ANALYSIS OF ONE-DIMENSIONAL TRANSFORMS IN CODING MOTION COMPENSATION PREDICTION RESIDUALS FOR VIDEO APPLICATIONS

Harley Zhang and Jae S. Lim, IEEE Fellow*

Research Laboratory of Electronics
Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology

ABSTRACT

In video coding, motion compensation prediction residuals are typically compressed by applying two-dimensional transforms such as the two-dimensional discrete cosine transform (2D-DCT). Using direction-adaptive one-dimensional discrete cosine transforms (1D-DCTs) can provide significant additional bitrate savings. However, this requires optimization over all available transforms to minimize the overall bitrate, which can be expensive in terms of time and computation.

We examine the use of only the horizontal and vertical 1D-DCTs in addition to the 2D-DCT for coding motion compensation residuals. By reducing the number of available transforms, the amount of required computation decreases significantly, with a potential cost in performance. We perform experiments using a modified H.264/AVC codec to compare the performance of using different sets of available transforms. The results indicate that for typical applications of video coding, most of the performance benefit from using directional 1D-DCTs can be retained by keeping only the horizontal and vertical 1D-DCTs.

Index Terms— Discrete cosine transforms, motion compensation, video coding

1. INTRODUCTION

The spatial characteristics of motion compensation (MC) residuals differ greatly from those of regular images, indicating that transforms such as the two-dimensional discrete cosine transform (2D-DCT) which are used for images may not provide good performance for coding MC residuals. MC residuals often display locally anisotropic features in the sense that on a local scale, the high-amplitude values of MC residuals display some particular directionality. These features are fundamentally different from the anisotropic features found in regular images; MC residuals have high-amplitude intensities at object and region edges and low-amplitude intensities in smooth, stationary regions.

Based on these observations, experiments have been performed by augmenting the H.264/AVC codec with a set of directional one-dimensional discrete cosine transforms (1D-DCTs) for use in MC residual coding [1]. The results have shown that using transforms that exploit the particular anisotropic properties of MC residuals can provide significant additional bitrate savings. However, extra calculations are required to optimize the selected transform for each block of every MC residual, which can be costly and time-consuming in practice.

In this paper, we analyze the performance of 1D-DCTs for coding MC residuals to determine if their usage can be modified to reduce computation while still maintaining significant bitrate savings. We explore the performance of the H.264/AVC codec with access to different sets of transforms for MC residual coding. Experimental results show that most of the bitrate savings can be preserved by keeping only the horizontal and vertical 1D-DCTs in addition to the 2D-DCT. These results indicate that significant increases in compression efficiency can be made with little added complexity in comparison to the unmodified codec, encouraging research in this direction for further improvements in video compression.

In Section 2, we review the results from using 1D-DCTs to code MC residuals. In Section 3, we describe the experimental setup and implementation of the system used in the experiments presented in this paper. Section 4 describes the new experimental results and further analyses of the impact of using 1D-DCTs on MC residuals. Finally, Section 5 concludes with a summary of the results, as well as directions for future related research.

2. PREVIOUS RESEARCH

Previous work has shown that MC residuals differ consistently from regular images in their statistical properties. As modeled in [2], MC residuals display stronger local 1D anisotropic features than regular images, most commonly oriented close to the horizontal and vertical directions. To exploit these characteristics of MC residuals, sets of directional 1D-DCTs were developed in [1] to code MC residual blocks

* Harley Zhang is currently at Oliver Wyman Financial Services.

using a modified H.264/AVC codec. These sets of transforms apply to 8x8 and 4x4 pixel blocks, corresponding to the block sizes used in H.264/AVC. Together, the directional 1D-DCTs in each set cover $180°$.

In H.264/AVC, for each macroblock, the block size for motion compensation is determined along with the best transform to apply to each block, using a Lagrangian-based rate-distortion optimization method that minimizes a mean squared error metric. The 2D-DCT may still work well for regions of the MC residual that do not have strong 1D anisotropic features, so it is also included in the set of available transforms. After selecting the optimum size and transform for a particular block, the transform is performed and the transform coefficients are quantized and entropy coded using context-adaptive variable length coding (CAVLC).

