

A NO-REFERENCE PERCEPTUAL BLOCKINESS METRIC

Hantao Liu¹ and Ingrid Heynderickx^{1,2}

¹ Department of Mediamatics, Delft University of Technology, The Netherlands

² Group Visual Experiences, Philips Research Laboratories, The Netherlands

ABSTRACT

A novel no-reference blockiness metric that can automatically and perceptually quantify blocking artifacts of DCT coding is presented. The proposed metric is built upon the specific structure information of the artifact itself combined with the properties of the human visual system (HVS) by means of a simple and efficient model of visual masking. Investigations are conducted to reduce the additional cost introduced by the human vision model, without compromising its overall prediction ability. The proposed metric is validated through comparing its performance to state-of-the-art HVS model based blockiness metrics with respect to accuracy, reliability and computational complexity.

Index Terms— Blocking artifact, human vision model, texture masking, luminance masking, image quality assessment

1. INTRODUCTION

Objective metrics have the aim to provide a quantitative measure for perceived image quality aspects. They are of fundamental importance to a broad range of image and video processing applications [1]. In current video chains, the receiving end, for example a TV-set, typically contains various objective quality metrics, which are implemented to improve the overall perceived quality of artifact impaired video. These metrics determine the quality of the incoming signal in terms of blockiness, ringing, blur, etc. and adapt the parameters in the video enhancement algorithms accordingly. Since in a digital chain the receiver usually has no access to the encoding parameters of the bit-streams, these objective metrics are constrained to a no-reference approach. This means that the impairment assessment only relies on the reconstructed image. Furthermore, taking into account the way human beings perceive quality aspects, while removing perceptual redundancies inherent in purely signal-based metrics can be greatly beneficial for matching objective quality prediction to human perception of quality. However, since an objective metric based on a model of the human visual system (HVS) often is computationally intensive for real-time applications, it is highly desirable to reduce the complexity of the HVS model embedded as well as the metric itself without compromising the overall performance.

A blocking artifact, which manifests itself as an artificial discontinuity between adjacent blocks, is known as the most annoying distortion type at low bit-rate DCT coding, and subjective experiments have indicated that blockiness is highly correlated with the overall perceived quality of MPEG-2 video e.g. [2]. Therefore, during the last decades a lot of research effort is devoted to the development of blockiness metrics. Most of these metrics, however, require a reference image or video (see for example [3] and [4]). Although designing no-reference metrics is

much more promising for real-world applications, it is still an academic challenge mainly due to the limited understanding of the human visual system [1]. Much work is done trying to incorporate relevant and accurately known properties of the HVS into no-reference blockiness metrics [5-7]. The generalized block impairment metric (GBIM) [6] expresses the blockiness as the inter-pixel difference across block boundaries scaled with a weighting function, which simply models the HVS masking effect. The total amount of blockiness is then normalized by the same measure calculated inside the coding blocks. The main drawbacks for GBIM are: (1) an efficient combination of different masking effects is not considered in modeling the HVS; and (2) the metric is designed such that the human vision model needs to be calculated for every pixel in an image, which is computationally very intensive. The blockiness metric based on a locally adaptive algorithm [7], which we refer to as LABM, examines the severity of blockiness block by block, and estimates its corresponding visibility by means of a just-noticeable-distortion (JND) profile. Subsequently, the resultant value is averaged over all the blocks in the image to yield the blockiness score. This metric is potentially more accurate, however, it exhibits several drawbacks: (1) as the blockiness is measured based on a fixed 8x8 block unit, the stability of this metric is doubted in case of spatial scaling; (2) calculating the JND profile is complex, and it cannot predict perceived annoyance above threshold; and (3) the JND is also calculated for every pixel in an image, which largely increases the computational cost of the metric. Furthermore, none of these metrics involves a detection phase for explicitly locating the precise grid position of block boundaries, which is practically needed in order to account for deviations in the blocking grid in the incoming signal or as a consequence of spatial scaling [8-10].

In this paper, we further rely on the locally adaptive approach taken in [7], but extend the idea by first detecting the precise grid position. Moreover, the extent of blockiness is locally measured, over each individual blocking artifact, with a purely pixel-based metric including the visibility of the distortion due to masking.

