Extending H.264/AVC by an adaptive coding of the prediction error

Matthias Narroschke

Institut für Informationsverarbeitung
University of Hannover, Appelstr. 9a, 30167 Hannover, Germany
narrosch@tnt.uni-hannover.de

Abstract. Standardized video coding algorithms are based on hybrid coding using motion compensated prediction and transform coding of the prediction error. The transform is efficient if the prediction error samples are correlated. For marginally correlated samples, the transform is inefficient. In this paper, adaptive prediction error coding in the spatial and in the frequency domain is investigated. In the spatial domain, scalar quantization and entropy coding using CABAC is applied. For each block of the prediction error, either standardized transform coding or spatial domain coding is used. The algorithm with lower rate-distortion costs is chosen. Furthermore, the combination of the adaptive coding with a displacement vector resolution of 1/8-pel is investigated, which further reduces the correlation between the prediction error samples and thus supports the coding in the spatial domain. For QCIF, CIF, and SDTV, the Y-PSNR is increased by up to 0.4 dB by the adaptive coding compared to the Key Technical Area Software of the ITU-T at the same bit rate. Combined with 1/8-pel displacement vector resolution, the increase is up to 1 dB.
Index Terms - H.264/AVC, video coding, prediction error coding

1 Introduction

In H.264/AVC [1, 2], the temporal redundancy is reduced by the use of predictive coding. In order to predict the current image signal s, a prediction image signal $\hat{s}$ is obtained from already reconstructed images by using block based motion estimation and motion compensated prediction. A displacement vector $\boldsymbol{D}$ is assigned to each block referring to the position of the block in an already reconstructed image. The prediction error signal e and the displacement vector $\boldsymbol{D}$ are encoded and transmitted. In order to reduce the spatial redundancy of the prediction error signal, transform coding is applied blockwise. The blocksize of the transform is 4x4 or 8x8 picture elements. The coefficients c are quantized with the step size controlled by the quantization parameter QP. The quantized coefficients c' are zigzag scanned starting at the DC-coefficient resulting a 1D array, see Figure 1. This array is

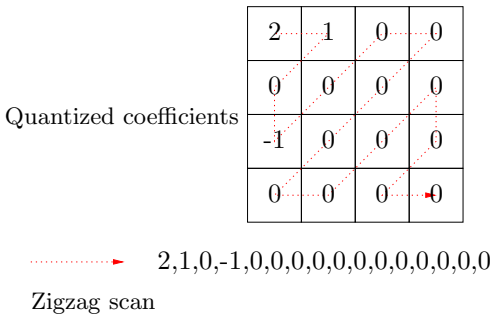

Fig. 1. Scan of quantized coefficients.

entropy-coded by the Context-Based Adaptive Binary Arithmetic Conding, CABAC [3], which is optimized for an array of coefficients with decreasing energy. The coding efficiency of the transform is high if the prediction error samples e are correlated. For marginally correlated samples the transform is inefficient. Therefore, the coding of the prediction error in the spatial domain is investigated in this paper as depicted in Figure 2. Each prediction error block is either coded in spatial or in the frequency domain. The information about the domain is sent as side information.
The remainder of this paper is structured as follows. Section 2 describes the quantizer design for the spatial domain. In Section 3, the entropy coding and in Section 4, the coding of the side information is described. The coder control is explained in Section 5. Experimental results and the conclusion are given in Sections 6 and 7.

2 Quantizer design for the spatail domain

H.264/AVC supports rate distortion optimized quantization of the coefficients c with two differ-

