regions. False salient regions could be detected in case the image contains the blackboard boundary edges, and edges due to presence of other objects. Figure 2(c) and 2(d) shows the $8 \times 8$ salient blocks. Further blockliness and blurriness estimation is carried out in these salient block regions.

Fig. 2. 8x8 saliency blocks representation for text detection. (a), original Image (b), Saliency map (c), Blocks representation (d) salient blocks.

2.2. Blockiness Estimation

Compression standards like JPEG and MPEG use blocks of $8 \times 8$ pixels for the DCT transform and quantization operations. This creates artifacts in the compressed images at the edges of these blocks. Here we propose to use an algorithm that detects edges at $8 \times 8$ block boundaries of the image and used to compute the degree of block processing in the image.
The extrapolated difference between adjacent blocks constituting the salient regions is calculated to estimate the blockiness effect on the perceived image quality.

Let $B_{ij}$ is an $8 \times 8$ block of pixels from the salient regions of the image.

The value of the blocking artifact $Blc_v$ across two horizontally adjacent blocks $B_{11}$ and $B_{12}$, as illustrated in Figure 3, represents a measure of the discontinuity at the vertical boundary between the two blocks. This value is computed in the following way, first the vertical discontinuity is evaluated for each line across the two blocks. This vertical discontinuity is computed as the absolute difference of the two extrapolated values, $(E^l)$ and $(E^r)$, across the boundaries of two adjacent blocks. $(E^l)$ and $(E^r)$ are calculated using first order extrapolator given as:

\[ E^l = \frac{3}{2} \cdot x_1 - \frac{1}{2} \cdot x_2 \quad (1) \]
\[ E^r = \frac{3}{2} \cdot y_1 - \frac{1}{2} \cdot y_2 \quad (2) \]

Where $x_1, x_2$ and $y_1, y_2$ are the pixel values at the boundary of the blocks as illustrated in Figure 3.

While the vertical artifact value is the mean of the eight discontinuities within a single block. Where $(E^r)_j$ is the $i^{th}$ row extrapolated values.

\[ Blc_v = \frac{1}{8} \sum_{j=0}^{7} |(E^r)_j - (E^l)_j| \quad (3) \]

Figure 3 shows the increase in the measured blocking artifacts with the addition of blockiness (due to compression) impairments in image.

The values for the horizontal artifacts can be calculated in similar fashion. A blockiness score can be estimated by summing up the vertical and horizontal blockiness artifacts.

\[ BS = Blc_v + Blc_h \quad (4) \]

Figure 4 shows the increase in the measured blocking artifacts with the addition of blockiness (due to compression) impairments in image.

2.3. Blur Estimation

Blur is usually caused by the quantization process or by the de-blocking filter. The high frequency information is associated with the detail of an image which is represented by the high frequency components. Quantization process removes such high frequency information detail from an image that results in a blurred image. Blur is actually hard to estimate accurately without reference image. One way of finding the amount of blur is to monitor the changes in the activity signal [6]. Blur is calculated across the horizontal and vertical boundaries of the $8 \times 8$ adjacent blocks. As we are only interested in the blocks that are text salient, so the occurrence of inaccurate results in regions with too complicated or too plain texture is ignored as in [20]. Local variance is used to estimate the blurriness in an image constituting the salient blocks. First the variance is calculated across the horizontal blocks and then across the vertical ones. The local variance at the first row of blocks $B_{11}$ and $B_{12}$ is given by:

\[ \sigma_{11} = \sqrt{\sum_{i=1}^{n} |x_i - y_i|} \quad (5) \]

Where $n = 2$ in the example in Figure 3. Next we compute the average of these local variances along row $j$ in the image, as follows:

\[ \Delta \sigma_j = mean\{\sigma_{ji} - \sigma_{j(i+1)}|i \in \{1, 2, ..., K\}\} \quad (6) \]

Where $K$ is the number of $8 \times 8$ blocks in the horizontal direction of the image. The sum of total blur across vertical direction is given by

\[ Blr_v = \sum_{j=1}^{N} \Delta \sigma_j \quad (7) \]
Where \( N \) is the total number of rows in the image.

The value for the horizontal blur is calculated in similar fashion.

### 2.4. Quality Prediction Score

An overall image quality value is obtained by combining the measured blockiness and blurriness from the dataset. First the average blocking and blurring values are obtained by combining the vertical and horizontal artifacts.

\[
Blc = \left( \frac{Blcv + Blch}{2} \right) \tag{8}
\]

\[
Blr = \left( \frac{Blrv + Blrh}{2} \right) \tag{9}
\]

Then the following prediction model is used to combine the artifacts

\[
QPM = 10 \times (\alpha + \delta \times Blc^a \times Blr^b) \times T^c \tag{10}
\]

The values of \( a = -0.24, b = -0.16, \) and \( c = 0.06 \) are estimated from the image dataset used to train the algorithm using a non-linear regression routine. Where \( \alpha = 0.5 \) and \( \delta = 2.356 \) are adjusted based on opinion score on training dataset. While \( T = 2 \) is a perceptual threshold obtained via the subjective test questionnaire and it represents the acceptable amount of blocking and blurring artifacts in an image. \( T \) is used to fine-tune the parameters \( \alpha \) and \( \delta \).