2. PROPOSED ALGORITHM

The schematic overview of the proposed algorithm is illustrated in Figure 1. Initially, a detection phase is adopted in order to identify the precise grid position of block boundaries. After locating all potential blocking artifacts in an image, a local level processing is carried out to individually examine each pre-detected blocking artifact by analyzing its surrounding content with a limited extent. This consists of two parallel steps: (1) a local blockiness metric (LBM), calculating the degree of distortion purely pixel-based; and (2) a human vision model, indicating the perceptual significance of the artifact to the human eye and outputting a visibility coefficient (VC). The resultant LBM and VC are integrated into a local

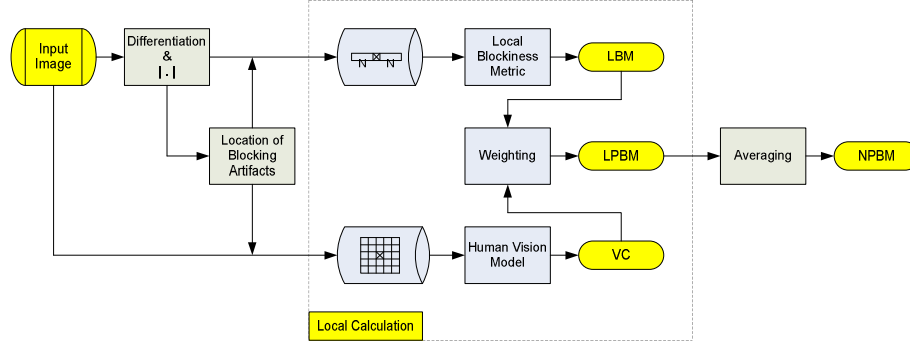


Fig. 1. Schematic overview of the proposed algorithm.

perceptual blockiness metric (LPBM). Finally, the LPBM is averaged over all detected artifacts in the image to produce a score of blockiness assessment (i.e. NPBM).

2.1. Blocking Artifact Location

In order to successfully measure blocking artifacts, one must be able to identify where they occur in a given image. Most, if not all, of the existing blockiness metrics simply assume that the grid of blocks of 8x8 pixels, on which blocking artifacts appear, starts exactly at the top-left corner of the image, however, this is not necessarily the case, e.g. after scaling decoded images, or due to a deviation in the incoming signal itself [8-10]. Hence, especially for a no-reference blockiness metric, where the bit-stream information is not available, a grid position detection approach is indeed required to ensure a reliable metric.

Based on the nature of the blocking artifact (i.e. spatial discontinuity) and the periodic property of the block-based coding scheme, a simple algorithm that maps the image into a 1-D signal profile, is proposed. Then, the precise grid position can be extracted from the amount of energy present in the gradient in 1-D signal profile in horizontal or vertical direction by means of the following procedure.

We denote the image of $M \times N$ (height x width) pixels as $I(i, j)$ for $i \in [1, M]$, $j \in [1, N]$, and calculate the gradient map

$$G_h(i, j) = |I(i, j+1) - I(i, j)|, \quad j \in [1, N-1] \quad (1)$$

The 1-D signal profile is then formed by summing G_h over all image rows M

$$S_h(j) = \sum_{i=1}^M G_h(i, j) \quad (2)$$

In the resulting 1-D signal profile, the grid position, which should be present as a periodic impulse train of signal peaks, might be masked by image content, such as object edges. The periodicity of the blocking grid, however, can be clearly revealed in the frequency domain using DFT, where the real block size (i.e. the period of the impulses) is obtained. The offset of the grid can then be directly retrieved from the 1-D signal. From the periodicity and the offset all blocking edges are calculated assuming a regular grid. The entire procedure is performed once in horizontal and once in vertical direction to address a possible asymmetry in the size or the offset of the blocks. The procedure has been shown to be robust against scaling and offset as occurring in daily-life applications.

2.2. Local Blockiness Metric

Since blocking artifacts intrinsically are a local phenomenon, their behavior can be reasonably described as a local distortion metric, indicating the relative signal discontinuity within a region of image content. Hence, a local blockiness metric (LBM), which examines each of the pre-detected blocking artifacts and individually provides a numerical measure of distortion, is proposed. This approach is potentially more accurate than a global approach, since the visual strength of the block discontinuity is primarily affected by its local surroundings. Furthermore, the local analysis based on each individual blocking artifact instead of on a fixed block unit, is practically more efficient in case of a deviating block size. In this paper, the blockiness is locally characterized as a blocking edge that stands out from its spatial vicinity, and is defined as the local gradient energy normalized by its neighboring pixels. This is done separately along each dimension. Figure 2 shows an example of a block discontinuity along the horizontal direction, where two adjacent blocks (i.e. A and B) are extracted from a real JPEG image.