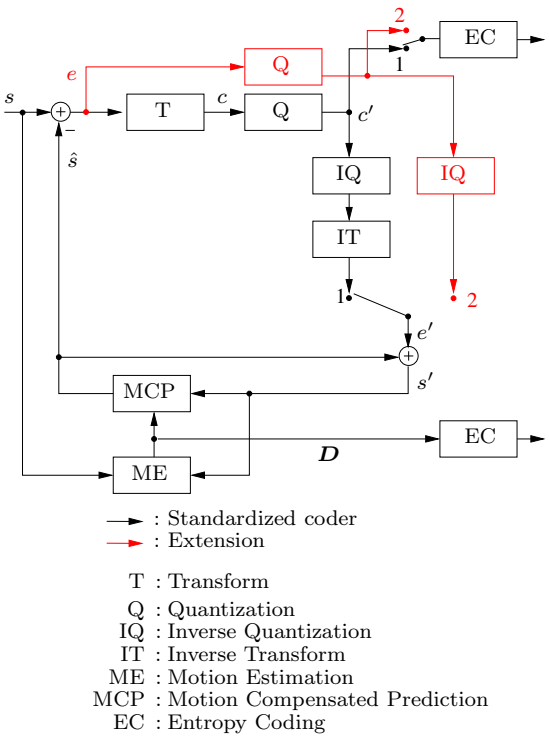

Fig. 2. Block diagram of the extended hybrid video encoder.

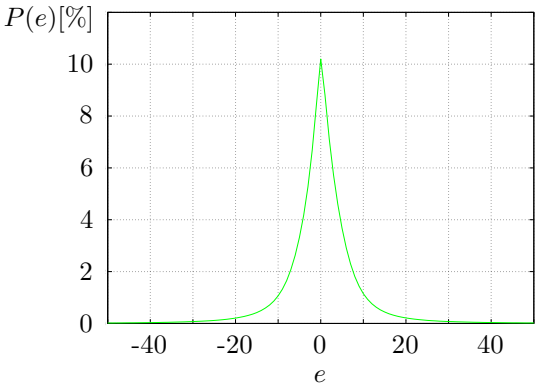

Fig. 3. Measured distribution of the prediction error signal e for the quantization parameter $QP = 26$.

ent distortion measures. The first measure is the mean squared quantization error, the second one is the subjectively weighted quantization error which is realized by the use of quantization weighting matrices. Correspondingly, two scalar rate distortion optimized quantizers are designed for the spatial domain. Since the distribution of the prediction error signal e is close to a Laplacian distribution, as shown in Figure 3, a scalar dead-zone plus uniform threshold quantizer is used in the case of mean squared quantization error optimization as proposed in [4]. Figure 4 visualizes the parameters α, β, and Δ of the quantization and inverse quantization. For each quantization parameter QP, an individual quantizer is designed whereas the parameters are determined by measurements. Table 1 shows the determined parameters α, β, and Δ for commonly used QPs. In the case of subjectively weighted quantization error optimization, a non uniform quantizer is applied with representative levels r_i, $-r_i$ and decision thresholds in the middle of adjacent r_i, which are also shown in Table 1 . In this case, visual masking [5] that occurs at luminance edges is exploited. Consequently, large quantization errors are allowed at edges, small ones if the image signal is flat. In Figure 5, the measured mean absolute reconstruction error $\overline{|q_{FD}(|e|)|}$ of subjectively weighted quantization in the frequency domain is shown for each absolute value of the prediction error $|e|$. It also shows the absolute reconstruction error of subjectively weighted quantization in the spatial domain $|q_{SD}(|e|)|$. The representative levels r_i are adjusted such that the mean absolute reconstruction error is the same for quantization in the frequency and in the spatial domain with respect to the quantization intervals in the spatial domain.

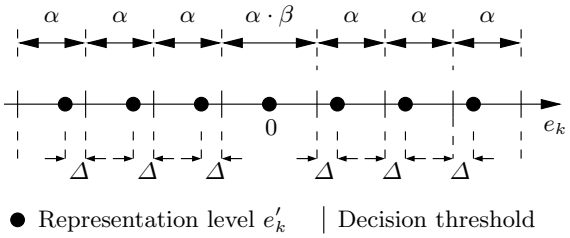

Fig. 4. Quantizer characteristic for the spatial domain in the case of mean squared error optimization.

3 Entropy coding

For entropy coding of the quantized samples in the spatial domain also CABAC is used as for the coding of the quantized coefficients in the frequency domain. However the scan is changed. Figure 6