The introduction of additional transforms means that side information must also be coded for each block to indicate the corresponding inverse transform. To minimize the overall number of side information bits, the transforms that are selected more frequently should be assigned shorter codewords. Based on the relative probabilities of selection for each transform, [1] uses a 1-bit codeword for the 2D-DCT and 5-bit codewords for each 1D-DCT in the case of 8x8 blocks. For 4x4 blocks, the same 1-bit codeword is used to indicate the 2D-DCT, and 4-bit codewords are used for each 1D-DCT. Using shorter side information codewords could reduce bitrate. This can be achieved by having fewer transforms available to code MC residual blocks. If the only available 1D-DCTs are the transforms in the horizontal and vertical directions, then each block only has three transform options, and the 1D-DCTs can be represented by 2-bit codewords.

The experimental results from [1] show that adding the 1D-DCTs provides bitrate savings in all cases under the range of picture qualities used, which ranged from about 30dB to 40dB in PSNR. Moreover, the bitrate savings increase with picture quality. At higher picture qualities, a larger fraction of the total bitrate is used to code MC residual transform coefficients, so the bitrate savings provided by the 1D-DCTs is amplified compared to lower picture qualities. The compression was performed in three modes with different available block sizes, and the average bitrate savings over a range of picture qualities varied from 4.1% to 11.4%. The average percentage of the bitrate that was used to code side information ranged from 3.6% to 5.9%, which suggests that cutting the number of side information bits by a significant amount would have a small but noticeable impact on the overall bitrate.

Various 1D transform techniques have been developed to take advantage of anisotropic features in images by filtering along directions of smaller intensity variation. These include methods based on DCTs [3], discrete wavelet transforms (DWTs) [4], and directionlets [5]. All of these techniques could be applied to MC residuals in addition to still images.

3. EXPERIMENTAL IMPLEMENTATION AND SETUP

To analyze the performance of different sets of transforms in coding MC residuals, we further augment the modified H.264/AVC codec used in [1], which can be referred to for details on 1D transform implementation, selection, and coding in the modified H.264/AVC codec. The primary differences compared to the new system used in this paper originate from the restriction of the set of available transforms. When rate-distortion optimization is performed to select a transform for each block of an MC residual, the optimization is only done over the transforms that are indicated as available.

The implementation of the selected transform remains the same as in [1], but the side information indicating the selected transform uses fewer bits when fewer transforms are available. The 2D-DCT is always assigned a 1-bit codeword, while each available 1D-DCT is assigned a codeword of length $1 + \log_2 N$, where N is the number of available 1D-DCTs for the block size being processed. This modified coding of side information is implemented at both the encoder and the decoder.

The modified H.264/AVC codec can be used in several different modes. For the results presented in this paper, the codec was run using three different available block sizes: 4x4 only, 8x8 only, and both 4x4 and 8x8. In all cases, a transform is selected for each block, rather than forcing all four 4x4 blocks in each 8x8 block to use the same transform as in [1].

For each set of transform block sizes, we consider three cases for the set of transforms available for coding MC residuals. Case 1 uses only the 2D-DCT, which produces the basic unmodified codec. Case 2 uses the 2D-DCT plus the vertical and horizontal 1D-DCTs. Case 3 uses the 2D-DCT and all directional 1D-DCTs, which produces a codec used in [1]. Since there are three options for available block sizes and three options for available MC residual transforms, we have nine different codecs.

We use these codecs on ten QCIF (177x144) resolution sequences at 30 frames per second (fps). The first 180 frames of each sequence are encoded, or the entire sequence if it comprises fewer than 180 frames. The first frame is encoded as an I-frame, and the rest of the frames are encoded as P-frames.

Many of the results in the remainder of this paper are presented as or derived from PSNR vs. bitrate rate-distortion data. The PSNR is given in dB, and it is calculated for the luminance component only, since the 1D-DCTs are only applied to the luminance component. The bitrate is given in kb/sec, representing the total bitrate for all encoded information. This includes transform coefficients for the luminance and chrominance components, displacement vectors, and side information. Since the chrominance components are typically undersampled in comparison with the luminance component, they contribute significantly less to the overall bitrate [1]. To obtain multiple PSNR vs. bitrate data points for each sequence, we vary the quantization parameter of the codec.

Lower values of the quantization parameter produce higher quality reconstructions, usually giving a higher PSNR in exchange for a higher bitrate.

4. RESULTS

4.1. Rate-Distortion Plots

Figure 1 shows the rate-distortion plots for the *Bridge close* QCIF video sequence using only 4x4 transform blocks. Three curves are shown, representing the three different sets of available transforms for MC residual coding. The shapes of the curves are similar for other test sequences, and for codecs with different available block sizes.

