Prediction of DCT Coefficients Considering Motion Compensation Error Distributions

Julia Schmidt
Gottfried Wilhelm Leibniz
Universitaet Hannover
Hannover, Germany
Email: schmidt@tnt.uni-hannover.de

Bernd Edler
International Audio Laboratories
Erlangen[1],
Erlangen, Germany
Email: Bernd.Edler@audiolabs-erlangen.de

Joern Ostermann
Gottfried Wilhelm Leibniz
Universitaet Hannover
Hannover, Germany
Email: office@tnt.uni-hannover.de

***Abstract*— Current video coding techniques use a Discrete Cosine Transform (DCT) to reduce spatial correlations within the motion estimation residual. Often the correlation cannot be completely eliminated leaving the transform coefficients statistically dependent. The presented paper proposes a method to predict these coefficients on a block level by using the distribution of the prediction error variance to improve coding efficiency. First experiments lead to a reduction in bit rate by 1.83% when compared to the standard JM 17.2 implementation results.**

I. Introduction

Hybrid video coders like H.264|AVC [1] use different intra and inter prediction techniques to exploit spatial and temporal dependencies. Only the difference between original and predicted images is then transmitted, leading to a gain in terms of needed bandwidth when compared with the sending of a whole picture. The computed residual however still contains spatial dependencies, which are often exploited by transforms such as the DCT. The resulting transform coefficients are eventually quantized and entropy coded [2][3]. One way to exploit temporal correlations is motion compensation. The main assumption is constant motion within a block. This motion is represented by a motion vector. Through this vector the next position of an object is approximated, so that the next frame can be estimated. The error resulting from this prediction was thought to be evenly distributed. Figure 1 on the other hand indicates the real distribution of the residual: The difference between motion compensated and original image is bigger towards the edges of the block, which makes it possible to model the variance of the error [4][5].

The spatially non-constant variance in the DCT input blocks leads to correlations between transform coefficients within the DCT output blocks, as had also been observed in audio coding [7]. To exploit these correlations this paper proposes a prediction mechanism for the transform coefficients within blocks, based on the distribution of the motion compensation error. In Section 2 the algorithm is explained, followed by experimental results in Section 3 and conclusions in Section 4.

II. Methodology

In our model, we assume the DCT input x_n to be a signal r with a constant variance multiplied by an error distribution function w as motivated by [4] and Figure 2:

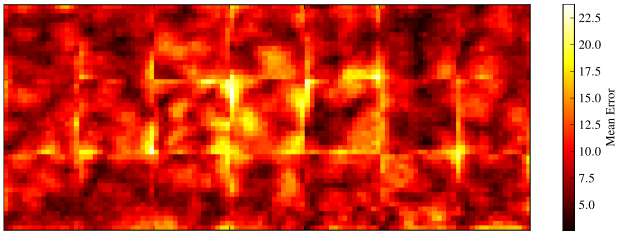

Fig. 1. Detail of the motion compensation error for the Kimono sequence [4]

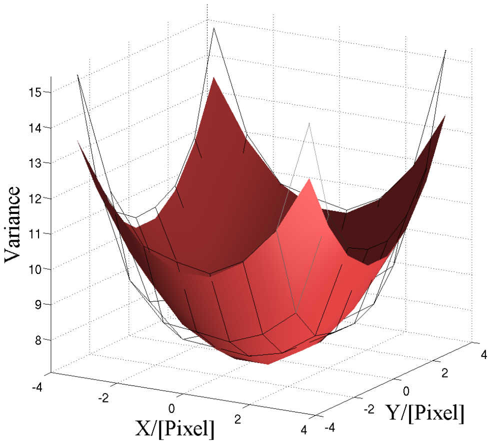

Fig. 2. Exemplary approximation of variance of Motion Compensation Error within a 8x8 block for Kimono sequence [4]

$$x_n = r_n w_n. \tag{1}$$

To calculate the prediction coefficients in the frequency domain, pairwise correlations in the DCT output X_k are used, which are elements of a covariance matrix. In the following, the frequency domain correlations are derived for a 1D DCT of length N, where the spectral coefficients X_k are

$$X_k = u_k \sqrt{\frac{2}{N}} \sum_{n=0}^{N-1} x_n cos \frac{\pi}{N} k(n + \frac{1}{2}) \tag{2}$$

[1] A joint institution of Fraunhofer IIS and Universitaet Erlangen-Nuernberg

$$\text{with } u_k = \begin{cases} \sqrt{\frac{1}{2}} & \text{for } k = 0 \\ 1 & \text{else} \end{cases} . \tag{3}$$