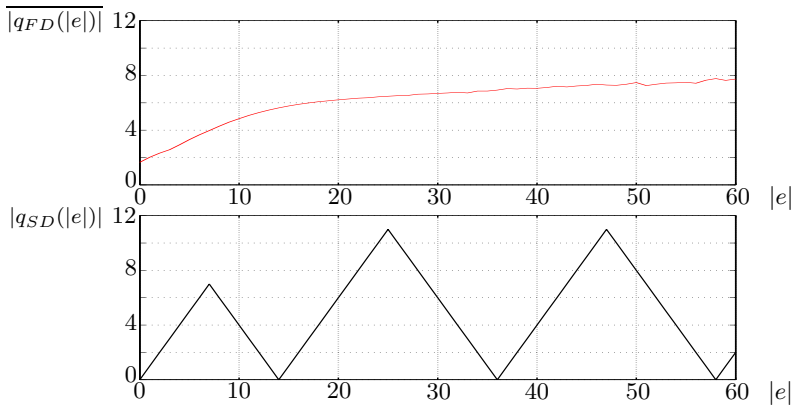

Fig. 5. Measured mean absolute reconstruction error $\overline{|q_{FD}(|e|)|}$ of subjectively weighted quantization in the frequency domain and absolute reconstruction error $|q_{SD}(|e|)|$ of subjectively weighted quantization in the spatial domain versus absolute value of the prediction error $|e|$. Quantization parameter $QP = 26$.

Table 1. Parameters α, β, and Δ of the quantizer in the case of mean squared quantization error optimization and representative levels r_i of the quantizer in the case of subjectively weighted quantization error optimization.

QP	α	β	Δ	r_1	r_2	r_3	r_4	r_5
23	9.6	1.6	2.7	0	11	28	46	66
26	14.8	1.4	4.8	0	14	36	58	110
29	22.2	1.4	6.9	0	20	54	92	148
32	30.2	1.4	9.3	0	28	76	130	220

shows that the conditional probability of a quantized predition error unequal zero increases with increasing gradients. Therefore, in order to arrange the samples in the 1D array with decreasing energy, the scan in the spatial domain is controlled by the magnitude of the gradient $G_{\hat{s}}$ of the prediction image $\hat{s}$ at the same spatial position as depicted in Figure 7. The samples to be coded are arranged in an order of decreasing gradients $G_{\hat{s}}$. The magnitude of the gradient at a spatial position x,y is calculated by

$$G_{\hat{s}} = \sqrt{g_x^2(x,y) + g_y^2(x,y)}. \tag{1}$$

g_x is the gradient in the horizontal and g_y the one in the vertical direction. They are calculated using the sobel operator:

$$\begin{aligned} g_x(x,y) = \tfrac{1}{8}[\, & 2\hat{s}(x+1,y) + \hat{s}(x+1,y-1) + \\ & \hat{s}(x+1,y+1) - 2\hat{s}(x-1,y) - \\ & \hat{s}(x-1,y-1) - \hat{s}(x-1,y+1)] \end{aligned}$$

$$\begin{aligned} g_y(x,y) = \tfrac{1}{8}[\, & 2\hat{s}(x,y+1) + \hat{s}(x+1,y+1) + \\ & \hat{s}(x-1,y+1) - 2\hat{s}(x,y-1) - \\ & \hat{s}(x+1,y-1) - \hat{s}(x-1,y-1)]. \end{aligned} \tag{2}$$

Separate probability models are used for the CABAC design in the frequency and in the spatial domain. The context modeling for the spatail domain is done in the same way as in the frequency domain.

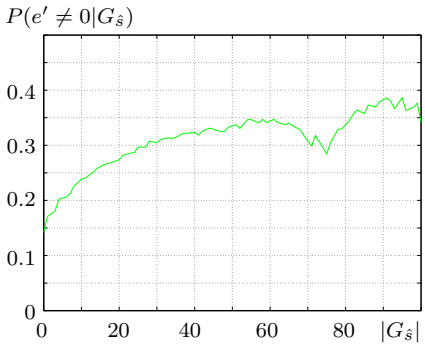

Fig. 6. Measured conditional probability of a quantized prediction error unequal zero given the magnitude of the gradient $G_{\hat{s}}$. Quantization parameter $QP = 26$.