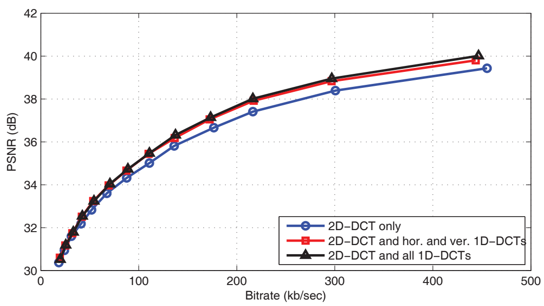

Fig. 1. Rate-distortion plots using codecs with access to different sets of transforms and using 4x4 transform blocks only for *Bridge close* test sequence.

These plots show that at a given bitrate, the achieved PSNR increases as the set of available transforms expands. Equivalently, at a fixed PSNR, the required bitrate decreases as the set of available transforms expands. As seen in [1], the percentage bitrate savings over Case 1 increases at higher video qualities for Cases 2 and 3. A larger percentage of the bitrate is used to code transform coefficients at higher qualities, so the benefits of improved transform coefficient compression becomes more significant at higher qualities.

From Figure 1, based on how close the curves for Cases 2 and 3 are in comparison to Case 1, it is clear that using only the 2D-DCT and the horizontal and vertical 1D-DCTs can provide nearly all of the bitrate savings of using the 2D-DCT and all of the directional 1D-DCTs. In a few cases, Case 2 even performs better than Case 3.

4.2. Bjontegaard-Delta Bitrate Results

To compare the separation between the curves shown in Figure 1 over the different video sequences, the Bjontegaard-Delta (BD) bitrate metric is used [6]. This metric provides a measure of the average bitrate savings for an encoder with respect to another encoder over a range of quality levels. Figure 2 shows the BD bitrate metrics comparing Case 2 with Case 1 and comparing Case 3 with Case 1 for the ten QCIF test sequences using 4x4 blocks only. The BD metrics are computed over a range of picture qualities corresponding to H.264/AVC quantization parameter values from 12 to 36.

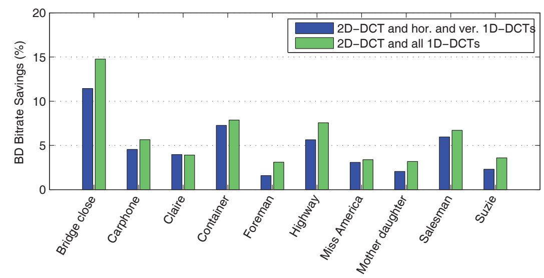

Fig. 2. Bjontegaard-Delta bitrate savings for ten QCIF resolution test sequences as compared to the total bitrate when only the 2D-DCT is used, 4x4 blocks only.

For nine out of ten sequences, the average bitrate savings over all picture qualities is higher when all of the 1D-DCTs are available, in comparison to when only the horizontal and vertical 1D-DCTs are available. For the one sequence in which Case 2 gives higher average bitrate savings than Case 3 (*Claire*), the BD metrics are nearly identical. The average BD metric is 4.78% for Case 2 and 5.97% for Case 3.

In the two other block size modes, the average bitrate savings is higher when all of the 1D-DCTs are available for all of the video sequences. When only 8x8 blocks are used, the average BD metric is 12.20% for Case 2 and 16.55% for Case 3. When both 4x4 and 8x8 blocks are used, the average BD metric is 6.12% for Case 2 and 8.12% for Case 3. The performances of Cases 2 and 3 are closest when only 4x4 blocks are used, and furthest when only 8x8 blocks are used.

Based on Figure 2, Case 3 performs better on average than Case 2 in terms of bitrate savings. This verifies that the benefit of having better adaptive transforms typically outweighs the extra bits of side information needed for each block. However, the data also indicate that using only the horizontal and vertical 1D-DCTs captures a significant portion of the bitrate savings afforded by using all of the directional 1D-DCTs. For all test sequences in all block size modes, Case 2 provides more than 50% of the average bitrate savings that Case 3 provides in terms of the BD metric.

