

Transforms for the Motion Compensation Residual

Fatih Kamisli and Jae S. Lim
Research Laboratory of Electronics
Massachusetts Institute of Technology
E-mail: fkamisli@mit.edu

Abstract—The Discrete-Cosine-Transform (DCT) is the most widely used transform in image and video compression. Its use in image compression is often justified by the notion that it is the statistically optimal transform for first-order Markov signals, which have been used to model images. In standard video codecs, the motion-compensation residual (MC-residual) is also compressed with the DCT. The MC-residual may, however, possess different characteristics from an image. Hence, the question that arises is if other transforms can be developed that can perform better on the MC-residual than the DCT. Inspired by recent research on direction-adaptive image transforms, we provide an adaptive auto-covariance characterization for the MC-residual that shows some statistical differences between the MC-residual and the image. Based on this characterization, we propose a set of block transforms. Experimental results indicate that these transforms can improve the compression efficiency of the MC-residual.

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

I. INTRODUCTION

The Discrete-Cosine-Transform (DCT) is the most widely used transform in image and video compression. Its use in image compression is often justified by its being the statistically optimal transform for a class of signals, known as the first-order Markov signals, which have been used to model images. In video coding, however, what is often transformed is the motion-compensation residual (MC-residual). In standard video codecs, the MC-residual is also transformed with the DCT. The MC-residual is related to images from which it has been obtained. However, its spatial characteristics may be different from that of an image. A number of studies report that the statistical characteristics of the MC-residual have some differences from those of images [1], [2], [3]. Hence, the question that arises is what those characteristics are, if the DCT can effectively exploit them, and if we can develop other transforms that can exploit these characteristics better.

Typically, motion compensation works well in smooth and slowly moving regions, and the MC-residual in such regions is small. Regions where motion compensation fails are occlusion regions. A major portion of regions where motion compensation works to some extent are object boundaries. Since motion-compensated prediction can only account for translational motion, whereas real-world objects also have other motions such as rotation, the shapes of objects tend to change slightly from frame to frame. As a result, the prediction around object boundaries is not successful. Therefore the high-magnitude pixels in the MC-residual tend to concentrate along object boundaries exposing one-dimensional structures in the

MC-residual. An example is shown in Figure 1. It appears that using two-dimensional transforms that have basis functions with square support is not the best choice.

Recently, there has been a great deal of research on transforms that can take advantage of locally anisotropic features in images [4], [5], [6], [7]. Conventionally, the 2-D DCT or the 2-D Discrete Wavelet Transform (DWT) is carried out as a separable transform by cascading two 1-D transforms in the vertical and horizontal directions. This approach does not take advantage of the locally anisotropic features present in images because it favors horizontal or vertical features over others. The main idea of these other approaches is to adapt to locally anisotropic features by performing the filtering along the direction where the image intensity variations are smaller. This is achieved by performing filtering and subsampling on oriented sublattices of the sampling grid [6], by directional lifting implementations of the wavelet transform [5], or by various other means. Even though most of the work is based on the wavelet transform, applications of similar ideas to DCT-based image compression have also been made [7]. However, it appears that the applicability of these ideas to modeling and compressing the MC-residual have not been investigated.

In this paper, we develop block transforms for the MC-residual. Using insights obtained from the research on direction-adaptive image transforms, we investigate how locally anisotropic features of images affect the MC-residual. In the next section, we obtain adaptive auto-covariance characterizations of the MC-residual and the image, which reveal some statistical differences between the MC-residual and the image.



Fig. 1. Motion compensation residual (Mobile sequence at CIF resolution, frame 8 and 10)

Based on this characterization, we propose in Section III a set of block transforms which can be used, together with the 2-D DCT, to compress the MC-residual. Experimental results, presented in Section IV, demonstrate that these transforms can improve the compression efficiency of the MC-residual.