### 3. EXPERIMENTAL SETUP

The psychophysical experiments were conducted on a Dell 2407 wide flat panel, with monitor white point set to D65, light intensity of 120 cd/m\(^2\) (376.8 lux) and a resolution of 1920x1200 pixels with 32bits color quality. The ambient light intensity was set to 200 lux. The images were shown in their original sizes with 786 x 560 average resolutions per image. Figure 5 shows samples of the images used.

Three different categories of images were created from 7 original images consisting of 3 green board images, 2 white and 2 black. The first two categories of images were having blocking and blurring artifacts respectively. Each category consists of 5 datasets having 10 images each. While the third category consists of 3 datasets having 10 images each. Table 1 shows the datasets classification.

| Table 1. Classification of artifacts into different categories |
|---|---|---|---|
| | Artifacts Type | Sets | Images | Total Image |
| Cat. 1 | blocking | 5 | 10 | 50 |
| Cat. 2 | blurring | 5 | 10 | 50 |
| Cat. 3 | block/blur | 3 | 10 | 30 |

During the psychophysical experiment a total of 130 images were shown in random fashion to 17 subjects without the reference images. The viewers were asked to rate the quality of each image based on the ease of readability of the text on a scale of 1 to 5. Where 1 means barely readable image (highly degraded) and 5 corresponds to easily readable one. Figure 6 shows the quality rating scale.

![Fig. 6. Quality rating scale.](image)

### 4. EXPERIMENTAL RESULTS

A high correlation value of 0.92 is obtained for 130 images in total, from three different categories, when comparing the quality prediction model (QPM) score with the Mean Opinion Score (MOS). The results were obtained from 17 non-expert subjects. The prediction model was trained on 65 images from all categories of dataset. There is also a significant in-
crease in correlation from 0.81 to 0.92 when comparing MOS with PSNR and QPM respectively as seen in Figure 7 and 8.

Fig. 7. Scatter of QPM score vs. MOS of 0.927.

Fig. 8. Scatter of PSNR vs. MOS = 0.815.

Table 2 shows the average correlation per artifact for datasets containing images impaired with blocking, blurring, and with both blocking and blurring artifacts in each image. The blocking artifact is introduced in images using imwrite routine in matlab.

It is really difficult to accurately estimate the amount of blurring artifact without the presence of reference image. For this reason in some images having blurring artifacts only, PSNR shows slightly better results. For the rest of the images in each category, QPM shows very high correlation. The results are shown in Table 3. In images with both the artifacts, the presence of one artifact masks the other. This is true for category 3 dataset images. Where, images were first introduced with blurring artifacts and then with the blocking. This results in the masking of blurring artifact by blocking. i.e. the blurring artifacts tends to fadeout. For these kind of images as well, QPM show significance increase in correlation with MOS than PSNR.

Table 2. Average correlation per artifacts.

<table>
<thead>
<tr>
<th></th>
<th>Category 1</th>
<th>Category 2</th>
<th>Category 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>blocking</td>
<td>blurring</td>
<td>blocks/blur</td>
</tr>
<tr>
<td>PSNR</td>
<td>0.837</td>
<td>0.752</td>
<td>0.853</td>
</tr>
<tr>
<td>QPM</td>
<td>0.972</td>
<td>0.842</td>
<td>0.932</td>
</tr>
</tbody>
</table>

Table 3. Individual dataset correlation coefficient

<table>
<thead>
<tr>
<th>Dataset No.</th>
<th>Blocking</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PSNR</td>
<td>QPM</td>
<td>PSNR</td>
<td>QPM</td>
</tr>
<tr>
<td>1</td>
<td>0.77</td>
<td>0.94</td>
<td>0.67</td>
<td>0.65</td>
</tr>
<tr>
<td>2</td>
<td>0.63</td>
<td>0.95</td>
<td>0.61</td>
<td>0.59</td>
</tr>
<tr>
<td>3</td>
<td>0.89</td>
<td>0.98</td>
<td>0.81</td>
<td>0.80</td>
</tr>
<tr>
<td>4</td>
<td>0.54</td>
<td>0.91</td>
<td>0.59</td>
<td>0.87</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>0.98</td>
<td>0.71</td>
<td></td>
</tr>
</tbody>
</table>

5. CONCLUSION AND FUTURE WORK

In this paper we studied different image distortions and impairments that affect the perceived quality of blackboard lectures images. And conducted a psychophysical tests where users were asked to rate the perceived quality of the impaired images. The test images were extracted from lecture videos stored in compressed format and thus suffered from different types of impairments introduced by the compression process. The MOS was calculated for the different types of impairments in the training dataset and used to develop a reference free objective quality metric. This metric was shown to correlates well with the subjective score obtained for the test dataset. From the experiments it was observed that people mostly focus on the text rather than the background. To exploit this, the QPM relies on text salient regions around which it estimates the degrees of blur and blockiness, which makes it correlate better with the Human Visual System. Comparison is made with the PSNR as it is still widely accepted quality metric when video compression is involved.

Even though the initial results obtained in this work are very encouraging, to adopt a metric for lecture images obtained from lecture videos in general is not easy. This is due to the sensitivity of the overall image quality for blackboard image to the text size, orientation and handwriting style. Moreover quality degradations in some image regions may be less noticeable due to uniform background with less text. Therefore, further studies are needed to improve the stability of the proposed quality metric and to cover different types of videos.
6. REFERENCES