As the Discrete Fourier Transform (DFT) is connected to the DCT by normalization and symmetric extension and has well known correspondences in time and frequency domain, it can be used to investigate the behavior of the DCT. To do so, first a symmetric extension, which we from now on will indicate by a tilde, has to be performed:
$\tilde{x}_n = \tilde{x}_{2N-1-n} = x_n \ , \quad 0 \leq n \leq N-1$
with
$\tilde{x}_n = \tilde{r}_n \tilde{w}_n \ , \ \tilde{r}_n = \tilde{r}_{2N-1-n} = r_n \ , \tilde{w}_n = \tilde{w}_{2N-1-n} = w_n.$

In a second step a DFT whose base function has a time offset of half a sample interval, a so called odd-time DFT or shifted DFT(SDFT), of length $2N$ has to be applied:

$$\begin{aligned} \tilde{X}_k = & \ \textstyle\sum_{n=0}^{2N-1} \tilde{x}_n e^{-j2\pi k(n+\frac{1}{2})/2N} \\ = & \ \textstyle\sum_{n=0}^{2N-1} \tilde{x}_n cos\frac{\pi}{N}k(n+\frac{1}{2}), \end{aligned} \tag{4}$$

resulting in the following relation of coefficients

$$\tilde{X}_k = \frac{\sqrt{2N}}{u_k} X_k, 0 \leq k < N. \tag{5}$$

From symmetry relations in the SDFT base function follows:
$\tilde{X}_k = \tilde{X}_{4N+k} = \tilde{X}_{4N-k}$, $\tilde{X}_{2N+k} = -\tilde{X}_k$,
$\tilde{X}_N = \tilde{X}_{3N} = 0$

Therefore $\tilde{X}_k$ can be calculated as a 4N-cyclic convolution in the spatial domain. Taking into account the symmetries and periodicities of the SDFT base functions, the formula for $\tilde{X}_k$ has to be

$$\begin{aligned} \tilde{X}_k = & \ \textstyle\frac{1}{4N}\sum_{l=0}^{4N-1} \tilde{R}_l \tilde{W}_{k-l}^{(4N)} \\ = & \ \textstyle\frac{1}{2N}\sum_{l=0}^{N-1} \tilde{R}_l q_{k,l} \end{aligned} \tag{6}$$

with

$$q_{k,l} = \begin{cases} \tilde{W}_k & \text{for } l = 0 \\ \tilde{W}_{|k-l|} + \tilde{W}_{k+l} & \text{for } k > 0 \cap k+l < N \\ \tilde{W}_{|k-l|} & \text{for } k > 0 \cap k+l = N \\ \tilde{W}_{|k-l|} - \tilde{W}_{2N-k-l} & \text{for } k > 0 \cap k+l > N \end{cases} ,$$

In matrix notation, with vector $\tilde{R}$ containing the first N SDFT-coefficients of $\tilde{r}_n$ and a NxN-matrix $\mathbf{Q}$ containing the elements $q_{k,l}$ and therefore describing the influence of $\tilde{w}_n$, this can be written as

$$\tilde{\mathbf{X}} = \frac{1}{2N}\mathbf{Q}\tilde{\mathbf{R}} \tag{7}$$

The covariance matrix can now be obtained as

$$\begin{aligned} E\{\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T\} = & \ \textstyle\frac{1}{4N^2} E\{\mathbf{Q}\tilde{\mathbf{R}}\tilde{\mathbf{R}}^T\mathbf{Q}^T\} \\ = & \ \textstyle\frac{1}{4N^2} \mathbf{Q} E\{\tilde{\mathbf{R}}\tilde{\mathbf{R}}^T\}\mathbf{Q}^T \\ = & \ \textstyle\frac{1}{4N^2} \mathbf{Q}\mathbf{C}_{RR}\mathbf{Q}^T. \end{aligned} \tag{8}$$

For uncorrelated spectral coefficients $\tilde{R}_k$, C_{RR} is a diagonal matrix with elements

$$d_k = E\{\tilde{R}_k^2\} \tag{9}$$

and the elements of the covariance matrix are

$$E\left\{\tilde{X}_k\tilde{X}_m\right\} = \frac{1}{4N^2}\sum_{l=0}^{N-1} d_l q_{k,l} q_{m,l} \tag{10}$$

To derive the covariance at the DCT output X_k, normalization has to be taken into account, altering (10) to:

$$\begin{aligned} E\{X_k X_m\} = & \ \textstyle\frac{u_k u_m}{2N} E\left\{\tilde{X}_k\tilde{X}_m\right\} \\ = & \ \textstyle\frac{u_k u_m}{8N^3}\sum_{l=0}^{N-1} d_l q_{k,l} q_{m,l} \end{aligned} \tag{11}$$

Using the error distribution w, the coefficients are now predicted by values within the same block as opposed to the prediction technique presented in [6], which uses neighbouring blocks. Compared to [7], where irrelevance reduction is achieved by shaping the quantization noise by applying forward prediction, in our approach a redundancy reduction is realized by using backward prediction, i.e. predicting from already quantized values, leading to an uncorrelated white quantization noise. If the error distribution is symmetric, the special properties of the DCT lead to correlations which are zero for all odd index differences.

