A locally adaptive algorithm for measuring blocking artifacts in images and videos


Institute for Infocomm Research, 21 Heng Mui Keng Terrace, Singapore 119613, Singapore

Received 27 October 2003; accepted 14 April 2004

Abstract

Block transform coding is the most popular approach for image and video compression. The objective measurement of blocking artifacts plays an important role in the design, optimization, and assessment of image and video coding systems. This paper presents a new algorithm for measuring image quality of a BDCT coded images or videos. It exhibits unique and useful features: (1) it examines the blocks individually so that it can measure the severity of blocking artifacts locally; (2) it is a one-pass algorithm in the sense that the image needs to be accessed only once; (3) it takes into account the blocking artifacts for high bit rate images and the flatness for the very low bit rate images; (4) the quality measure is well defined in the range of 0–10. Experiments on various still images and videos show that the new quality measure is very efficient in terms of computational complexity and memory usage, and can produce consistent blocking artifacts measurement.

Keywords: Blocking artifacts; Quantitative impairment metric; Blockiness measure; Image coding

1. Introduction

Block transform coding is the most popular approach for image and video coding. Most of the current image and video coding standards, such as JPEG, H.26x and MPEG-1/2/4, make use of the block-based discrete cosine transform (BDCT) [4,6]. In BDCT coding, DCT coefficients are calculated over small non-overlapping blocks, and the 2-D blocks of transform coefficients are then quantized. In the decoder, the quantized transform coefficients are de-quantized, and inverse transformed to recover the original data (with certain distortion). At low bit rate image and video coding, a large quantization value is used. Therefore, the decompressed image and video exhibits various kinds of distorted artifacts. One of the most noticeable artifacts is the “blocking artifact,” which is an artificial discontinuity between adjacent blocks and is a direct result of the independent quantization of the BDCT coefficients [9–11]. However, blocking artifacts are structural disturbance, and are sometimes “buried” in the massively accumulated across-the-board pixel-wise error. Therefore their
significance in perceptual visual quality assessment is not reflected correctly in the conventional PSNR (peak signal-to-noise ratio) measure. An effective measure of blocking artifacts can be used as a quality metric alone [10] or a factor in quality evaluation [2,3,7,8,11,12].

There are generally two types of objective image and video quality metrics, i.e., referenced and un-referenced approach. In a referenced approach, the access to original images is required. By using the original and reproduced image as inputs, the system outputs a numerical value that quantifies the visibility of blocking artifacts in the reproduced image [2]. This approach is not very useful in applications such as image and video communication, where original image and video is not accessible. On the other hand, the un-referenced approach [9,10] is of more interests because of its computational efficiency and wider scopes of potential applications, including in-service visual quality monitoring and post-processing for decoded signal. However, designing un-referenced objective quality metrics is very difficult due to the limited understanding of the human vision system (HVS). It is believed that effective un-referenced objective quality metrics are only feasible when the prior knowledge about the image distortion types is available [9–11].

The generalized block-edge impairment metric $M_{GBIM}$ proposed in [10] evaluates the visual significance of block-edge artifacts in a given image by taking into account the luminance masking effects in extreme bright/dark areas in a reconstructed image. However, this algorithm is mathematically ill defined such that for certain images, the denominator used to normalize the metric becomes zero. Also, the value range of this metric is not well defined. Wang [9] has proposed an un-referenced perceptual quality assessment of JPEG compressed images based on the measures of blockiness and image signal activity, and these two measures are then combined in a model whose parameters are estimated from the subject test data. Again, the value range of this metric is not well defined. As the choice of their model is empirical, and because the model parameters are derived from the training data, one may worry about its robustness—its performance on the unknown data because these data may be rather different from the training data statistically.

In this paper, a new algorithm is proposed to calculate the blocking artifacts in BDCT-coded image and video. Compared with the existing algorithms [9,10], this new method exhibits unique features: (1) it examines block by block to individually decide if it has blocking artifacts. This is useful in detecting blocking artifacts locally, since each macro block in the image may have different quantization steps and the same extent of blockiness results in different degree of disturbance in human eyes due to the masking effect; (2) it is a one-pass algorithm in the sense that the image needs to be accessed only once and this makes it very suitable for real time implementation; (3) It takes into account the blocking artifacts for high bit rate images and the flatness for the very low bit rate images; (4) The measure is well defined in the range of 0–10. Experiments on various still images and videos show that the new measure can produce consistent prediction of blockiness.

