A Fast Intra 4x4 Mode Decision Algorithm for H.264/AVC Down Rate Transcoding

Zhihang Wang*[a], Luhong Liang[a], Shengfu Dong[b], Wen Gao[b], Debin Zhao[c], Qingming Huang[d]

[a]Key Laboratory of Intelligent Information Processing, Institute of Computing Technology,Chinese Academy of Sciences, Beijing, 100190, China
[b]Institute of Digital Media, Peking University, Beijing 1000871, China
[c]Department of Computer Science, Harbin Institute of Technology, Harbin, 150001, China
[d]Graduate University of Chinese Academy of Sciences, Beijing, 100039, China

ABSTRACT

H.264/AVC adopts 9 intra 4x4 prediction modes to improve the intra frame compression efficiency. This makes the intra frame coding and transcoding very complex. In this paper, a fast intra 4x4 mode decision algorithm is proposed to reduce the intra transcoding complexity. Firstly, the surrounding modes in the high bitrate and the modes of the neighboring blocks in the current bitrate are used to form the candidate modes. Secondly, the variance and mean of the intra prediction residual are used to further reduce the size of the candidate mode set. Thirdly, the intra mode context model is built from the high bitrate modes and updated by the low bitrate modes to order the candidate modes. An early skip scheme is used to terminate the mode decision process. Experimental results show that about 50% intra 4x4 modes can be saved with a penalty on coding efficiency less than 0.1dB. The transcoding performance can even gain up to 0.6 dB on some sequences.

Keywords: transcoding, fast intra mode decision , down rate transcoding

1. INTRODUCTION

H.264/AVC[1] is the latest international video coding standard which improves the compression efficiency significantly. It adopts intra prediction technology to improve the compression efficiency of I frames at the expense of complexity.[1] There are 9 intra 4x4 prediction modes and 4 intra 16x16 intra prediction modes. In the FRExt[2] profile of H.264/AVC, another 9 intra 8x8 intra prediction modes are enabled. Thus the complexity of I frame is very high. Numerous research in recent years[3] have been done to reduce the complexity of the intra mode decision process. On the other hand, because of the different bandwidth of the internet, transcoding the bitstream from high bitrate to low bitrate in real time is important to transport the video on the internet. Low complexity transcoding algorithm is needed to meet the application requirement.

Basically, there are two methods to reduce the intra transcoding complexity. One is using the fast intra mode decision algorithms. For example, the edge information can be used to speed up the intra mode decision process.[4] Kim *et al.*[3] proposed to use the SAD and SATD to speed up the intra mode decision process. But it will cause more than 0.1 dB loss in the rate distortion performance. The intra mode context information is also used[5,6] to reduce the complexity. Sim's algorithm[5] cause up to 0.4 dB loss in the rate distortion performance. Kai's *et al.* method provides very good performance.[6] But the adjustment method of the γ value which influences the performance greatly is still not known. Although these fast intra mode decision algorithms can be used in the transcoder to speed up the intra frame transcoding, they mainly only focus on the encoding process. In the transcoder, the mode information in the high bitrate bitstream can be used to further improve the intra mode decision performance. But so far their is little research on this topic. This is the main focus of this paper.

In this paper, a novel *practical* fast intra 4x4 mode decision algorithm is proposed. Firstly, the surrounding modes of current block in the high bitrate bitstream and the modes of neighboring blocks in the current bitrate bitstream are used to form the candidate mode set so that the candidate mode search space size is reduced. Secondly the variance and mean of the intra prediction residual are used to further filter out some candidate

*zhihangwang@jdl.ac.cn; www.jdl.ac.cn

Visual Communications and Image Processing 2010, edited by Pascal Frossard, Houqiang Li, Feng Wu, Bernd Girod, Shipeng Li, Guo Wei, Proc. of SPIE Vol. 7744, 77442H · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.863464