II. CHARACTERIZATION OF THE MOTION COMPENSATION RESIDUAL

Characterizations of the MC-residual are more focused on representing its auto-covariance with functions that provide a close fit to experimental data [1], [2], [3]. These studies have used one global model for the entire MC-residual. However, the lessons learned from the research on direction-adaptive image transforms indicate that adapting the representation to local anisotropic features in images may improve the compression performance. We consider how such local anisotropic features of images affect the MC-residuals. To analyze this, we characterize the auto-covariance of the image and the MC-residual using a more generalized version of the conventionally used separable first-order Markov model. The separable model and the generalized model are given below in (1) and (2), respectively.

$$R(I, J) = \rho_1^{|I|} \rho_2^{|J|}. \quad (1)$$

$$R(\theta, I, J) = \rho_1^{|I \cos(\theta) + J \sin(\theta)|} \rho_2^{|-I \sin(\theta) + J \cos(\theta)|}. \quad (2)$$

The generalized model is a rotated version of the separable model, where θ represents the amount of rotation. Setting it to zero reduces the generalized model to the separable model.

We estimate the parameters ρ_1 and ρ_2 for the separable model, and the parameters ρ_1 , ρ_2 and θ for the generalized model from blocks of 8x8 pixels of an image (Mobile sequence, frame 10) and its MC-residual. We use the biased estimator to estimate the auto-covariance of the blocks and then find the parameters ρ_1 , ρ_2 and θ that minimize the mean-square-error between the auto-covariance estimate and the models in (1) and (2). Figures 2-a and 2-b show scatter plots of ρ_1 and ρ_2 estimated from the image for the separable and the generalized auto-covariance models. Figures 2-c and 2-d show scatter plots of the same parameters estimated from the MC-residual. We use the Mobile sequence for these experiments because this sequence has various moving objects causing unsuccessful motion-compensated prediction at object boundaries. Characteristics similar to those in the plots of Figure 2 have been observed with other sequences as well.

The scatter plots of ρ_1 and ρ_2 obtained from the image using the separable and the generalized auto-covariance models are plotted in Figures 2-a and 2-b. In the plot of the generalized model (Figure 2-b), the region along the $\rho_1 = \rho_2$ line is less populated by data points relative to the plot of the separable model (Figure 2-a). On the other hand, there are almost no data points beyond the $\rho_1 = 0.8$ and $\rho_2 = 0.8$ lines in the separable model. From these observations, we can see that in the generalized case, the data points lying below the $\rho_1 = \rho_2$ line have been shifted along the ρ_1 axis towards the right, and the data points lying above the $\rho_1 = \rho_2$ line have been shifted along the ρ_2 axis towards the top. This

implies that by allowing an additional parameter θ in the model, higher correlation coefficients ρ_1 or ρ_2 are more likely. The parameter θ adjusts itself such that either ρ_1 or ρ_2 point along more regular variations than in the separable model, which is consistent with the resampling and lifting methods in [4] and [5].

From the plots for the MC-residual in Figure 2, we can observe that in the plot with the separable auto-covariance model (Figure 2-c), the points seem to fill up evenly a region between a $\rho_2 = k/\rho_1$ curve and the axes. In the plot with the generalized auto-covariance model (Figure 2-d), the points seem to concentrate towards the tails of that curve, forming semi-disks centered on the axes. Better concentration of data points in a region means better predictability of data, which in turn implies that the model used in (2) can provide a more faithful characterization of the data. The significant improvement of the characterization with an additional parameter, θ , is important since a better parametrization of a source can potentially lead to a better compression of the source.

Figure 2 also illustrates the effect of the locally anisotropic features of images on the MC-residual. Consider the generalized auto-covariance characterization of the image and the MC-residual (Figure 2-b and 2-d). While the characterization of the image has data points which, if roughly described, fill up two disks tangent to the ρ_1 and ρ_2 axes, the characterization of the MC-residual seems to have its data points concentrated in two semi disks centered on those axes. In other words, the data points move closer to the axes in the MC-residual case. This means that given any data point (ρ_1, ρ_2) in the image characterization, the smaller covariance factor becomes even smaller in the MC-residual characterization. This is how locally anisotropic features of images affect or propagate to the MC-residual. It is a major difference in the statistical characteristics between the image and the MC-residual. Therefore, direction-adaptive transforms proposed for images may not work as well on MC-residuals. Indeed, in Section IV, we compare the performance of the direction-adaptive block-based image transform in [7], with the transforms we propose. The results show the superiority of the proposed transforms on the MC-residual.