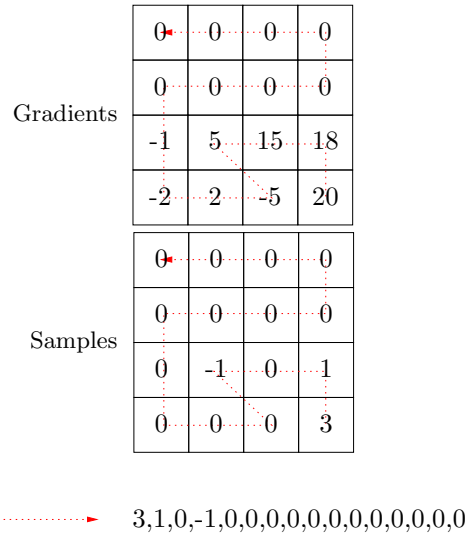

Fig. 7. Scan of quantized samples in the spatial domain according to the magnitude of the gradient in the prediction image at the same spatial position

4 Coding of the side information

For each individual macroblock, a flag indicates if all blocks of the current macroblock, which are of the size 4x4 or 8x8 picture elements, are coded in the frequency domain or if at least one block is coded in the spatial domain. This flag is coded conditioned by the flags of the already coded neighboring macroblocks to the top and to the left. If at least one block of the macroblock is coded in the spatial domain, it is indicated by a flag for each block of the macroblock if coding in the frequency or in the spatial domain is applied. This flag is coded conditioned by the flags of the already coded neighboring blocks to the top and to the left.

5 Coder control

For all luminance prediction error blocks of each macroblock either coding in the frequency or in the spatial domain is applied according to the lower rate-distortion costs $C = SSD + \lambda \cdot R$. SSD is the sum of squared differences between e and e', R the required rate including the side information and λ the Lagrange parameter as commonly used for the coder control of H.264/AVC [6]. Since the chrominance signal is not considered in this investigation yet, all chrominance blocks are unchanged coded in the frequency domain.

6 Experimental results

For experimental investigations, the current version jm10.1aif of the Key Technical Area Software [7] is merged with the current version jm10.2 of the reference software of H.264/AVC [8] and used as reference. In the following, it is referred to as KTA. The adaptive prediction error coding is integrated into KTA. High Profile with CABAC and 5 reference frames is applied. In the case of progressive video sequences, the frame structure I B RB B P ... with hierarchical B-pictures is used. RB indicates a B-picture which is used as a reference picture. Since KTA supports this hierarchical structure only for progressive sequences, the frame structure I B B P ... is used the case of interlaced video sequences. For interlaced video sequences, macroblock adaptive frame-field coding is applied. Rate-distortion optimization in the "High Complexity Mode" is turned on. Both kinds of quantization are testet. In Figures 8-15, operational rate-distortion curves are shown for the mean squared quantization error optimization for various test sequences with the spatial resolutions QCIF, CIF, SDTV, and HDTV. The operational rate distortion curves measured for the mean squared quantization error optimization and for the subjectively weighted quantization error optimization are approximately the same. For QCIF, CIF, and SDTV, the PSNR is increased by up to 0.4 dB by the adaptive coding compared to H264/AVC at the same bit rate. For HDTV, the PSNR is increased by up to 0.05 dB.
Furthermore, the adaptive prediction error coding is tested in combination with an increased displacement vector resolution of 1/8-pel [9], which improves the motion compensated prediction and reduces the correlation of the prediction error samples which is a benefit for the coding in the spatial domain. In combination with 1/8-pel displacement vector resolution, the PSNR is increased by up to 1 dB for QCIF, CIF, and SDTV and by up to 0.06 dB for HDTV.

7 Conclusion

In this paper, adaptive prediction error coding in the spatial and in the frequency domain is investigated in an H.264/AVC environment. For each block of the luminance signal, either frequency domain or spatail domain coding is applied. The Y-PSNR is increased by up to 0.4 dB for QCIF,

CIF and SDTV. In combination with 1/8-pel motion vector resolution, the Y-PSNR is increased by up to 1 dB for QCIF, CIF and SDTV. Since this technique appears very attractive for internet and mobile television it was recently presented at the ITU-T meeting [10].