The rest of the paper is organized as follows. Section 2 presents the one-pass algorithm to detect the blocking artifacts at the block level and at the image level. Section 3 presents the experimental results, and its comparison with some of the existing schemes. Section 4 provides the conclusions of the paper.

2. Algorithm description

As mentioned in the Introduction, the blocking artifacts manifest itself as an artificial discontinuity between neighboring blocks. However this is only true at high bit rate. At very low bit rate, this artificial discontinuity becomes less severe and many blocks might even merge together to form a large uniform region. Therefore flatness across the block boundary is another important attribute to the overall effect of blocking artifacts. Furthermore, the independent processing of these blocks that does not take into account the between-block pixel correlations tells that the blocking artifacts is due to the local behaviors of the block and its neighboring ones, i.e., the blocking artifacts can be detected locally.
2.1. One-pass blockiness measure of a single block

Fig. 1(a) shows an example of $8 \times 8$ block and its neighboring blocks. In order to decide whether Block A has blocking artifacts, it is sufficient to check its relationship with the neighboring blocks, i.e., Block B, C, D and E. However, in the circumstance of real time image processing or video coding, only Block B and C are available at the time of processing Block A. Extensive experiments on test images and video sequences have shown that there are approximately similar amount of discontinuity between Block A and Blocks B, C, and that between Block A and Blocks D, E, since the discontinuity is caused by the high quantization value during the coding process where all the blocks are subject to similar distortions. Therefore it is reasonable to quantify the blocking artifacts of Block A by just examining the inter-pixel discontinuity between Block A and its neighbors Blocks B and C. Note that the other two borders of Block A are to be considered when blocking artifacts are identified for Block D and Block E.

As is denoted in Fig. 1(b), we define the horizontal inter-block difference $B_h$ between Blocks A and B as

$$B_h = \begin{cases} N_h & \text{if } D_h \neq 0, \\ D_h & \text{if } D_h = 0, \end{cases}$$

where, $N_h$ and $D_h$ are defined as

$$N_h = \gamma_1 \times \sum_{i=1}^{8} |a_{i1} - b_{i8}|,$$

$$D_h = \gamma_2 \times \sum_{i=1}^{8} \left( \sum_{j=5}^{7} |b_{(i+1)j} - b_{ij}| + \sum_{j=1}^{3} |a_{(i+1)j} - a_{ij}| \right) + \sum_{j=1}^{8} |a_{i1} - b_{i8}|.$$

It should be noted that,

1. The weighting coefficients $\gamma_1$ and $\gamma_2$ in Eqs. (2) and (3) are chosen such that the value of $B_h$ is defined within the range of 0–10, i.e., $\gamma_1 = 10$ and $\gamma_2 = 1.5$. This can be seen from Eqs. (1)–(3) that, $B_h$ equals to 0 if and only if all the terms in Eq. (3) are equal to 0, and $B_h$ equals to 10 if and only if the first term in Eq. (3) is 0.

2. $B_h \leq 1$ indicates that there is no inter-block discontinuity horizontally. It happens to the image where there is either a constant difference in the horizontal direction or the inter-pixel differences of cross-block pixels are small than that of the in-block pixels; Similarly, $B_h = 10$ indicates that the horizontal inter-block discontinuity is most severe, which happens when the inter-pixel differences of cross-block pixels are none-zero while that of the in-block pixels are all zero.

3. $D_h$ in Eq. (3) is the weighted average of the inter-pixel differences at and near the block boundary, and serves as the normalization factor. This average covers only half of Block B and half of Block A. The small area of average could reduce the amount of calculation involved, and avoid the influence of possible texture changes that are far away from the block boundary—this is in line with the human eyes’ characteristics indicating that texture masking only occurs in the proximity of the masker.

The vertical inter-block difference $B_v$ between Block A and Block C are defined in the similar way. Therefore the measure for blocking artifacts of an $8 \times 8$ block can be summarized as,

$$B_{BLK} = \frac{B_h + B_v}{2}.$$

Fig. 1. (a)–(b) $8 \times 8$ block and its neighboring blocks.
Here we assume that the sensitivity of HVS to horizontal and vertical blocking artifacts are similar.

The measure, $B_{BLK}$, of Eq. (8) tells the severity of blocking artifacts of an individual $8 \times 8$ block. Its value ranges from 0 to 10. The higher the value of $B_{BLK}$ is, the greater the severity of the blocking artifacts.