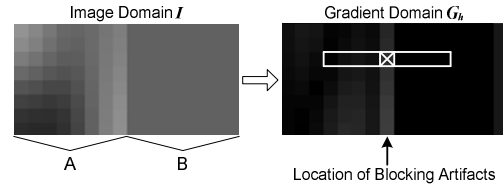


Fig. 2. Local blockiness metric (LBM).

LBM is estimated within a template centered on the blocking artifact at pixel location (i, j) as shown in Figure 2

$$LBM_h(i, j) = \begin{cases} \omega \times BG_h & \text{if } NBG_h = 0, BG_h \neq 0 \\ \frac{BG_h}{N BG_h} & \text{if } NBG_h \neq 0, BG_h \neq 0 \\ 0 & \text{if } NBG_h = 0, BG_h = 0 \end{cases} \quad (3)$$

where ω ($\omega = 1$ in our experiments) is used to adjust the amount of energy present in the gradient, BG_h and NBG_h are

$$BG_h = G_h(i, j) \quad (4)$$

$$N BG_h = \frac{1}{2n} \sum_{x=-n \dots n, x \neq 0} G_h(i, j+x) \quad (5)$$

and the template size of $(2n+1)$ depends on the detected block size.

2.3. The Human Vision Model

A simplified human vision model based on the spatial masking properties of the HVS is already proposed in [11]. It contains texture masking and luminance masking, which both are highly relevant to the perception of blocking artifacts. Both masking effects are implemented, based on the local characteristics of image signals, using spatial filtering followed by a weighting function. Then, they are efficiently combined into a single visibility coefficient (VC), which reflects the perceptual significance of the artifact quantitatively. In the following, the block discontinuity along the horizontal direction is discussed, but it can be done in the same way for the vertical direction.

2.3.1. Local Visibility due to Texture Masking

Texture masking is modeled simply calculating a visibility coefficient (VC_t); the higher its value, the smaller the masking effect, and hence, the stronger the visibility of the artifact. The procedure involves: (1) calculating the local background activity using a texture detector; (2) a classification scheme to capture the active background regions by thresholding; and (3) a visibility transform function (VTF) to obtain a visibility coefficient.

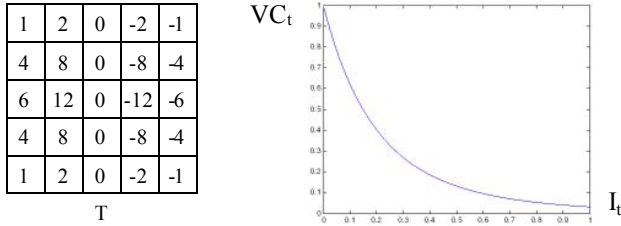


Fig. 3. High-pass filter and visibility transform function used.

As shown in Figure 3, a high-pass filter T is used to measure the background activity (i.e. $t(i, j)$) for each blocking artifact, and then a pre-defined threshold Thr ($Thr=0.15$ in our experiments) is applied to classify the background into “flat” or “texture”, resulting in an activity value $I_t(i, j)$, which is given by

$$I_t(i, j) = \begin{cases} 0 & \text{if } t(i, j) < Thr \\ t(i, j) & \text{otherwise} \end{cases} \quad (6)$$

$$t(i, j) = \frac{1}{48} \sum_{x=1}^5 \sum_{y=1}^5 I(i-3+x, j-3+y) \cdot T(x, y) \quad (7)$$

A visibility transform function (VTF) is proposed in accordance to human perceptual properties, where the visibility coefficient is inversely proportional (nonlinear) to the activity value. Figure 3 shows an approximation of the VTF, considered to be good enough, and already introduced in [11]. It is defined as

$$VC_t(i, j) = \frac{1}{(1 + I_t(i, j))^\alpha} \quad (8)$$

where $\alpha > 1$ ($\alpha = 5$ in our experiments) is used to adjust the nonlinearity.

2.3.2. Local Visibility due to Luminance Masking

It was found that the human visual system’s sensitivity to variations in luminance depends on (is a nonlinear function of) the local mean luminance [11]. Modeling the luminance masking

consists of: (1) calculating the local averaged background luminance; and (2) a visibility transform function (VTF) to obtain a visibility coefficient (VC_l) based on empirically driven properties of the HVS for luminance masking [11].