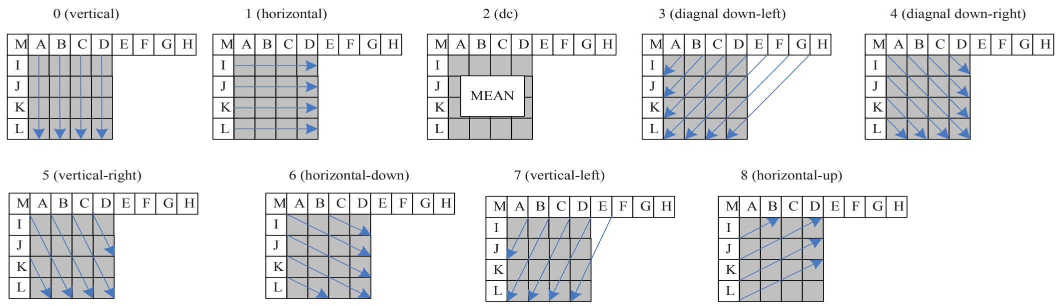

Figure 1. the H.264/AVC intra prediction mode

modes. Intuitively, if the variance and mean of the residual are small enough, the corresponding mode can be selected as the best mode. Finally, the neighboring mode context model is built using the mode information in the high bitrate stream and updated using the low bitrate stream to track the mode changes. These context models can order the modes in the candidate mode set according to their probabilities. A threshold is used to early terminate the mode decision process.

The rest of the paper is organized as follows. Section 2 reviews the intra 4x4 prediction technology in H.264/AVC. In Section 3, the fast intra 4x4 mode decision algorithm is proposed. Experimental results are presented in Section 4. Section 5 concludes this paper.

2. INTRA 4X4 MODE DECISION IN H.264/AVC

There are total 9 intra 4x4 prediction modes in H.264/AVC as shown in Fig. 1. Except for the DC mode, the other 8 modes have different prediction directions. If the mode is available, the RDCost (rate distortion cost) of this mode is calculated as

$$J = D + \lambda \times R \tag{1}$$

Here D represents the distortion caused by the quantization process and R is the bits used to encode this mode. Assuming that every intra 4x4 block is independent,[8] after calculating the RDcost of every mode, the mode corresponding to the *minimal* J is selected as the best intra prediction mode. There are two problems in this method. One is complexity. Obtaining D and R demands huge computations. Almost every intra 4x4 block will be encoded 9 times. The intra mode decision process is very complex. The other one is that the final mode is not the *really* optimal mode because of the assumption. Thus it is possible to find more suitable modes to improve the compression performance.

3. PROPOSED ALGORITHM

The proposed algorithm is designed according to the different characters of the intra prediction residual and the mode correlation between the high bitrate bitstream and the low bitrate bitstream. It mainly consists of three parts as shown in Fig. 2. Firstly, the neighboring modes in the high rate and current rate are used to form the candidate mode set. Secondly, we use the variance and the mean of the intra prediction residual to speed up some intra 4x4 blocks. For the remaining 4x4 blocks, the context mode model built from the high bitrate is used to order the candidate modes. A threshold is calculated according to QP to early terminate the intra mode decision process.

3.1 Surrounding Modes Based Candidate Mode Decision

Generally speaking, the intra modes between the high bitrate bitstream and the low bitrate bitstream have high correlation. Most of the modes are the same between different bitrate bitstreams as shown in Table 1. But it is obvious that the mode will change. This is because the rate distortion characters will change even for the same residual. The residual of every mode is also changed. It is very difficult to track these changes analytically. Here we use the heuristic method to track the mode change. We suppose that the intra 4x4 modes has high correlation