References

1. ITU-T Recommendation H.264 and ISO/IEC 14496-10 (MPEG-4) AVC: Advanced Video Coding for Generic Audiovisual Services (2005)
2. J. Ostermann, J. Bormans, P. List, D. Marpe, M. Narroschke, F. Pereira, T. Stockhammer, and T. Wedi: Video coding with H.264/AVC: Tools, Performance, and Complexity. IEEE Circuits and Systems Magazine **4** (2004) 7–28
3. D. Marpe, H. Schwarz, and T. Wiegand: Context-Based Adaptive Binary Arithmetic Coding in the H.264/AVC Video Compression Standard. IEEE Transactions on Circuits and Systems for Video Technology **13** (2003) 620–636
4. G. J. Sullivan, S. Sun: On Dead-Zone Plus Uniform Threshold Scalar Quantization. In: Proc. of Visual Communications and Image Processing (VCIP), Beijing (2005)
5. A. Fiorentini, M. T. Zoil: Detection of a Target Superimposed to a Step Pattern of Illumination. II. Effects of a Just-Perceptible Illumination Step. Atti. Fond. G. Ronchi **22** (1967) 577–586
6. K.-P. Lim, G. Sullivan, T. Wiegand: Text Description of Joint Model Reference Encoding Methods and Decoding Concealment Methods. Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, doc. JVT-O079, Busan, Korea (2005)
7. Key Technical Area Software of the ITU-T, Version jm10.1aif, SG16, Q6: downloadable at http://iphome.hhi.de/suehring/tml/download/KTA/ (2005)
8. Reference software of H.264/AVC, version jm10.2: downloadable at http://iphome.hhi.de/suehring/tml/download/ (2006)
9. T. Wedi, H.G. Musmann: Motion- and Aliasing-Compensated Prediction for Hybrid Video Coding. IEEE Transactions on Circuits and Systems for Video Technology **13** (2003) 577–586
10. M. Narroschke, H.G. Musmann: Adaptive prediction error coding in spatial and frequency domain for H.264/AVC. ITU-T Q.6/SG16, doc. VCEG-AB06, Bangkok, Thailand (2006)

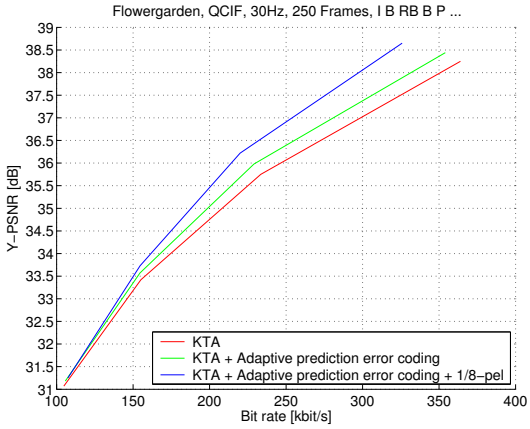

Fig. 8. Operational rate distortion curves for the test sequence *Flowergarden* in QCIF.

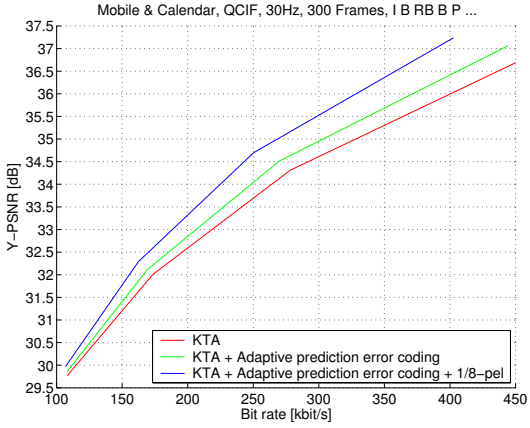

Fig. 9. Operational rate distortion curves for the test sequence *Mobile & Calendar* in QCIF.

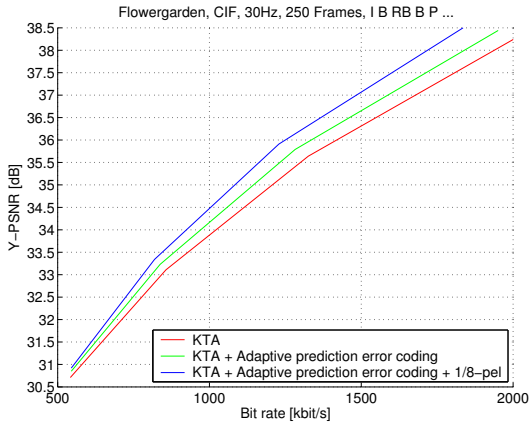

Fig. 10. Operational rate distortion curves for the test sequence *Flowergarden* in CIF.