2.2. One-pass flatness measure of a single block

In an image coded at very low bit-rate, many blocky blocks would merge to form large flat regions with uniform gray level. The blockiness across block boundaries is zero because there is no inter-pixel gray level difference. However, human eyes are not so comfortable with this phenomenon which contributes significantly to the distortion of the image. To design a quality assessment algorithm, we have to count in this factor.

The flatness is measured in terms of the proportion of zero crossings from pixel to pixel locally. That means only the $8 \times 8$ region near the block boundary is considered, ranging from four pixels to the left of the boundary to four pixels to the right of the boundary. The total number of zero-crossings is obtained and divided by the total number of crossings to give the flatness measure in that particular region.

Defining a function for zero crossing as

$$z(x, y) = \begin{cases} 1 & \text{if } |x - y| = 0, \\ 0 & \text{otherwise}. \end{cases}$$

Then,

$$Z_h = \frac{10}{56} \sum_{i=1}^{8} \left( \sum_{j=5}^{7} z(b_j, b_{j+1}) + \sum_{j=1}^{3} z(a_j, a_{j+1}) \right)$$

$$+ \frac{10}{56} \sum_{i=1}^{8} z(a_{i1}, b_{i8}),$$

where $Z_h$ is the horizontal inter-block flatness measure across blocks A and B, and 56 is the total number of crossings across different pixels, and 10 is the scale factor such that $Z_h$ ranges from 0 to 10, with 10 to be the worst case. The vertical inter-block flatness measure $Z_v$ could be obtained similarly. The average of the two is used to estimate the flatness near the region of block A.

$$Z_{BLK} = \frac{Z_h + Z_v}{2}. \quad (7)$$

2.3. Measuring blocking artifacts of an image

The overall measure of blocking artifacts for an image block is decided by the dominating values of $B_{BLK}$ and $Z_{BLK}$, by taking into account of the local contrast masking and spatial masking as follows:

$$Q_{BLK} = \begin{cases} \max(B_{BLK}, Z_{BLK}) & \text{if } \max(B_{BLK}, Z_{BLK}) > T_{JIND}, \\ 0 & \text{otherwise}. \end{cases} \quad (8)$$

Note that $T_{JIND}$ is a local pixel activity threshold decided by the average local just noticeable distortion (JND) as defined in [12], and

$$T_{JIND} = \lambda \times \text{average (JND}_{fb}(i,j))$$

and the average is taking place among all the pixels $a_{ij}$, $b_{ij}$ and $c_{ij}$ as in Fig. 1. $T_{JIND}$ will help to mask those $B_{BLK}$ and $Z_{BLK}$ that are below the noticeable values. $\lambda$ is a scale factor which is chosen to make the measure $Q_{BLK}$ to correlate well with subjective test results. We have tested our algorithm based on subjective video quality experiments similar to that conducted for the evaluation of the JVT sequences [1]. Here, $\lambda$ is set to 0.05 which provides the best subjective–objective correlation.

Therefore the quality measure of an image can be easily decided by the average of $Q_{BLK}$ of all the blocks in the image.

$$Q_{IMAGE} = \text{average (} Q_{BLK} \text{).} \quad (10)$$

3. Experiments

In the following experiments, we used a number of still images, as well as frames from the test video sequences. These images have different resolutions, ranging from $176 \times 144$ to $720 \times 480$ to $512 \times 512$. We also compared our results with those from other objective quality metrics such as
PSNR, the quality metrics $M_{\text{GBIM}}$ of [10], and the NR quality metrics $S$ of [9]. In order to plot all these metrics in the same figure, we scale PSNR by dividing a factor of 5. Note that according to [10], there is no defined value range $M_{\text{GBIM}}$, but the higher the $M_{\text{GBIM}}$ value is above one, the greater the severity of the blocking effect. On the other hand, according to [9], the smaller the $S$ is, the greater the severity of the blocking effect is. Experiments on different images show that $S$ ranges from $-246$ to 13 and $M_{\text{GBIM}}$ ranges from 0.98 to $\infty$.

Fig. 2 shows an example of such comparison. Note that in these figures, ‘$Q_{\text{IMAGE}}$’, ‘$M_{\text{GBIM}}$’, ‘$S$’ and ‘PSNR’ represent the quality metric of this paper, blockiness measure of [10], the NR quality metrics of [9] and PSNR, respectively. From the figure we can see that, our blockiness measure $Q_{\text{IMAGE}}$ has a very well defined range of values. Also our blockiness measure responds to the increase of quantization value faster than the rest, which makes it more suitable in controlling the coding quality in real-time.

