Journal of Zhejiang University SCIENCE A
ISSN 1009-3095 (Print); ISSN 1862-1775 (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

A no-reference blocking artifact metric for B-DCT video*

YANG Fu-zheng[†1], WAN Shuai[2], CHANG Yi-lin[1], LUO Zhong[3]

([1]*National Key Lab on ISN, Xidian University, Xi'an 710071, China*)

([2]*Multimedia and Vision Research Laboratory, Electronic Engineering Department Queen Mary, University of London, London E1 4NS,UK*)

([3]*Huawei Technology Limited Cooperation, Shenzhen 518057, China*)

[†]E-mail: fzhyang@mail.xidian.edu.cn

Received Nov. 30, 2005; revision accepted Feb. 17, 2006

Abstract: A new no-reference blocking artifact metric for B-DCT compression video is presented in this paper. We first present a new definition of blocking artifact and a new method for measuring perceptive blocking artifact based on HVS taking into account the luminance masking and activity masking characteristic. Then, we propose a new concept of blocking artifact cluster and the algorithm for clustering blocking artifacts. Considering eye movement and fixation, we select several clusters with most serious blocking artifacts and utilize the average of their blocking artifacts to assess the total blocking artifact of B-DCT reconstructed video. Experimental results illustrating the performance of the proposed method are presented and evaluated.

Key words: Blocking artifact cluster, Blocking artifact metric, HVS, Video quality assessment
doi:10.1631/jzus.2006.AS0095 **Document code:** A **CLC number:** TN919.8

INTRODUCTION

The block-based discrete cosine transform (B-DCT) scheme is a fundamental component of many image and video compression standards including JPEG, H.263, MPEG-1, MPEG-2, MPEG-4, H.264, and others, used in a wide range of applications. The coding algorithm exploits the local spatial redundancies of input video to achieve a low bit rate, which may cause visible coding distortions in reconstructed sequences due to the lost coding nature, such as blocking artifact, ringing, mosquito effect, MC mismatch, blurring, and color bleeding (Yuen and Wu, 1998). Among B-DCT digital-video coding distortions, blocking artifact is of particular importance. As shown in (Karunasekera and Kingsbury, 1995), blocking artifact and its propagation through reconstructed video sequences, are the most significant of all coding artifacts. The blocking artifact thus represents a major type of distortion in the reconstructed video compressed with B-DCT algorithm, which will be particularly annoying in the case of high compression ratio.

A great deal of research has been carried out measuring blocking artifact. Karunasekera and Kingsbury introduced a distortion measure of blocking artifact based on human visual system (HVS) (Wu, 1995). And a perceptual blocking distortion metric is introduced which is based on more complex HVS models (Yu *et al*., 2002). These quantitative measures require both the original and reconstructed images. However, in the absence of the original image, the above distortion measures cannot be used to evaluate the quality of a reconstructed image, which makes them impractical for real time transmission applications. People have long been investigating no-reference blocking artifact metrics (Vlachos, 2000; Gao *et al*., 2002; Triantafyllidis *et al*., 2002). However, HVS should be taken into account in the measure of blocking artifact. A no-reference measure of blocking artifact is also proposed with some inadequate models of HVS (Wu and Yuen, 1997).

In this paper, a no-reference blocking artifact metric is presented which does not need reference

* Project (No. YJCB2003017MU) supported by Huawei Technology Fund, China

image at all. We present a new definition of blocking artifact and a new method for measuring the perceptive blocking artifact based on HVS. Considering eye movement and fixation, we also introduce a new concept of blocking artifact cluster, and propose a no-reference blocking artifact metric for B-DCT reconstructed video based on it.

The rest of this paper is organized as follows. In Section 2, the perceptive blocking artifact measure is proposed. Section 3 presents the new concept of blocking artifact cluster and the new blocking artifact metric for B-DCT reconstructed video. Experimental results given in Section 4 evaluate the performance of the metric. Finally, conclusions are drawn in Section 5.

