Video compression with 1-D directional transforms in H.264/AVC

<table>
<thead>
<tr>
<th>Citation</th>
<th>Kamisli, Fatih, and Jae S. Lim. “Video Compression with 1-D Directional Transforms in H.264/AVC.” IEEE International Conference on Acoustics Speech and Signal Processing, 2010. 738–741. © Copyright 2010 IEEE</th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1109/ICASSP.2010.5495034">http://dx.doi.org/10.1109/ICASSP.2010.5495034</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>Institute of Electrical and Electronics Engineers (IEEE)</td>
</tr>
<tr>
<td>Version</td>
<td>Final published version</td>
</tr>
<tr>
<td>Accessed</td>
<td>Sun Apr 07 07:29:01 EDT 2019</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/72132">http://hdl.handle.net/1721.1/72132</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td></td>
</tr>
</tbody>
</table>
VIDEO COMPRESSION WITH 1-D DIRECTIONAL TRANSFORMS IN H.264/AVC

Fatih Kamisli and Jae S. Lim

Research Laboratory of Electronics
Massachusetts Institute of Technology

ABSTRACT

Typically the same transforms, such as the 2-D Discrete Cosine Transform (DCT), are used to compress both images in image compression and prediction residuals in video compression. However, these two signals have different spatial characteristics. In [1], we analyzed the difference between these two signals and proposed 1-D directional transforms for prediction residuals. In this paper, we provide further experimental results using these transforms in the H.264/AVC codec and present other related information which can provide insights in understanding the use of these transforms in video coding applications.

Index Terms— Discrete cosine transforms, Motion compensation, Video coding

1. INTRODUCTION

An important component of image and video compression systems is a transform. A transform is used to transform image intensities. A transform is also used to transform prediction residuals of image intensities, such as the motion-compensation residual (MC-residual), the resolution-enhancement residual in scalable video coding, or the intra-prediction residual in H.264/AVC. Typically, the same transform is used to transform both image intensities and prediction residuals. For example, the 2-D Discrete Cosine Transform (2-D DCT) is used to compress image intensities in the JPEG standard and MC-residuals in many video coding standards. However, prediction residuals have different spatial characteristics from image intensities [1, 2]. Therefore, it is of interest to study if transforms better than those used for image intensities can be developed for prediction residuals.

Recently, new transforms have been developed that can take advantage of locally anisotropic features in images [3, 4, 5, 6]. A conventional transform, such as the 2-D DCT or the 2-D Discrete Wavelet Transform (2-D DWT), is carried out as a separable transform by cascading two 1-D transforms in the vertical and horizontal dimensions. This approach does not take advantage of locally anisotropic features present in images because it favors horizontal or vertical features over others. The new transforms adapt to locally anisotropic features in images by performing the filtering along the direction where image intensity variations are smaller. This is achieved, for example, by directional lifting implementations of the DWT [6]. Even though most of the work is based on the DWT, similar ideas have been applied to DCT-based image compression [4].

Inspection of prediction residuals shows that locally anisotropic features are also present in prediction residuals. However, unlike in image intensities, a large number of pixels in prediction residuals have zero amplitude. Pixels with nonzero amplitude concentrate in regions which are difficult to predict, such as moving object boundaries, edges, or texture regions. Therefore a major portion of the signal in prediction residuals concentrates along such object boundaries and edges, forming 1-D structures along them. As a result, anisotropic features in prediction residuals typically manifest themselves as locally 1-D structures at various orientations. This is in contrast to image intensities, which have 2-D anisotropic structures. This difference can be quantified using an auto-covariance analysis [1].

We proposed 1-D directional block transforms for the MC-residual in [1] and showed potential gains achievable with one example set of such 1-D transforms on 8x8-pixel blocks within the H.264/AVC codec. In this paper, we provide further experimental results that show achievable gains using 1-D transforms defined on both 8x8-pixel blocks and 4x4-pixel blocks, which are available transform sizes in H.264/AVC. We also provide other useful information obtained from these experiments. These include the comparison of the achievable gains at relatively lower and higher picture qualities, the fraction of the bitrate used to code the selected transform for each block, and the average number of times each transform is selected.