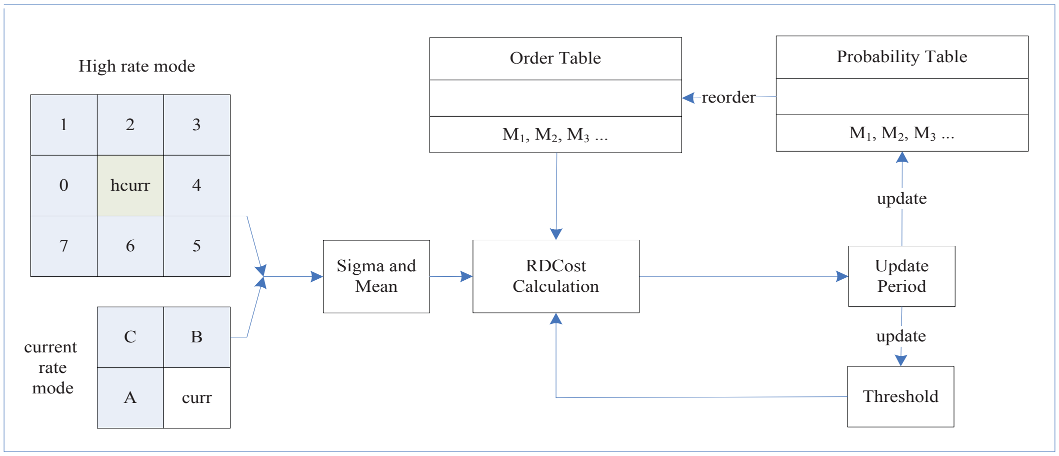

Figure 2. the context adaptive fast intra mode decision algorithm

with the edge information in the image. Because there usually exists strong correlation in the small area of the image, the mode of the current block and the mode of the surrounding blocks in the high bitrate are used as the candidate modes. These modes consist of the first part of the modes in the candidate mode set. For example, the modes of block *hcurr* and block 0, 1, 2, 3, 4, 5, 6, 7 as shown in Fig. 2 are used as the first part modes of the candidate mode set. According to the intra prediction mode direction characters, a neighbor mode table is defined as shown in Table 2. This table is defined according to Fig. 3. These neighbor modes are also used as the candidate modes. They consist of the second part of the modes in the candidate mode set. For example, if the mode of block 0 is the vertical mode (0), intra mode 5 and intra mode 7 are also used as the candidate modes. In order to track the mode change, the third part of the candidate mode set contains the modes of block A, B, C as shown in Fig. 2. If the mode of block *hcurr* is different from all its surrounding modes, all the 9 intra modes will be used as the candidate modes. The reason is that in this case it denotes there is little correlation between current mode and its surrounding modes. The DC mode is always enabled because it contains no edge information.

3.2 Variance and Mean Based Candidate Mode Decision

After the candidate mode set is determined using the method in 3.1, we use the variance and mean of the intra 4x4 mode prediction residual to further filter out some intra 4x4 modes because the variance of a signal plays an important role in the rate distortion performance according to the rate distortion theory. If x is a memoryless Gaussian source with variance σ_x^2, its rate-distortion function is[9]

$$R(D) = \begin{cases} \frac{1}{2}\log_2(\sigma_x^2/D), & \text{if } D \le \sigma_x^2 \\ 0, & \text{if } D > \sigma_x^2 \end{cases} \tag{2}$$

If y is also a memoryless Gaussian source, then we can get that if $\sigma_y > \sigma_x$ and $D_y = D_x$, then $R_y > R_x$. This motivates us to use σ^2 to decide the best mode. In practical video coding, the mean of the residual is the DC coefficient and should be encoded. It will also influence the residual rate. This parameter is also used to decide the final mode. After many experiments, we use T_σ and T_{mean} as the threshold. They are calculated as

$$T_\sigma = \frac{\sum_{i=1}^{n} \sigma_i^2}{3n} \tag{3}$$

$$T_{mean} = \frac{\sum_{i=1}^{n} |mean_i|}{2n} \tag{4}$$

Table 1. Intra Mode 0 Change Trend of QCIF Video Sequences from QP 16 to QP 24

sequence	0→0	0→1	0→2	0→3	0→4	0→5	0→6	0→7	0→8
akiyo	79.56%	3.87%	5.52%	2.21%	0.74%	4.05%	0.55%	2.76%	0.74%
coastguard	58.46%	15.38%	12.31%	4.62%	0.00%	3.08%	3.08%	1.54%	6.15%
container	67.03%	13.74%	3.85%	3.30%	2.20%	1.65%	2.20%	1.10%	4.95%
flower	87.02%	3.19%	1.91%	1.49%	0.85%	2.34%	0.64%	1.49%	1.06%
foreman	57.14%	8.67%	12.24%	2.55%	2.55%	5.61%	2.04%	7.14%	2.04%
mobile	55.83%	3.75%	10.00%	3.75%	6.67%	6.25%	4.17%	6.25%	3.33%
news	74.48%	11.30%	2.93%	1.46%	1.67%	2.09%	0.63%	3.77%	1.67%
stefan	52.69%	18.28%	10.75%	3.76%	2.15%	3.23%	4.30%	2.15%	2.69%
average	60.04%	10.50%	8.98%	2.88%	3.26%	4.30%	2.79%	4.83%	2.43%