As this is the case with the error distribution functions presented in [4] and [5], only coefficients with even index differences need to be taken into account for prediction. An example with three predictor coefficients is depicted in Figure 3, which requires an extension of the above to 2D. The prediction for a coefficient $X_{k,l}$ denoted as O in Figure 3 is

$$\hat{X}_{k,l} = a_{k,l}X_{k-2,l} + b_{k,l}X_{k,l-2} + c_{k,l}X_{k-2,l-2} \qquad \text{for } k,l \geq 2, \tag{12}$$

with a, b and c being the prediction coefficients obtained as the solution of the following equation

$$\begin{pmatrix} \varphi_{(k-2,l),(k-2,l)} & \varphi_{(k,l-2),(k-2,l)} & \varphi_{(k-2,l-2),(k-2,l)} \\ \varphi_{(k-2,l),(k,l-2)} & \varphi_{(k,l-2),(k,l-2)} & \varphi_{(k-2,l-2),(k,l-2)} \\ \varphi_{(k-2,l),(k-2,l-2)} & \varphi_{(k,l-2),(k-2,l-2)} & \varphi_{(k-2,l-2),(k-2,l-2)} \end{pmatrix} \cdot \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} \varphi_{(k,l),(k-2,l)} \\ \varphi_{(k,l),(k,l-2)} \\ \varphi_{(k,l),(k-2,l-2)} \end{pmatrix}, \tag{13}$$

in which $\varphi_{(k,l),(i,j)}$ denotes the covariance $E\{X_{k,l}X_{i,j}\}$. For the border cases with $k < 2$ and $l < 2$, $X_{k,l}$ only depends on 1 or 0 previously transmitted transform coefficients. Due to the needed neighbouring transform coefficients, a prediction for the first four upper left transform coefficients isn't possible, leaving the corresponding transform coefficients zero.

If the motion compensation error distribution does not show symmetry characteristics, (12) has to be changed accordingly.

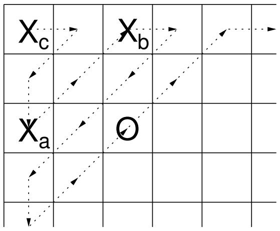

Fig. 3. To predict coefficient O, coefficients marked with X are used.

TABLE I
RESULTS FROM INTEGRATION INTO JM 17.2 FOR QPS 20, 26, 28, 32, 36 COMPUTED THROUGH USE OF BJØNTEGAARD DELTA

Testsequence	Approach	Bit rate [%]	PSNR [dB]
Kimono	pred	-0.9870	0.0438
	pred+sz	-0.7161	0.0244
	pred+oc	-0.1977	0.0092
Raven	pred	-0.9198	0.0453
	pred+sz	-0.7021	0.0357
	pred+oc	-0.2409	0.0133
BasketballDrive	pred	-0.0804	0.0023
	pred+sz	0.2056	-0.0069
	pred+oc	0.6347	-0.0203

III. EXPERIMENTAL VALIDATION

The described algorithm was implemented in the H.264|AVC reference software JM17.2 and tested on three sequences, one with a relatively small motion prediction error (Kimono), one with lots of motion and a higher prediction error (BasketballDrive) and one with less but harder to predict motion (Raven). The coding scheme used for the presented results is IPPP with a fixed block size of 8x8 (extended profile). Rate-Distortion-Optimized quantization was disabled because other functions are used here, as well as IntraInInter. Results in Table I, which are computed through the use of the Bjøntegaard Delta and show the difference to the JM 17.2 implementation, are achieved by three different variations of the algorithm:

- The base version (pred), in which three blocks of prediction coefficients a,b,c per frame are calculated,
- a version in which only the first 16 coefficients are predicted and the rest is set to zero (pred+sz),
- and a version where three blocks of prediction coefficients are calculated for usage within the whole sequence as opposed to just one frame (pred+oc).