The remainder of the paper is organized as follows. In Section 2, we discuss the 1-D directional transforms that were used in the experiments. Some details regarding the integration of these transforms into the H.264/AVC codec (JM reference software 10.2) are discussed in Section 3. We then present the experimental results and insights obtained from these results in Section 4. We conclude the paper in Section 5.

2. 1-D DIRECTIONAL TRANSFORMS

As described in Section 1, a large number of local regions in prediction residuals consist of 1-D structures, which follow object boundaries or edges present in the original image. It appears that using 2-D transforms with basis function that have 2-D support is not the best choice for such regions. Therefore we propose to use transforms with basis functions whose support follow the 1-D structures of the prediction residuals. Specifically, we propose to use 1-D directional transforms for prediction residuals. In our initial paper [1], we used 1-D transforms on 8x8-pixel blocks for the experiments. In this paper, we use 1-D transforms on both 8x8-pixel and 4x4-pixel blocks since both block sizes are available in H.264/AVC. We note that the idea of 1-D transforms for prediction residuals can also be extended to wavelet transforms [7].

The 1-D directional transforms that we use in our experiments are shown in Figure 1. Figure 1-a shows the first five 1-D block transforms defined on 8x8-pixel blocks. The remaining eleven are symmetric versions of these five and can be easily derived. Figure 1-b shows the first three 1-D block transforms defined on 4x4-pixel blocks. The remaining five are symmetric versions of these three and can be easily derived.
Each of the 1-D block transforms consists of a number of 1-D patterns which are all roughly directed at the same angle. For example, all 1-D patterns in the fifth 1-D block transform defined on 8x8-pixel blocks or the third 1-D block transform defined on 4x4-pixel blocks are directed towards south-east. The angle is different for each of the 1-D block transforms and altogether they cover 180°, for both 8x8-pixel blocks and 4x4-pixel blocks. Each 1-D pattern in any 1-D block transform is shown with arrows in Figure 1 and defines a group of pixels over which a 1-D DCT is performed.

3. INTEGRATION OF 1-D TRANSFORMS INTO THE H.264/AVC CODEC

To integrate these transforms into a codec, a number of related aspects need to be carefully designed. These include implementation of the transforms, quantization of the transform coefficients, coding of the quantized coefficients, and coding of the side information which indicates the selected transforms for each block.

In H.264/AVC, transform and quantization are merged together so that both of these steps can be implemented with integer arithmetic using addition, subtraction and bitshift operations. This has many advantages including the reduction of the computational complexity [8]. The computational complexity is not our concern in this paper, and we use floating point operations for these steps (including the implementation of 2-D DCT). This does not change the results. We note that it is possible to merge the transform and quantization steps of our proposed 1-D transforms so that these steps can also be implemented with integer arithmetic.

3.1. Coding of 1-D transform coefficients

We use CAVLC (context-adaptive variable-length codes) mode to perform entropy coding. To code the quantized transform coefficients in this mode, H.264/AVC zigzag-scans these coefficients. It uses a run-length coding algorithm to code the positions of the nonzero quantized coefficients and a context-adaptive method to code the amplitudes. Note that these methods are tailored to the statistics of the transform coefficients which are obtained from an approximation of the 2-D DCT. Ideally, it would be best to design a new method which is adapted to the characteristics of the coefficients of the proposed 1-D transforms. We are working on such designs. For the experiments in this paper, however, we use the same method of H.264/AVC with the exception of the scan. We use different scans for each of the 1-D transforms.

Figure 2-b shows the scans for the 1-D transforms defined on 4x4-pixel blocks shown in Figure 1-b. These scans were designed so that coefficients less likely to be quantized to zero are closer to the front of the scan and coefficients more likely to be quantized to zero are closer to the end of the scan. Scans for the remaining 1-D transforms defined on 4x4 blocks are symmetric versions of those in Figure 2-b.