PERCEPTIVE BLOCKING ARTIFACT

Since blocks of pixels are treated as single entities and coded separately, correlation among spatially adjacent blocks is not taken into account in coding, which results in block boundaries being visible when the decoded image is reconstructed. The blocking artifact is commonly defined as the discontinuities at the boundaries. However, the term can be generalized to mean an overall difference between adjacent blocks. In experiments, we observed that the block is especially annoying to human eyes when it is discontinuous with its spatially adjacent blocks at more than one boundary. Here, we call it mosaic block. Therefore, we define the blocking artifact of a block as the total discontinuity of its four boundaries with the adjacent blocks in this paper. Then the blocking artifact $d_{i,j}$ of block $B_{i,j}$ is defined by

$$d_{i,j} = \sum_{l=1}^{4} d_{l,i,j}, \tag{1}$$

where $d_{l,i,j}$ (l=1, …, 4) represent the discontinuities at $B_{i,j}$'s four boundaries with spatially adjacent blocks $B_{i,j-1}$ (left), $B_{i,j+1}$ (right), $B_{i-1,j}$ (upper) and $B_{i+1,j}$ (lower) respectively. $d_{l,i,j}$ is called "edge artifact" in this paper.

The human observer is the end user of most image information. Therefore, human perception should be taken into account in the measure of blocking artifact. The models of HVS generally consist of opposite-colors space, contrast sensitivity, multi-resolution architecture, masking, pattern adaptation, pooling and cognitive processes (Winkler, 1999). However, the human visual system is extremely complex, and many of its properties are not well understood even today.

In this paper, straightforward but effective HVS models are used to get the perceptive blocking artifact measure. Because the multichannel process is very complex and the contrast gain control at interchannel is not very clear, most attention is focused on background luminance (brightness) masking and activity masking in this paper. Human vision has lower acuity in the color pathway. Winkler (2000) conducted research on the influence of the choice of color space on the performance of models. It has been shown that it is possible for the vision models to work on luminance component only without a dramatic degradation in prediction accuracy. Therefore, the metric introduced in this paper uses only the luminance component. Take the perceptive edge artifact $d_{1,i,j}$ for example, we will introduce in detail how to obtain the edge artifact with consideration of HVS in the following.

Original edge artifact

Let us consider an image F (size of $L\times C$) which consists of K blocks, each of size $N\times N$. A pixel luminance value at (n,m), $0\le n,m\le N-1$, in the block $B_{i,j}$ ($0\le i<L/N$, $0\le j<C/N$) is denoted by $f_{i,j}(n,m)$.

The original edge artifact $ds_{1,i,j}$ of $B_{i,j}$ at the left boundary with $B_{i,j-1}$ is defined by

$$ds_{1,i,j} = \begin{cases} \dfrac{1}{N}\left|\sum_{n=0}^{N-1} f_{i,j}(n,0) - \sum_{n=0}^{N-1} f_{i,j-1}(n,N-1)\right|, & j>0, \\ 0, & j=0. \end{cases} \tag{2}$$

Luminance masking

The response of the human visual system depends much less on the absolute luminance than on the relation of its local variations from the surrounding luminance. Unfortunately, a common definition of contrast suitable for all stimuli does not exist. We use Weber contrast here that can be expressed as

$$C_{\mathrm{W}} = \Delta L / L. \tag{3}$$

Weber contrast is most appropriate for patterns consisting of a single increment or decrement ΔL to an otherwise uniform background luminance L, and is suitable for masking edge artifact. Contrast sensitivity remains nearly constant over an important range of intensities. Then the luminance masking of edge artifact is modelled by

$$db_{1,i,j} = \frac{ds_{1,i,j}}{1+\left(\frac{2|b-b_0|}{b_0}\right)^{r_1}}, \tag{4}$$

where $db_{1,i,j}$ denotes the edge artifact masked with luminance, parameter r_1 is estimated experimentally, and b is the local background luminance determined as follows.

$$b = \frac{b_{i,j} + b_{i,j-1}}{2}, \tag{5}$$

where $b_{i,j}$ is the luminance of $B_{i,j}$

$$b_{i,j} = \frac{1}{N^2}\sum_{n=0}^{N-1}\sum_{m=0}^{N-1} f_{i,j}(n,m). \tag{6}$$

Because HVS has the characteristic of adaptation to luminance, b_0 is the average luminance of image F.

$$b_0 = \frac{1}{K}\sum\sum b_{i,j}. \tag{7}$$

Activity masking

The visibility of an edge artifact will be affected by a surrounding spatial region of limited extent. For example, the edge artifact is more noticeable in smooth areas than that near edges. Therefore, the edge artifact is also masked by the activity in the background. The model of activity masking is given by

$$d_{1,i,j} = \frac{db_{1,i,j}}{a_0 + (m/m_0)^{r_2}}. \tag{8}$$

where m_0 is the average activity, parameters a_0 and r_2 are selected to match the variation of human visual sensitivity with the background activity, and m is the sum of weighted activities of $B_{i,j}$ and $B_{i,j-1}$.