A block diagram illustrates the process in Fig. 4. The input data for the prediction process are 8x8 blocks of the transformed and quantized motion estimation residual. Depending on which version is used, either three prediction coefficient blocks per frame or per sequence have to be transmitted, to decode the sent data properly. In the base version, the image is passed once to compute the covariance matrix. In a second pass the prediction coefficients derived from the matrix are used to predict the DCT coefficients. When rate-distortion-optimized quantization is used, the frame is passed several times either way, so that those passes could be used to execute the calculation and prediction. An implementation like this would not introduce a great amount of additional complexity, so that the overall encoding time basically stays the same. The information gained during the rate-distortion optimization process would also help to decide whether or not a prediction of coefficients is useful, so that losses like those caused by prediction in the BasketballDrive sequence would be prevented. As transform coefficients get smaller from the upper left side of a block to the lower right side, an idea to reduce computational effort and the data to be transmitted is to predict only the first 16 coefficients. This saves rate for the transmission of predictor coefficients. The results of this test can be found under pred+sz in Table I. To further reduce the data rate, the last variation on the original idea is to only calculate three prediction coefficient sets through averaging over all frames and use these coefficients throughout the whole sequence as described for the original mode version (pred+oc). A thorough look at the rate-distortion function implies, that the higher the bit rate, the less useful DCT prediction gets, whereas in regions with higher QPs the results show a reduction in bit rate in most of the cases. When using the Bjøntegaard Delta for the computation of QPs higher than 28, gains around 2.36% (computed with QP 28, 32, 36 for the base algorithm) when it comes to bit rate reduction are achieved for the Kimono sequence. If a side information of 8 bit per predictor coefficient using PCM is taken into consideration, the gain of the base implementation would of course be reduced, leaving an overall gain of around 1.83%. Lower side information rates are likely to be achieved by applying more sophisticated coding.

A comparison of the results from the tested sequences shows, that the higher the motion compensated prediction error, the harder a prediction of the transform coefficients becomes. The cause for this observation might be that the motion compensation error in this case is distributed in a way different from the symmetric assumption we made before. Reasons might be occlusions and lighting effects violating the symmetry assumption. An alternative computation of predictor coefficients which doesn't regard symmetry might help here. If only one prediction coefficient set is used for the whole sequence, a gain in bit rate and PSNR is still existing, but noticeably smaller than in the more suitable pred version.

The motion compensated prediction error is, in this implementation, only taken into account in terms of the symmetry characteristic.

IV. CONCLUSION

Prior work in the field of motion compensation assumed the prediction residual to be evenly distributed. As this is not the case, the remaining spatial dependencies can be used to predict the DCT output to eventually reduce the data to be transmitted. The presented paper points out a way to calculate the DCT coefficients using the symmetry characteristics of the motion

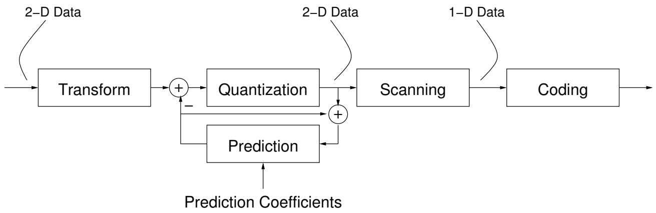

Fig. 4. Prediction process in the JM17.2 context. 2-D Data = 8x8 blocks

prediction error on a block level. Coefficients are predicted from preceding values within a block as opposed to the technique described in [6], where the surrounding blocks are used. The implementation of the algorithm provides promising results with gains around 1.83% data reduction compared to the JM 17.2 reference implementation when considering QPs larger than 28 and should be further examined.

References

[1] ITU-T Rec. H.264 and ISO/IEC 14496-10 (MPEG4-AVC), *Advanced Video Coding for Generic Audiovisual Services*, v1, May, 2003; v2, Jan. 2004; v3 (with FRExt), Sept. 2004; v4, July 2005.

[2] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, A. Luthra, *Overview of the* H.264/AVC *video coding standard*, IEEE Transactions on Circuits and Systems for Video Technology, vol. 13, no. 7, pp. 560-576, July 2003.

[3] F. Pereira, T. Ebrahimi, *The* MPEG-4 *Book*, Prentice Hall, ISBN 0-130-61621-4, 2002, chapter 8.2.3.2

[4] S. Klomp, M. Munderloh, J. Ostermann, *Block size dependent error model for motion compensation*, Proceedings ICIP 2010

[5] L. Falkenhagen, T. Wedi, “Improving Block-Based Disparity Estimation by Considering the Non-Uniform Distribution of the Estimation Error”, in *3D Structure from Multiple Images of Large-Scale Environments*, proceedings of SMILE workshop, pp. 93-108, June 6-7, 1998, Freiburg, Germany, ISBN 3-540-65310-4 Springer Verlag, 1998.

[6] A. Puri, R. L. Schmidt and B. G. Haskell, *Improvements in* DCT-*based video coding*, Proc. SPIE 3024, 676 (1997)

[7] J. Herre, *Temporal noise shaping, quantization and coding methods in perceptual audio coding: A tutorial introduction*, AES 17th International Conference on High Quality Audio Coding, Florence, Italy, September 2-5, 1999