III. ONE-DIMENSIONAL TRANSFORMS

The ρ_1 vs ρ_2 scatter plot of the MC-residual obtained with the generalized model in Figure 2-d indicates that often one of the two correlation coefficients, either ρ_1 or ρ_2 , is significantly larger than the other. Hence, we propose to decorrelate the data along the direction with the larger correlation coefficient. Unlike the direction-adaptive transforms applied to images, we choose not to perform any decorrelation along the other direction with the smaller correlation coefficient. For low correlation coefficients, decorrelation may not perform well. It may even worsen the energy-compaction, especially for such signals as the MC-residual. Unlike an image, the energy of the MC-residual is not uniformly distributed in the spatial domain. Even within a block, many pixels may have zero intensity and the energy may often be concentrated in a region of the

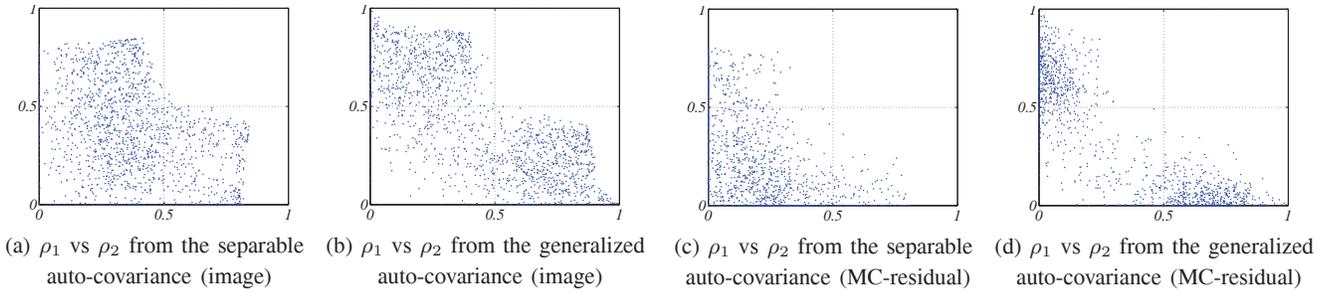


Fig. 2. Scatter plots of estimated parameters ρ_1 and ρ_2 using the separable and generalized auto-covariance models from an image (mobile sequence at CIF resolution, frame 10) and an MC-residual (mobile sequence at CIF resolution, frame 8 and 10).

block (see Figure 1). Performing a spatial transform with a support on the whole block may not produce the best energy compaction.

The transforms that we propose to use for the MC-residual, in addition to the 2-D DCT, are shown in Figure 3. They are one-dimensional DCT's over groups of pixels in 8×8 blocks. One group of pixels, in the first transform for example (most left in Figure 3), consists of the eight leftmost pixels sitting on top of each other. The groups of pixels are chosen such that they form lines roughly pointing at one direction, which is the direction of the large correlation coefficient. We choose sixteen directions that cover 180° and we have sixteen transforms. Only the first five are shown in Figure 3 due to space limitations. Each group of pixels for each of the shown transforms are represented with arrows that traverse those pixels.

The scanning patterns of the resulting transform coefficients depend on the orientation of the transforms. They are designed such that lower frequency coefficients are scanned before higher frequency coefficients and such that they are continuous.

The best transform for each block is chosen in a Rate-Distortion (RD) optimized manner. The block is encoded with each available transform, including the 2-D DCT. Then a cost function is formed using a linear combination of the distortion (MSE) of the block and the number of bits spent on the block (both to encode the quantized coefficients and the side information) for each transform. The transform which has the smallest cost function is chosen for each block.

The chosen transform for each block is encoded using variable-length-codes (VLC). Since the 2-D DCT is chosen about half of the time on average, we use a 1-bit codeword to represent the 2-D DCT and 5-bit codewords to represent each of the sixteen one-dimensional transforms. This choice was made because of its simple implementation and because it is close to the optimal Huffman codes for the average frequencies of the transforms. More sophisticated methods may result in improvements in coding the side information.

IV. EXPERIMENTAL RESULTS

We present some experimental results to illustrate the performance of the proposed transforms within the H.264/AVC

reference software. We use QCIF resolution sequences at 30 frames-per-second. Some of the important encoder parameters are as follows. The first frame is encoded as an I-frame, and all the remaining frames are coded as P-frames. Quarter-pixel-resolution full-search motion-estimation with adaptive block-sizes ($16 \times 16, 16 \times 8, 8 \times 16, 8 \times 8$) are used. Entropy coding is performed with context-adaptive variable-length-codes (CAVLC).