4.3. Side Information Bitrates

As picture quality increases (or the quantization parameter decreases), the percentage of the bitrate used for side information generally increases and then decreases. The initial increase is partially due to more blocks selecting a transform, rather than being quantized to all zeros. The H.264/AVC encoder checks to see if a block is quantized to all zeros, in which case no transform needs to be performed and thus no side information needs to be sent. In addition, the increasing frequency with which 1D-DCTs are selected as compared to the 2D-DCT contributes to the rise in percentage of total bitrate used by side information, since the 1D-DCTs use more bits of side information than the 2D-DCT.

Beyond a certain picture quality, nearly all of the MC residual blocks that could benefit from using a transform are already being transformed, so the additional quality improvement comes primarily from finer quantization. More accurate quantization requires more bits to represent transform coefficients, increasing the total bitrate. At the same time, the amount of side information does not change significantly, since the number of blocks that are not zero-coefficient blocks is almost constant. Therefore, the side information takes up a decreasing fraction of the total bitrate.

Similar results hold for all three block size modes. Among the block size modes, the difference between Cases 2 and 3 is larger for 8x8 blocks than for 4x4 blocks. Case 3 always results in a higher percentage of the total bitrate being used to transmit side information. As we will examine further below, this is primarily due to codeword length, since the 1D-DCTs as a whole are selected at similar frequencies in the two cases.

4.4. Frequencies for Selection of Transforms

For each MC residual block that the H.264/AVC encoder processes, rate-distortion optimization is performed to select a transform. When a block consists entirely of transform coefficients that would be quantized to zero, there is no need to encode the block with any particular transform, so the encoder sends a codeword to indicate this information to the decoder. Therefore, each MC residual block is assigned one of the available transforms or marked as a zero-coefficient block.

For each sequence at a given quantization parameter level, the percentage of zero-coefficient blocks is similar among all three cases. This indicates that the addition of more transforms does not significantly affect blocks that were originally considered to be zero-coefficient blocks. This is consistent with the fact that many MC residual blocks are actually very close to being entirely zero, since motion compensation produces good predictions in most regions.

Another observation is that the percentages of blocks that select any 1D-DCT are fairly close for Cases 2 and 3. This indicates that a large fraction of the blocks that choose a non-horizontal, non-vertical 1D-DCT in Case 3 end up choosing either the horizontal or vertical 1D-DCT in Case 2. Therefore, while the horizontal and vertical 1D-DCTs are less efficient in terms of compression for these blocks, they still provide better performance than the 2D-DCT in most instances.

5. CONCLUSIONS

We report experimental results demonstrating that significant bitrate savings can be achieved by adding the horizontal and vertical 1D-DCTs in coding MC residuals. Moreover, for quality levels found in most typical video applications, these bitrate savings are close to the bitrate savings obtained by adding a full set of directional 1D-DCTs. By using only the horizontal and vertical 1D-DCTs as opposed to all of the 1D-DCTs, the number of required computations is reduced while still maintaining comparable compression efficiency. Future research in this direction can further improve compression efficiency, for example by taking advantage of correlations among the transform directions of neighboring blocks and by adaptively varying the directions of the available directional transforms. Additional details on the work reported in this paper can be found in [7].

6. REFERENCES

[1] F. Kamisli and J.S. Lim, "Video compression with 1-d directional transforms in h.264/avc," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2010*. IEEE, 2010, pp. 738–741.

[2] F. Kamisli and J.S. Lim, "Transforms for the motion compensation residual," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2009*. IEEE, 2009, pp. 789–792.

[3] B. Zeng and J. Fu, "Directional discrete cosine transforms for image coding," *Multimedia and Expo, 2006 IEEE International Conference on*, pp. 721–724, July 2006.

[4] C.-L. Chang and B. Girod, "Direction-adaptive discrete wavelet transform for image compression," *Image Processing, IEEE Transactions on*, vol. 16, no. 9, pp. 1289–1302, May 2007.

[5] V. Velisavljevic, B. Beferull-Lozano, M. Vetterli, and P.L. Dragotti, "Directionlets: anisotropic multidirectional representation with separable filtering," *Image Processing, IEEE Transactions on*, vol. 15, no. 7, pp. 1916–1933, July 2006.

[6] G. Bjontegaard, "Calculation of average psnr differences between rd-curves," *VCEG Contribution VCEG-M33*, 2011.

[7] H. Zhang, "Analysis of one-dimensional transforms in coding motion compensation prediction residuals for video applications," M.Eng thesis, Dept. Elect. Eng. and Comp. Sci., MIT, Cambridge, MA, 2011.