Fig. 3 shows the experimental results of applying $Q_{\text{IMAGE}}$ to different images with different spatial complexity and resolution. It can be shown from these results that in general, $Q_{\text{IMAGE}}$ achieves consistent results for all these images. However, $Q_{\text{IMAGE}}$ is slightly higher for the image from the test sequence ‘Garden’ at high bit rate (bbp > 2.5). This is due to the fact that in that image, the rooftop, and the grass show the similar structure as small blocks. This coincides also with subjective test of the same sequence because the area of rooftop and the grass in the test sequence ‘Garden’ appears “blocky” even before the actual blocking artifacts are introduced during coding process. Similarly, $Q_{\text{IMAGE}}$ is slightly lower for the image from the test sequence ‘Coastguard’, which is caused by the strong spatial masking effect of the water area. However, at very low bit rate (bbp < 0.5), $Q_{\text{IMAGE}}$ is high for ‘Mobile’, which is again in line with human eyes’ characteristics indicating there is little texture masking in ‘Mobile’ due to the many uniform areas in that image. In the worst cases, the maximum differences of $Q_{\text{IMAGE}}$ for all the images and video sequences we have tested are less than 1. (Table 1)

Figs. 4 and 5 show an original frame of test sequence ‘Akiyo’ and its reconstructed frame after BDCT coding. From these two images, one can see that Fig. 5 has very obvious degradation compared
to Fig. 4. This is reflected by our quality measure $Q_{\text{IMAGE}}$ as shown in Table 2, where Fig. 4 scores 0.886 and Fig. 5 scores 2.885. However by using [9,10] metrics, these two images are both considered of very good quality. Therefore our quality measure can truly reflect the actual degradation of an image and can respond to the quality degradation faster, which makes it more suitable in controlling the coding quality in real-time coding. Fig. 6 shows another example of non-BDCT coded image “Books”. As this image has not gone through any BDCT based coding, there are no visible blocking artifacts in this image; however, the rating by [9] is $-245.891$, which is obviously an outlier point as their training data do not cover this type of images. Also, by comparing the perceptual quality of Figs. 5 and 6, one can easily tell that Fig. 6 has a better perceptual quality than Fig. 5, but [10] gives the similar rating. However our proposed quality measure can rate them more accurately compared to the subjective test.

Table 2 shows the Pearson correlation between various quality metrics and the subjective ratings of the JPEG database provided by LIVE [5]. It can be seen that our metrics has a similar correlation
value with the subjective data compared to the other two metrics. Note also that the difference in the correlation values among all three metrics are within the tolerance range of the subjective data, and the robustness of the JPEG database provided by LIVE [5] has yet to prove. On the other hand, the advantages of our algorithm, such as locally adaptive, fast response to blocking artifacts, and suitable for real-time implementation, etc., have made it more attractive in real-life applications.

4. Conclusions

This paper presents an effective and efficient un-referenced approach for measuring blocking artifacts in BDCT coding, especially in low bit-rate cases. The novelty of the algorithm is the adaptive determination of blockiness via block-by-block analysis, addressing the effects caused by different quantization steps in neighboring blocks and HVS’s luminance and spatial masking characteristics. A very important feature of the proposed measure is that it captures both the blockiness and flatness of BDCT-coded image, where most of the existing algorithms focus on the blockiness only. An additional advantage of the proposed measure is its well-defined value range and very simple computationally.

Experiments show that the locally adaptive measure exhibits consistent results with various types of images and frame in video sequences. The new measure can be used as a quality metric, a contributing factor in a metric, or a parameter controlling an encoding/post-processing process in real-time, in both referenced and un-referenced situations.

Table 2
Pearson correlation and Spearman for LIVE database

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Pearson correlation</th>
<th>Fitted Pearson correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{Q}_{\text{IMAGE}}$</td>
<td>-0.921</td>
<td>0.923</td>
</tr>
<tr>
<td>$M_{\text{GRIM}}$ [10]</td>
<td>-0.794</td>
<td>0.943</td>
</tr>
<tr>
<td>$S$ [9]</td>
<td>0.953</td>
<td>0.976</td>
</tr>
</tbody>
</table>

Fig. 5. Reconstructed frame of Fig. 4 after BDCT coding.

Fig. 6. Another example of non-BDCT-coded image “Books”.
References