$$m = 0.5m_{i,j} + 0.5m_{i,j-1}, \tag{9}$$

where $m_{i,j}$ is the activity of the corresponding block $B_{i,j}$. Strongest masking occurs when the interacting stimuli have similar characteristics, i.e., similar frequencies, orientations, etc. Therefore, for vertical edges, the activity $m_{i,j}$ is defined by

$$m_{i,j} = \sqrt{\frac{1}{N}\sum_{m=0}^{N-1}\left(\sum_{n=0}^{N-1} f_{i,j}(n,m) - Nb_{i,j}\right)^2}. \tag{10}$$

After the original edge artifact $ds_{1,i,j}$ was masked by luminance and activity, we obtain the edge artifact $d_{1,i,j}$ that is consistent with human vision characteristics. $d_{2,i,j}$, $d_{3,i,j}$, $d_{4,i,j}$ can be obtained in a manner similar to that for $d_{1,i,j}$. For the horizontal edge artifacts $d_{3,i,j}$ and $d_{4,i,j}$, activity $m_{i,j}$ is defined by

$$m_{i,j} = \sqrt{\frac{1}{N}\sum_{n=0}^{N-1}\left(\sum_{m=0}^{N-1} f_{i,j}(n,m) - Nb_{i,j}\right)^2}. \tag{11}$$

BLOCKING ARTIFACT METRIC OF IMAGE

In experiments, we also find that several adjacent blocks with serious blocking artifacts have a stronger impact on the subjective quality of the picture than the dispersed ones do, as shown in Fig.1. We consider those adjacent blocks with serious blocking artifacts as a blocking artifact cluster (BAC). Then the blocking artifact of a picture is calculated by the average of blocking artifacts of several BACs in the picture.

Fig.1 The impact of the BAC on image subjective quality

Classical approaches in computer vision consider the task of vision as the construction of a detailed representation of the physical world, obtained by looking at the input images as static entities, to be processed in a passive way. However, it has been shown that only a small fraction of the information can reach levels of processing at any given time. The understanding of the whole image relies on eye movement and fixation. Whereas an average of three eye fixations per second generally occurs during active looking, the optimal spatial collapsing normally involves some form of worst case processing (Privitera and Stark, 2000; Boccignone *et al*., 2003). This is because localized impairments tend to draw the focus of the viewer, making the worst part of the picture the predominant factor in the subjective quality decision. Accordingly, we select several BACs with serious blocking artifacts to determine the total blocking artifact of the whole image.

The proposed algorithm to measure the total blocking artifact of the whole image is as follows:

(1) All the blocks of the image $B_{i,j}$ ($0\le i<L/N$, $0\le j<C/N$) are ordered by their blocking artifact values calculated by Eq.(1) to form an ascending sequence $B^{(0)}$, $B^{(1)}$, …, $B^{(K-1)}$, in which $d^{(k)}$ is the blocking artifact of the kth block $B^{(k)}$. So $d^{(k)}\ge d^{(l)}$ when $0\le j<l<K$, where K is the total number of blocks in the image.

(2) Set a threshold N_{T1}, where $N_{T1}\le K$. In the block set of ($B^{(0)}$, $B^{(1)}$, …, $B^{(N_{T1}-1)}$), the adjacent blocks are clustered to form BACs by the following algorithm:

(a) Search for a block $B^{(i)}$ ($0\le i<N_{T1}-1$), which has the maximum blocking artifact and does not belong to any BACs. Regard it as the core of a new BAC C_m.

(b) Search for blocks that are spatially adjacent to C_m and do not belong to any other BACs, from which the block $B^{(i)}$ ($0\le i<N_{T1}-1$) with the maximum blocking artifact is selected. Then it is clustered with C_m.

(c) Repeat Step (b) until no more blocks can be clustered with C_m or the number of blocks in C_m is larger than a threshold N_{T2}.

(d) Repeat Steps (a), (b) and (c) until all the blocks in $B^{(0)}$, $B^{(1)}$, …, $B^{(N_{T1}-1)}$ are clustered.

