PERCEPTUALLY WEIGHTED DISTORTION MEASURE IN LOW BITRATE BLOCK-BASED VIDEO CODERS

Jaehan In [1], Ali Jerbi [1], and Foued Ben Amara [2]

[1] UB Video Inc., 1038-1040 Hamilton St., Vancouver, BC V6B 2R9, Canada
[2] The Univ. of Toronto, 5 King's College Rd., Toronto, Ontario M5S 3G8, Canada
Email: [1] {jaehan, ali}@ubvideo.com, [2] foued@mie.utoronto.ca

ABSTRACT

Visual artifacts found in videos compressed at low bit-rates may be the results of various decisions made in the encoding process. Although many post-processing techniques are available to detect and reduce such artifacts, they often lead to undesired blurriness and require extensive computations. A better approach is to avoid such artifacts in the first place by making better decisions in the encoding process. In this paper, a Perceptually Weighted Distortion (PWD) measure is presented. This distortion measure, which outperforms the conventional distortion measures in terms of subjective quality, leads to an acceptable increase in computational complexity.

1. INTRODUCTION

The block-based hybrid-coding framework has been widely employed in various video compression standards from ISO/IEC MPEG-1/2/4, ITU-T H.261/263/263+ to the most recent H.264/MPEG-4 Part 10 AVC. A typical hybrid video coding process consists of four steps as shown in Figure 1. First, temporal and spatial redundancies are reduced by motion estimation and intra prediction. In motion estimation, each block in the source video frame is predicted from a coded block at a different temporal location. In intra prediction, on the other hand, the pixels are predicted from the neighboring pixels at the same temporal location. The residual block, which contains the prediction error between the original and the predicted block, is then transformed to the frequency domain for more efficient coding. The resulting transform coefficients are then quantized and finally entropy-coded to achieve further compression. Various types of artifacts found in the compressed video are generated in steps 1 and 3 of the coding process. Some of the most annoying artifacts are the blocking artifacts. A number of post-processing techniques have been proposed to detect and reduce such artifacts [1]. While these techniques have been proven to be effective in most cases, they are decoder-dependent and perform less efficiently for frames coded using a large quantization step size that results in a reconstructed frame consisting mainly of the predicted pixels. It follows that the type of artifacts that appear in the decoded frame is strongly dependent on the accuracy of the prediction. This suggests that one way to prevent the appearance of visual artifacts is to carefully select a better prediction for a given source block during the encoding process [2]. A common method in selecting the best prediction is to compare, for each source block, the associated distortions among the available prediction candidates. Some of the popular distortion measures include mean squared error (MSE) and peak signal-to-noise ratio (PSNR). These distortion measures are not representative of the distortions perceived by the Human Visual System (HVS). In practice, sum of absolute differences (SAD) or mean absolute difference (MAD) are used due to their low computational cost. Although distortion measures that better match the perceived distortion have been proposed in recent studies [3], they are usually computationally intensive. In this paper, we propose to deal with the blocking artifacts during the encoding process by introducing a perceptually weighted distortion measure that outperforms the popular distortion measure at the cost of a reasonably increased complexity. In the next section, we discuss the proposed distortion measure. Experimental results and conclusions are presented in Sections 3 and 4, respectively.

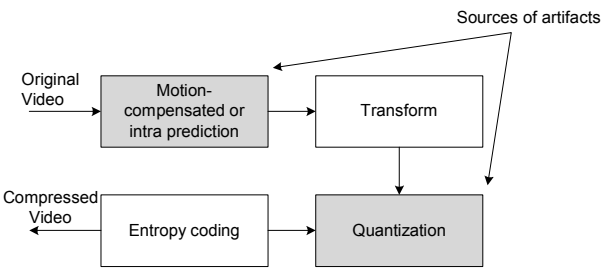

Figure 1: Hybrid video coding framework

0-7803-8554-3/04/$20.00 ©2004 IEEE.

2. PERCEPTUALLY WEIGHTED DISTORTION

Quantifying the perceived distortion is a challenging task due to the very complex nature of the HVS. Although various aspects of the HVS are still not well understood, there are a few well-defined characteristics such as the difference threshold or the just noticeable difference (JND). The JND is the minimum intensity difference between two stimuli that is detectable by the human eye. For closely located intensities whose differences are below the JND, the human eye does not distinguish between the individual intensities, but perceives their average. This behavior of the human eye plays a major role in the design of the perceptual distortion measure outlined in the following.