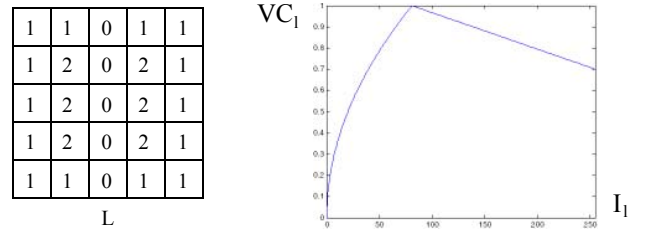


Fig. 4. Low-pass filter and visibility transform function used.

The local luminance $I_l(i, j)$ is calculated using a weighted low-pass filter as shown in Figure 4

$$I_l(i, j) = \frac{1}{26} \sum_{x=1}^5 \sum_{y=1}^5 I(i-3+x, j-3+y) \cdot L(x, y) \quad (9)$$

The relationship between the visibility coefficient and the local luminance is modeled based on the VTF shown in Figure 4 and discussed in [11]

$$VC_l(i, j) = \begin{cases} \left(\frac{I_l(i, j)}{81} \right)^{1/2} & \text{if } 0 \leq I_l(i, j) \leq 81 \\ \left(\frac{1-\beta}{174} \right) \cdot (81 - I_l(i, j)) + 1 & \text{otherwise} \end{cases} \quad (10)$$

where $0 < \beta < 1$ ($\beta = 0.7$ in our experiments) is used to adjust the slope of the linear part of this function

2.3.3. Integration Strategy

Since the visibility of an artifact depends on various masking effects co-existing in the HVS, an efficient integration strategy is needed to obtain an accurate perceptual model. In our approach, based on the local image content surrounding a stimulus, first texture masking is calculated. In case the local activity in the area is larger than a given threshold (see equation (6)), a visibility coefficient VC_t is applied, followed by the application of a luminance masking coefficient VC_l . In case the local activity in the area is low, only VC_l is applied. The application of VC_l , where appropriately combined with VC_t , results in an output value VC .

2.4. The Perceptual Blockiness Metric

The output of the proposed human vision model (VC) is used to locally weight the pixel-based blockiness metric LBM , resulting in a local perceptual blockiness metric ($LPBM$). Since the horizontal and vertical blocking artifacts are calculated separately, the $LPBM$ for the block discontinuity along the horizontal direction is described as

$$LPBM_h(i, j) = VC(i, j) \times LBM_h(i, j) \quad (11)$$

which is then averaged over all detected blocking artifacts in the whole image to determine the blockiness metric

$$NPBM_h = \text{mean}(LPBM_h) \quad (12)$$

A metric $NPBM_v$ can be similarly defined for the blockiness along the vertical direction. Assuming no interaction and no difference in sensitivity to blockiness in horizontal and vertical

direction, the two metrics are added together to give the resultant blockiness score

$$M_{NPBM} = \frac{NPBM_h + NPBM_v}{2} \quad (13)$$

Since, in our case, both the local blockiness metric and the human vision model are calculated at the locations of the blocking artifacts only, and not for all pixels in an image, the computational load of this metric is largely reduced.

3. PERFORMANCE EVALUATION

Subjective ratings resulting from psychovisual experiments are widely accepted as the benchmark for evaluating objective quality metrics. In our work, the LIVE JPEG database [12], which consists of 233 JPEG images with their Mean Opinion Score (MOS), is adopted to validate various objective blockiness metrics. An evaluation is carried out comparing two most closely related blockiness metrics, namely GBIM and LABM (as implemented based on the descriptions in [6] and [7], respectively), to our proposed metric NPBM for the same image database. The performance of an objective metric can be quantitatively evaluated in terms of accuracy, monotonicity, and consistency with the Pearson linear correlation coefficient, the Spearman rank order correlation coefficient, and the outlier ratio between the MOS scores and the metric's predictions, respectively [13]. As suggested in [13], nonlinear correlations using mapping functions for the objective predictions may be computed as alternatives. They usually yield higher correlation coefficients as demonstrated with our results reported in [11]. They, however, have the disadvantage of minimizing performance difference between metrics. Hence, to make a more critical comparison, only the linear correlations are calculated in this paper (to make a comparison with nonlinear regression possible, it should be noted that NPBM results in a Pearson correlation of **0.94** in that case).