(3) We suppose that C_m consists of M blocks $B^{(l_0)}$, $B^{(l_1)}\cdots$, $B^{(l_{M-1})}$. Then the blocking artifact of C_m is defined as

$$dc_m = M^{-\frac{2}{3}}\left(\sum_{s=0}^{M-1} d^{(l_s)}\right). \tag{12}$$

(4) The blocking artifacts of BACs are ordered to form an ascending sequence $dc^{(0)}$, $dc^{(1)}$, $dc^{(2)}$, $\cdots$, from which the first N_{T3} BACs ($dc^{(0)}$, $dc^{(1)}$, $\cdots$, $dc^{(N_{T3}-1)}$) are selected. We define the average blocking artifacts of the first N_{T3} BACs as the blocking artifact of the whole image.

$$FB = \frac{1}{N_{T3}}\sum_{n=0}^{N_{T3}-1} dc^{(n)}. \tag{13}$$

Then the blocking artifact metric (BAM) to evaluate the perceptive blocking artifact of the video sequence is calculated as the mean value of the blocking artifacts (FBs) through the entire video sequence.

$$BAM = \frac{1}{N_F}\sum FB_i, \tag{14}$$

where N_F is the total number of images in the video.

EXPERIMENTAL RESULTS

In our experiments, the “Foreman”, “Claire”, “Carphone” and “News” sequences (CIF) are compressed using H.263 Intra-mode with QP from 5 to 25. The model parameters are set as follows. In Eq.(4), r_1 is set to 2, because the luminance (ΔL and L) are measured in cd/m^2 in Eq.(3), whereas b and b_0 in Eq.(4) refer to the luminance measured in grey levels. In Eq.(8), r_2=1.4, a_0=0.3. They are determined by the training sequence “Missa” and “Paris”. When the viewing distance is approximately four times the size of the image, the horizontal visual angle of a block in a CIF image is 2arctan[1/(44×4×2)]≈0.65°. And the entire fovea covers approximately 2° of visual angle. Therefore, in the clustering process, N_{T1}= 15%K and N_{T2}=2%K. N_{T3}=5 because an average of three eye fixations per second generally occurs during active looking. The reconstructed sequences are assessed with subjective evaluation and the blocking artifact metric proposed in this paper respectively.

Single stimulus method (SSM) (ITU-T Recommendation BT.500-10, 2000) with five-grade impairment scale was used in our subjective test, in which 20 non-expert viewers were asked to evaluate the blocking artifacts of all the video sequences distributed randomly. The subjective mean opinion scores (MOS) S_i are linearly scaled to a nominal range of [0, 1], where zero represents the best rating and one represents the worst rating. The scaled subjective mean opinion scores $\overline{S}_i$ are given by

$$\overline{S}_i = \frac{S_i - S_{\text{best}}}{S_{\text{worst}} - S_{\text{best}}}. \tag{15}$$

In order to demonstrate that the BACs have a stronger impact on the subjective quality of the picture than those dispersed block artifacts do, we introduce the mean perceptive blocking artifact metric (MBAM) for comparison. The MBAM is calculated as the average value of the perceptive blocking artifacts in a video sequence, which is given by

$$MBAM = \frac{1}{N_F}\sum\left(\frac{1}{K}\sum_i\sum_j d_{i,j}\right), \tag{16}$$

where the perceptive blocking artifact $d_{i,j}$ can be obtained by the methods described in Section 2. N_F and K respectively represent the number of images in the sequence and the number of blocks in the image.

Next, linear polynomial fit is used for the objective MBAM scores and BAM scores.

$$\overline{MBAM}_i = a_1 MBAM_i + b_1, \tag{17}$$

$$\overline{BAM}_i = a_2 BAM_i + b_2. \tag{18}$$

We utilize the accuracy evaluation metrics suggested by the Video Quality Experts Group (VQEG) (VQEG Report, 2003) to evaluate the performance of the MBAM and the proposed BAM. The root-mean-squared error (RMSE), Pearson correlation coefficient (PCC), and Spearman rank order correlation coefficient (SCC) are computed between fitted objective data and corresponding subjective data. The results are shown in Table 1. As shown experimentally, the MBAM and the BAM proposed in this paper all achieve good agreement with the MOS. Furthermore, BAM has a lower RMSE, which indicates a lower prediction error compared with MBAM. And the small increase in Pearson correlation and Spearman correlation indicates that BAM has a better performance in prediction monotonicity than MBAM.

Table 1 Performance comparison of the objective data and subjective data

	RMSE	*PCC*	*SCC*
MBAM	0.0655	0.9129	0.9261
BAM	0.0205	0.9524	0.9446

The scatter plots of MOS versus the scaled objective scores MBAM and BAM are shown in Figs.2a and 2b respectively. The scatter plots demonstrate that the BAM supplies remarkably good prediction of the subjective scores and is more consistent with subjective quality than MBAM.