The results of the experiments are shown with the Bjontegaard-Delta (BD) bitrate metric [8] using the following quantization parameters: 24, 28, 32, 36. The BD-bitrate metric indicates the average bitrate savings (in terms of percentage) of the codec with the proposed transforms with respect to the codec with the conventional transform, the 2-D DCT. Figure 4-a shows the results obtained with the proposed transforms. The upper graph shows results that are obtained when taking the side information, which signals the chosen transform for each block, into account. The lower graph shows results obtained without taking the side information into account. The lower graph is included to show the ultimate performance of the proposed transforms and scans. These performances may be achievable if the side information is coded more intelligently and becomes negligibly small. For QCIF resolution sequences, the proposed transforms can achieve bitrate savings upto 25.8% with an average of 12%. When the side information bits are neglected, the ultimate achievable bitrate savings can be as large as 40.3% with an average of 21.7%.

We have also performed experiments to report the performance of the direction-adaptive image transforms in [7] on the MC-residual. In [7], the performance on images is reported. The transforms in [7] are 2-D directional DCT's together with a DC separation and Δ DC correction method borrowed from [9]. For the experiments, we have complemented the six transforms in [7] with another eight transforms to achieve finer directional adaptivity, which is comparable to the adaptivity of our proposed transforms. We also used the scans proposed in [7]. All remaining parts of the system were unchanged. The results of these experiments are shown in Figure 4-b. The average bitrate savings are 4.8%, which is less than half of the 12% achieved with the proposed transforms.

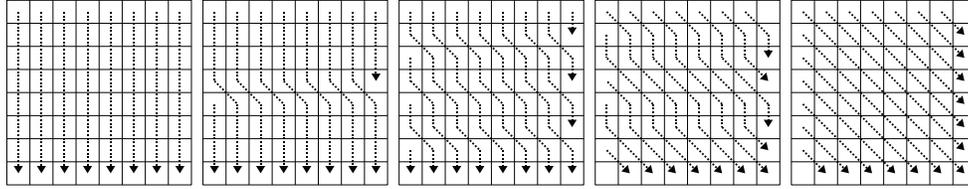


Fig. 3. One-dimensional transforms for the motion compensation residual. The arrows traversing a set of pixels indicate a one-dimensional DCT on those pixels. A total of 16 transforms, directed at different directions cover 180° . Only the first five transforms are shown here. The remaining transforms are symmetric versions of these five and can be easily derived.