The assumption upon which the proposed measure is based is as follows: Among the predicted blocks chosen based on the traditional distortion measures, the one whose average intensity is the closest to that of the source block will for the most part result in the best subjective quality. To illustrate this point, consider the example shown in Figure 2, where four 32x32 blocks (a), (b), (c) and (d) are given. Block (a) is a flat block consisting of pixel values 128. The neighboring block to the right, denoted by (b) consists of intensity values of 128 and 144 arranged as shown in the matrix at the bottom right of Figure 2. On the other hand, block (c), above block (a), is represented by pixel values of 120 and 136 as given by the matrix at the top right of Figure 2. Finally, block (d), located below block (a) with pixel intensities of 116 and 139, arranged in the matrix shown in the bottom of the Figure 2. It can be noted that the edge between block (a) and block (b) is much more vivid than that between either block (a) and block (c) or block (a) and block (d). In other words, if we were to find a match for block (a), blocks (c) and (d) will better represent block (a) perceptually. However, if we use the SAD or MSE measures, block (b) will result in the smallest SAD or MSE and will be chosen as the best match for block (a). Now, if we were to consider the smallest average of the residual values as the distortion measure, then block (c), with a zero average, will be the best match for block (a), which is considered the best perceptual prediction.

Motivated by the example above, we propose a perceptually weighted distortion measure (PWD), $\widetilde{D}$, as follows:

$$\widetilde{D} = D + D1 + D2$$

where D is the traditional distortion measure such as SAD or MAD; $D1$ and $D2$ are two perceptual distortion measures discussed below. We choose SAD as the traditional distortion measure in the following description of our proposed distortion measure. Let $x_k(i,j)$ be the original pixels in the k^{th} 4x4 block in the block of size $L \times M$ $(L,M \geq 4)$, $y_k(i,j)$ be the predicted pixels, and $r_k(i,j)$ be the residual pixel obtained using $r_k(i,j) = x_k(i,j) - y_k(i,j)$. Then

$$\text{SAD} = \sum_{k=0}^{N-1} \sum_{i=0}^{3} \sum_{j=0}^{3} \left| r_k(i,j) \right|$$

where N is the number of 4x4 blocks in the $L \times M$ block. We compute the $L \times M$ SAD as the sum of the 4x4 block SADs such that the residual 4x4 blocks can be used in the computation of $D1$ as follows:

$$D1 = \sum_{k=0}^{N-1} \alpha \cdot \left| \sum_{i=0}^{3} \sum_{j=0}^{3} r_k(i,j) \right|$$

where α is the weighting parameter whose value depends on the quantization step size and brightness of the block being measured. $D1$ represents the absolute value of the sum of the differences in a 4x4 block. Adding $D1$ to the SAD measure will essentially results in choosing a prediction block that minimizes the SAD as well as the average residual values. As a consequence, the average of the predicted block will be forced to get as close as possible to that of the source block. As a result, the decoded 4x4 blocks will get perceptually closer to those in the source frame as discussed earlier.

However, selecting the best prediction that is based solely on $D1$ may lead to degradation in the perceived quality in cases where the spatial properties of the predicted and the original blocks are different, but the residuals have an average of zero. Consider a 4x4 block in the original frame with pixel luma intensities given as [128, 128 ; 128, 128]. Assume that the predicted block consists of the values [0, 0 ; 256, 256]. Even though the residual block [128, 128 ; -128, -128] has a luma average of zero, choosing such prediction will obviously lead to degradation in the perceptual quality. In general, when the intensity differences between the pixels in the original block are below JND and those in the predicted block are above JND (or vice versa), the perceived difference by the HVS is more vivid. Minimizing the differences between neighboring residuals will preserve the JND properties between the source block and its corresponding prediction. Therefore $D2$ is defined as:

$$D2 = \sum_{k=0}^{N-1} \beta \cdot \left(\sum_{i=0}^{2} \sum_{i=0}^{2} \left(\left| h_k(i,j) \right| + \left| v_k(i,j) \right| \right) \right)$$

where

$$h_k(i,j) = r_k(i,j) - r_k(i+1,j)$$

$$v_k(i,j) = r_k(i,j) - r_k(i,j+1)$$

and β is the weighting parameter whose value also depends on the quantization step size and brightness of the block. The impact of D_2 is to minimize a variance-like measure of the residual values within each 4x4 block. This will lead to a prediction that preserves details present in the source block while avoiding the false edges.