For transforms defined on 8x8-pixel blocks, H.264/AVC generates four length-16 scans instead of one length-64 scan, when entropy coding is performed in CAVLC mode. Figure 2-a shows the four length-16 scans for each of the 1-D transforms defined on 8x8-pixel blocks shown in Figure 1-a. These scans were designed based on two considerations. The first is that coefficients less likely to be quantized to zero are closer to the front of the scan and coefficients more likely to be quantized to zero are closer to the end of the scan. The second consideration is that neighbouring 1-D patterns are grouped into one scan. The 1-D structures in prediction residuals are typically concentrated in one region of the 8x8-pixel block and the 1-D transform coefficients representing them will therefore be concentrated in a few neighbouring 1-D patterns. Hence, grouping neighbouring 1-D patterns into one scan enables capturing those 1-D transform coefficients in as few scans as possible. More scans that consist of all zero coefficients can lead to more efficient overall coding of coefficients.

3.2. Coding of side information

The selected transform for each block is encoded as side information using variable-length codes (VLC). We performed experiments to estimate the probabilities of selection of transforms. From these experiments and the effect of the side information on the increase of the overall bitrate, we decided to use the following codeword assignments. If a macroblock uses 8x8-pixel transforms, then for each 8x8-pixel block, the 2-D DCT is represented with a 1-bit codeword, and each of the sixteen 1-D transforms is represented with a 5-bit codeword.

If a macroblock uses 4x4-pixel transforms, then for each 4x4-pixel block, the 2-D DCT can be represented with a 1-bit codeword.
and each of the eight 1-D transforms can be represented with a 4-bit codeword. Alternatively, four 4x4-pixel blocks within a single 8x8-pixel block can be forced to use the same transform, which allows to represent the selected transforms for these four 4x4-pixel blocks with a single 4-bit codeword. This reduces the average bitrate for the side information but will also reduce the flexibility of transform choices for 4x4-pixel blocks. We use the alternative method in our experiments because it usually gives slightly better results.

4. EXPERIMENTAL RESULTS

We present experimental results to illustrate the performance of the proposed transforms within the H.264/AVC codec (JM reference software 10.2). We use 9 QCIF resolution sequences at 30 frames-per-second (fps), 4 CIF resolution sequences at 30 fps, and one 720p resolution sequence at 60 fps. We encode 20 frames for the 720p sequence and 180 frames for all other sequences.

Some of the important encoder parameters are as follows. The first frame is encoded as an I-frame, and all remaining frames are encoded as P-frames. Quarter-pixel-resolution full-search motion-estimation with all available block-sizes are used. Entropy coding is performed in CAVLC mode. Rate-distortion (RD) optimization is done in the high-complexity mode, which encodes all possible macroblock coding options and chooses the best. Selection of the best transform for each block is also performed in an RD optimized way by encoding each block with every available transform and choosing the transform with the smallest RD cost.

We encode all sequences at four different picture quality levels (with quantization parameters 24, 28, 32 and 36) using 6 different encoders, each having access to a different set of transforms. These encoders are

- 4x4-dct
- 4x4-1D (includes 4x4-dct)
- 8x8-dct
- 8x8-1D (includes 8x8-dct)
- 4x4-and-8x8-dct
- 4x4-and-8x8-1D (includes 4x4-and-8x8-dct).

4.1. Bjontegaard-Delta (BD) bitrate results

To present the gains achievable with the 1-D transforms, we compare the above encoders using the BD-bitrate metric [9]. This metric measures the average bitrate savings of one encoder with respect to another encoder within the picture quality range covered by four quantization parameters, which in our case roughly corresponds to 30dB to 40dB.