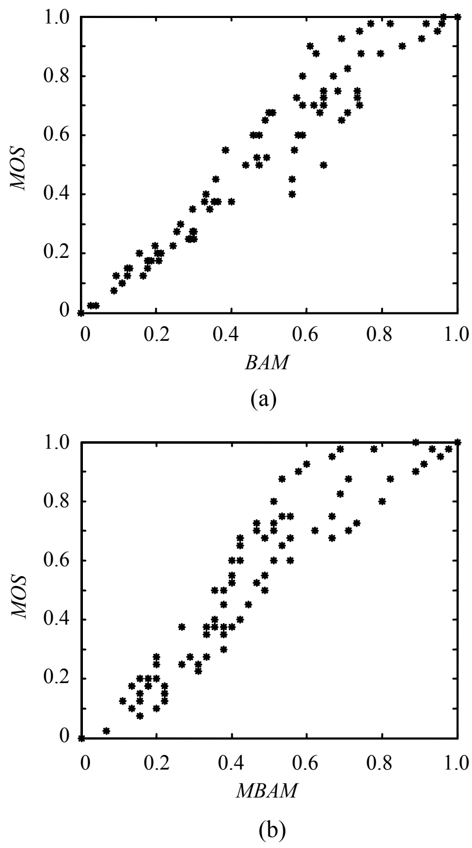

Fig.2 Scatter plot of MOS versus the scaled BAM (a) and the scaled MBAM (b)

CONCLUSION

We proposed a novel no-reference blocking artifact metric with consideration of HVS. The performance of the proposed metric was evaluated through subjective experiments, where the results show high correlations with subjective data and low prediction errors.

For accurate video quality assessment, it is important to recognize that different distortions are predominant in different regions of digitally compressed images. However, the blocking artifact represents a major type of distortion in the reconstructed video compressed with B-DCT algorithm. Therefore the proposed BAM can be used as an important part in video quality assessment. For the reconstructed videos with the same compression algorithm, the blocking artifact can reflect well the video quality. So the proposed BAM can also be used to evaluate the quality of the videos with certain compression algorithm.

Without any need of original reference video at all, the proposed BAM can satisfy the requirement of video assessment in real time. It can be used to monitor the video quality in real time transmission applications such as videoconference.

References

Boccignone, G., Marcelli, A., Somma, G., 2003. Foveated Analysis of Video. IEEE Proceeding of ICIAP, p.638-643.

Gao, W., Mermer, C., Kim, Y., 2002. A de-blocking algorithm and a blockiness metric for highly compressed images. *IEEE Trans. CSVT*, **12**:1150-1159.

ITU-T Recommendation BT.500-10, 2000. Methodology for the Subjective Assessment of the Quality of Television Pictures.

Karunasekera, S.A., Kingsbury, N.G., 1995. A distortion measure for blocking artifacts in images based on human visual sensitivity. *IEEE Trans. Image Processing*, **4**(6): 713-724. [doi:10.1109/83.388074]

Privitera, C.M., Stark, L.W., 2000. Algorithms for defining visual regions-of-interest: comparison with eye fixations. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **22**(9):970-982. [doi:10.1109/34.877520]

Triantafyllidis, G.A., Tzovaras, D., Strintzis, M.G., 2002. Blocking artifact detection and reduction in compressed data. *IEEE Trans. CSVT*, **12**:877-890.

Vlachos, T., 2000. Detection of blocking artifacts in compressed video. *Electronics Letters*, **36**(13):1106-1108. [doi:10.1049/el:20000847]

VQEG Report, 2003. Draft Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment.

Winkler, S., 1999. Issues in vision modeling for perceptual video quality assessment. *Signal Processing*, **78**(2): 231-252. [doi:10.1016/S0165-1684(99)00062-6]

Winkler, S., 2000. Quality Metric Design: A Closer Look. Proc. SPIE Human Vision and Electronic Imaging V, **3959**: 37-44.

Wu, H.R., 1995. Analysis of Video Reconcstruction Artifacts and Quality Metrics. Proc. Australian Telecommunications Network and Application Conf., p.191-194.

Wu, H.R., Yuen, M., 1997. A generalized block-edge impairment metric for video coding. *IEEE Signal Processing Letters*, **4**(11):317-320. [doi:10.1109/97.641398]

Yu, Z., Wu, H.R., Winkler, S., Chen, T., 2002. Vision-model-based impairment metric to evaluate blocking artifacts in digital video. *Proceedings of the IEEE*, **90**:154-169. [doi:10.1109/5.982412]

Yuen, M., Wu, H.R., 1998. A survey of hybrid MC/DPCM/DCT video coding distortion. *Signal Processing*, **70**(3):247-278. [doi:10.1016/S0165-1684(98)00128-5]