In summary, the new distortion measure uses, in addition to the well-known SAD, the average and the gradient information of the residual values of each 4x4 block. Minimizing the three components will result in a predicted block corresponding to the source block that approximates as much as possible the average and the spatial distributions present in the source block, leading to a better perceptual prediction.

3. EXPERIMENTAL RESULS

In this section, we present the performance of the proposed perceptual distortion measure using the latest H.264/AVC encoder [4]. In the following experiments, a highly optimized H.264 Baseline Profile encoder [5] is used that uses SAD as the distortion measure in three locations of the encoding process: comparisons among intra predictions, among motion compensated predictions, and between the best intra and motion compensated predictions. While keeping the rest of the encoding process identical, we replace SAD with the perceptually weighted distortion measure. We also use a fixed quantization step size and disable the de-blocking filter in our experiments such that any changes in the subjective quality and the bit-rate are solely caused by the distortion measure. Figures 3 and 4 show the decoded intra and inter frames of Foreman coded at QP = 36 using SAD and PWD with $\alpha = 0.0625$ and $\beta = 0.125$, repectively. The proposed distortion measure results in no or slight increase in bit-rate (< 4%). However, as evident in Figures 3 and 4, the subjective quality of the frames encoded using PWD is significantly better than that using SAD.

4. CONCLUSION

We demonstrated the performance of the perceptually weighted distortion measure that leads to better subjective quality. Our proposed distortion measure is designed as a combination of the traditional measure and the weights whose values vary according to the HVS. The proposed distortion measure results in an increase in complexity in the order of 40%. Reduction of this figure to 25% is achieved when our implementation is optimized using SIMD extensions. We believe that further reduction of computational complexity is possible through algorithmic improvements such as selective application of PWD in blockiness-prone areas and is left for future work.

5. REFERENCES

[1] S. Liu and A. C. Bovik, "Efficient DCT-Domain Blind Measurement and Reduction of Blocking Artifacts," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 12, pp. 1139–1149, Dec. 2002.

[2] F. B. Amara, A. Jerbi, J. Au, and F. Kossentini, "Trailing Artifact Avoidance for Low Bit-rate Block-Based Video Coder," *First International Symposium on Control, Communications and Signal Processing*, Hammamet, Tunisia, 21-24 March 2004.

[3] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image Quality Assessment: From Error Measurement to Structural Similarity," *IEEE Transactions on Image Processing,* vol. 13, no. 1, pp. 1−14, Jan. 2004.

[4] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra, "Overview of the H.264/AVC Video Coding Standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, July 2003.

[5] A. Joch and F. Kossentini, *Demonstration of a Computation-Optimized JVT Codec*, Doc. JVT-C148, Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG (ISO/IEC JTC1/ SC29/WG11 and ITU-T SG16 Q.6), Fairfax, Virginia, USA, 6-10 May, 2002.

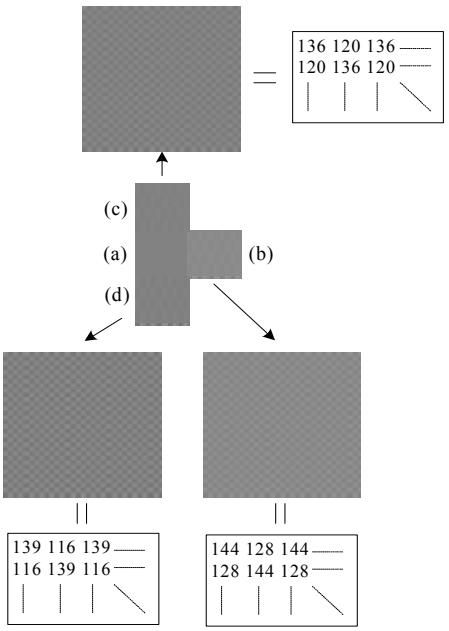

Figure 2: (a) 32x32 block whose pixel values are all 128, predictions with average residual of (b) 8, (c) 0, and (d) 0.5

(a)

(b)

Figure 3: Intra-frame of Foreman coded at 147 Kbps (QP=36) using (a) SAD and (b) PWD

(a)

(b)

Figure 4: Inter-frame of Foreman coded at 147 Kbps (QP=36) using (a) SAD and (b) PWD