The BD-bitrate savings results are shown in Figure 3. We compare several encoders which have access to same block size transforms. Figure 3-a compares 8x8-dct to 8x8-1D. This is what was reported in [1]. Figure 3-b compares 4x4-dct to 4x4-1D and Figure 3-c compares 4x4-and-8x8-dct to 4x4-and-8x8-1D. The average bitrate savings are 11.4% for Figure 3-a, 4.1% for Figure 3-b, and 4.8% for Figure 3-c.

The bitrate saving is largest when comparing encoders which have access to only 8x8-pixel block transforms and smallest when comparing encoders which have access to only 4x4-pixel block transforms. This is in part because the distinction between 2-D transforms and 1-D transforms becomes less when block-size is reduced. For example, for 2x2-pixel blocks, the distinction would be even less, and for the extreme case of 1x1-pixel blocks, there would be no difference at all.

4.2. Bitrate savings at high and low picture qualities

The BD-bitrate metric gives savings averaged over a wide range of picture qualities. To determine bitrate savings for a specific range, Figure 4 shows RD curves of several encoders for the foreman sequence at QCIF resolution. It can be observed that bitrate savings are higher at high picture qualities than at low picture qualities. For the comparison between 4x4-and-8x8-dct and 4x4-and-8x8-1D, the savings are 2.2% at 32dB, 8.7% at 38dB and 6.3% when averaged over the range of 32dB to 40dB. This trend, which implies larger savings at higher picture qualities and smaller savings at lower picture qualities, is present in many sequences. This is in part because at higher picture qualities the bitrate used for coding the side information becomes a smaller fraction of the entire bitrate.

4.3. Bitrate used for coding the side information

The average fraction of the total bitrate used for coding the side information is 3.6% for 4x4-1D, 5.9% for 8x8-1D and 4.4% for 4x4-and-8x8-1D. These are averages obtained from all sequences at all picture qualities. The lowest fraction is used by 4x4-1D and the highest fraction is used by 8x8-1D. The reason is that 4x4-1D uses a 1-bit (2-D DCT) or a 4-bit (1-D transforms) codeword for every four 4x4-pixel blocks with coded coefficients, and the 8x8-1D uses a 1-bit or a 5-bit codeword for every 8x8-pixel block with coded coefficients. In addition, the probability of using a 1-D transform is
higher in 8x8-1D than in 4x4-1D.

4.4. Probability of selection of transforms

Information about how often each transform is selected is presented in Figure 5. These numbers depend on the encoded sequence and picture qualities. Figure 5 shows the probabilities obtained from all sequences for the 4x4-and-8x8-1D encoder at low and high picture qualities. It can be observed that the 2-D DCT’s are chosen more often than the other transforms. A closer inspection reveals that using a 1-bit codeword to represent the 2-D DCT and a 4-bit codeword (5-bit in case of 8x8-pixel transforms) to represent the 1-D transforms is consistent with the numbers presented.

At low picture qualities, the probability of selection is 58% for both 2-D DCT’s, and 42% for all 1-D transforms. At high picture qualities, the probabilities are 38% for both 2-D DCT’s, and 62% for all 1-D transforms. The 1-D transforms are chosen more often at higher picture qualities. Choosing the 2-D DCT costs 1-bit, and any of the 1-D transforms 4-bits (5-bits for 8x8-pixel block transforms). This is a smaller cost for 1-D transforms at high bitrates relative to the available bitrate.

5. CONCLUSIONS

We report experimental results that show achievable gains when 1-D directional transforms are used in addition to the 2-D DCT within the state-of-the-art H.264/AVC codec. Experimental results also provide other useful information in understanding the use of 1-D transforms within the codec, such as the bitrate savings at higher and lower picture qualities, the fraction of bitrate used to code the side information which indicates selected transforms, and probabilities of selection for all transforms. Future research efforts focus on designing more efficient coefficient coding methods adapted to the characteristics of 1-D transforms, and on applying 1-D transforms on other prediction residuals.

6. REFERENCES