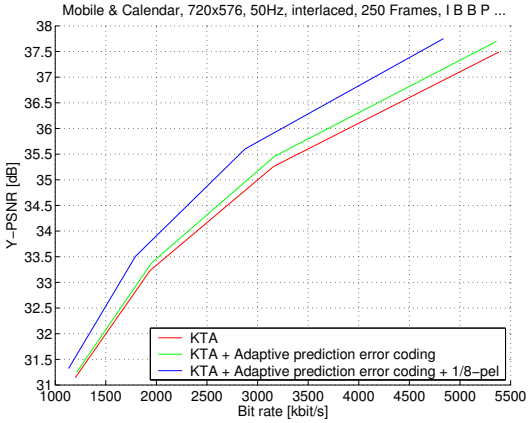

Fig. 13. Operational rate distortion curves for the test sequence *Mobile & Calendar* in SDTV.

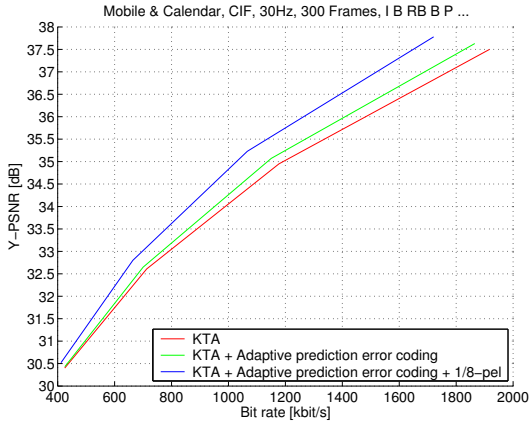

Fig. 11. Operational rate distortion curves for the test sequence *Mobile & Calendar* in CIF.

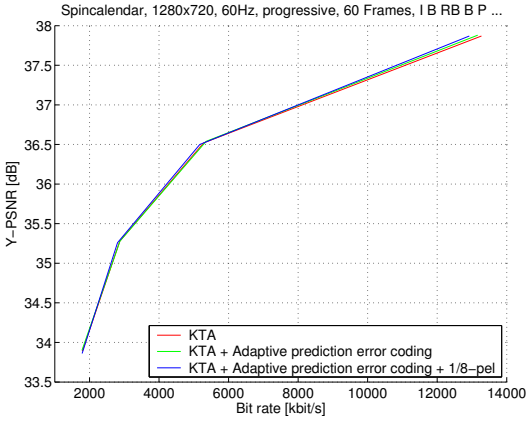

Fig. 14. Operational rate distortion curves for the test sequence *Spincalendar* in HDTV.

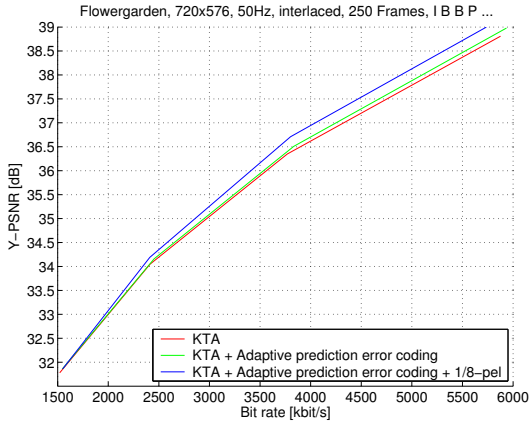

Fig. 12. Operational rate distortion curves for the test sequence *Flowergarden* in SDTV.

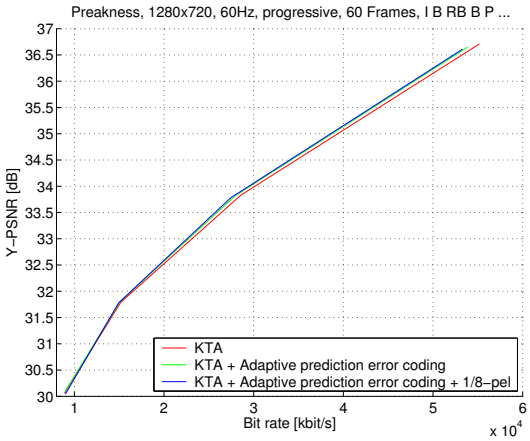

Fig. 15. Operational rate distortion curves for the test sequence *Preakness* in HDTV.