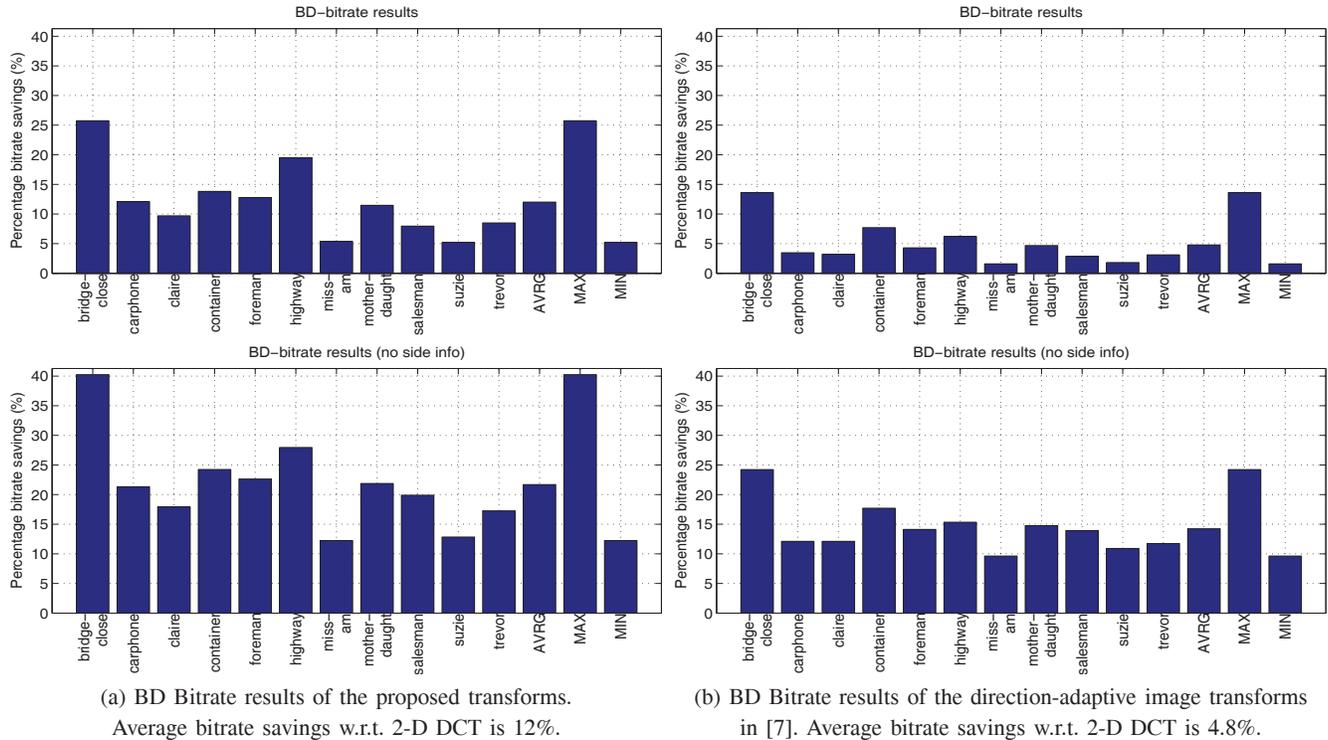


Fig. 4. Bjontegaard-Delta Bitrate results for QCIF resolution sequences.

V. CONCLUSION

We have analyzed the effect of locally anisotropic features of images on the motion-compensation residual (MC-residual). The analysis reveals statistical differences between the MC-residual and the image. We have proposed transforms based on this analysis and reported experimental results which indicate that these transforms can improve the compression efficiency of the MC-residual. Future research efforts focus on improving these transforms and on investigating the effects of locally anisotropic features of images on other prediction residuals such as the intra prediction residual in H.264/AVC, resolution enhancement residual in scalable video coding and the disparity compensation residual in multiview video coding.

REFERENCES

- [1] C.-F. Chen and K. Pang, "The optimal transform of motion-compensated frame difference images in a hybrid coder," *Circuits and Systems II: Analog and Digital Signal Processing, IEEE Transactions on*, vol. 40, no. 6, pp. 393–397, Jun 1993.
- [2] W. Niehsen and M. Brunig, "Covariance analysis of motion-compensated frame differences," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 9, no. 4, pp. 536–539, Jun 1999.
- [3] K.-C. Hui and W.-C. Siu, "Extended analysis of motion-compensated frame difference for block-based motion prediction error," *Image Processing, IEEE Transactions on*, vol. 16, no. 5, pp. 1232–1245, May 2007.
- [4] E. Le Pennec and S. Mallat, "Sparse geometric image representations with bandelets," *Image Processing, IEEE Transactions on*, vol. 14, no. 4, pp. 423–438, April 2005.
- [5] C.-L. Chang and B. Girod, "Direction-adaptive discrete wavelet transform for image compression," *Image Processing, IEEE Transactions on*, vol. 16, no. 5, pp. 1289–1302, May 2007.
- [6] V. Velisavljevic, B. Beferull-Lozano, M. Vetterli, and P. Dragotti, "Directionlets: anisotropic multidirectional representation with separable filtering," *Image Processing, IEEE Transactions on*, vol. 15, no. 7, pp. 1916–1933, July 2006.
- [7] B. Zeng and J. Fu, "Directional discrete cosine transforms for image coding," *Multimedia and Expo, 2006 IEEE International Conference on*, pp. 721–724, 9–12 July 2006.
- [8] G. Bjontegaard, "Calculation of average psnr differences between rd-curves," *VCEG Contribution VCEG-M33*, April 2001.
- [9] P. Kauff and K. Schuur, "Shape-adaptive dct with block-based dc separation and dc correction," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 8, no. 3, pp. 237–242, Jun 1998.