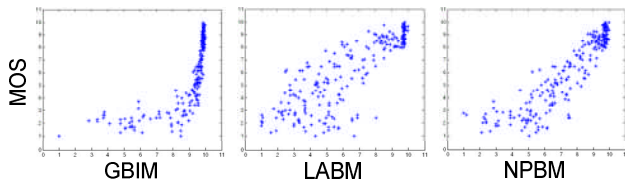


Fig. 5. Scatter plots of MOS vs. blockiness metrics.

Metric	Pearson Linear Correlation	Spearman Rank Order Correlation	Outlier Ratio
GBIM	0.736	0.912	0.099
LABM	0.834	0.832	0.009
NPBM	0.900	0.904	0

Table 1. Performance comparison of blockiness metrics.

Figure 5 shows the scatter plots of the MOS versus GBIM, LABM and NPBM, respectively, and the performance evaluation results are listed in Table 1. It demonstrates that our proposed metric outperforms the other two in the prediction of perceived quality. Furthermore, the most promising feature achieved with NPBM is that its major computational cost introduced by the human vision model is reduced, by simplifying the model itself as

well as minimizing the number of times the model is calculated in the whole process of determining the blockiness score for an image. For example, in our experiments, the total amount of time needed to compute the vision model in NPBM, is only 1/8 of that required in both GBIM and LABM.

4. CONCLUSIONS

A novel no-reference blockiness metric that is highly consistent to the perception of blocking artifacts is presented. The proposed metric involves a grid detection phase, which is used to account for a block size change or grid shift, and intrinsically ensures the subsequent local processing of the blocking artifacts. For each pre-detected blocking artifact the blockiness is individually calculated as a signal discontinuity relative to its local content and its visibility due to masking is locally estimated. Combining the results in a simple way yields a metric that shows a promising performance with respect to practical reliability, prediction accuracy, and computational efficiency.

5. REFERENCES

- [1] Z. Wang and A. Bovik, *Modern Image Quality Assessment*, Morgan & Claypool, USA, 2006.
- [2] C. C. Koh, S. K. Mitra, J. M. Foley and I. Heynderickx, "Annoyance of Individual Artifacts in MPEG-2 Compressed Video and Their Relation to Overall Annoyance," *Proc. SPIE*, vol. 5666, pp. 595-606, Jan. 2005.
- [3] Z. Yu, H. R. Wu, S. Winkler and T. Chen, "Vision-Model-Based Impairment Metric to Evaluate Blocking Artifacts in Digital Video," *Proceedings of the IEEE*, vol. 90, pp. 154-169, Jan. 2002.
- [4] S. A. Karunasekera and N. G. Kingsbury, "A Distortion Measure for Blocking Artifacts in Images Based on Human Visual Sensitivity," *IEEE Trans. Image Processing*, vol. 4, no. 11, pp. 713-724, June 1995.
- [5] S. Liu and A. C. Bovik, "Efficient DCT-Domain Blind Measurement and Reduction of Blocking Artifacts," *IEEE Trans. on CSVT*, vol. 12, no. 12, pp. 1139-1149, Dec. 2002.
- [6] H. R. Wu and M. Yuen, "A Generalized Block-edge Impairment Metric for Video Coding," *IEEE Signal Processing Letters*, vol. 70, no. 3, pp. 247-278, Nov. 1998.
- [7] F. Pan, X. Lin, S. Rahardja, W. Lin, E. Ong, S. Yao, Z. Lu and X. Yang, "A locally adaptive algorithm for measuring blocking artifacts in images and videos," *Signal Processing: Image Communication*, vol. 19 i6, pp. 499-506, 2004.
- [8] E. Lesellier and J. Jung, "Robust wavelet-based arbitrary grid detection for MPEG," *Proc. IEEE ICIP*, pp 417-420, 2002.
- [9] R. Muijs and I. Kirenko, "A No-reference Blocking Artifact Measure for Adaptive Video Processing," *Proc. EUSIPCO*, 2005.
- [10] S. Tjoa, W. S. Lin, H. V. Zhao and K. J. R. Liu, "Block Size Forensic Analysis in Digital Images," *Proc. ICASSP*, 2007.
- [11] H. Liu and I. Heynderickx, "A Simplified Human Vision Model Applied to a Blocking Artifact Metric," *Proc. CAIP*, LNCS 4673, pp. 334-341, 2007.
- [12] LIVE: <http://live.ece.utexas.edu/research/quality>.
- [13] VQEG: <http://www.vqeg.org>.