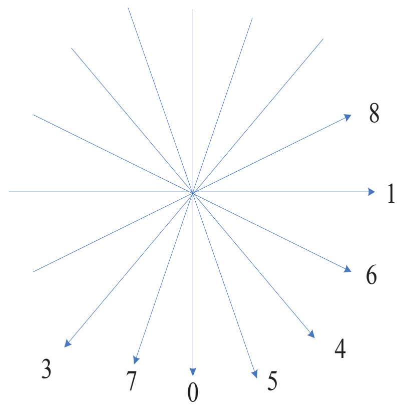

Figure 3. The intra mode directions

Here n is the total available intra 4x4 mode number. If $\sigma_i^2 < T_\sigma$ and $mean_i < T_{mean}$, mode i will be selected as the candidate mode.

3.3 Context Model Based Mode Ordering

Now the candidate mode set has been determined. We try to order the modes in the candidate mode set so that the most probably mode will be checked first.

The neighboring modes posse high correlation.[6] This relation can be modeled as a 3-order Markov random field.

$$P(M_c|(M_A, M_B, M_C)) \equiv P(Mode_{curr} = M_n|Mode_A = M_A, Mode_B = M_B, Mode_C = M_C) \quad (5)$$

where $Mode_{curr}$, $Mode_A$, $Mode_B$ and $Mode_C$ are random variables that represent the best mode of current block and its neighboring blocks denoted in Fig. 2. A probability table as show in Fig. 2 is used to store this information. All the items in the probability table are initialized to be 1. This means that the probability of every mode for current block is the same. During the transcoding process, the probability table can be built online.

The probability table is built from the mode information in the high bitrate bitstream. Every occurrence of the four tuple (M_c, M_A, M_B, M_C) in the high bitrate bitstream will cause the corresponding item in the probability table increase 1. In order to track the mode change trend, the probability table is updated after

Table 2. Neighbor Mode Table

mode	0	1	2	3	4	5	6	7	8
neighbor mode	5,7	6,8	0,1	0,7	5,6	0,4	1,4	0,3	1,2

encoding K 4x4 blocks. All the intra 4x4 modes will be checked every K 4x4 blocks. In our experiments, K is set to be 50. After encoding every K 4x4 blocks, if the best mode is the same as the corresponding element in the probability table, it means that the probability table can predict this mode correctly. Then this element in the probability table is increased by 1. The modes are reordered according to the corresponding item in the probability table. Because of the candidate mode set determined before, the γ value[6] is not needed. This makes the algorithm practical.

After ordering the modes, a threshold is used to early terminate the intra mode decision process. The threshold should be selected carefully to ensure little loss in the rate distortion performance. Now the initial value of the threshold is calculated as[6]

$$T_4 = 2^{0.33 \times QP - 1.265} \tag{6}$$

The threshold is updated with the mode probability table. After encoding every K 4x4 blocks, all the modes are checked. If the minimal RDCost is greater than T_4, T_4 is updated as

$$T_4 = T_4 + 2 \times 0.33 \times 2^{0.33 \times QP - 1.265} \tag{7}$$

If the minimal RDCost is less than T_4, T_4 is updated as

$$T_4 = T_4 \times 0.4 \tag{8}$$

4. EXPERIMENTAL RESULTS

The proposed algorithm is implemented on JM 15.1. Only intra 4x4 mode prediction is enabled to test its performance. The original intra mode decision method is treated as the anchor. We encode 150 frames of every sequence. All the frames are encoded as I frames. The frames are first encoded with QP set to 16. Then the generated bitstream is decoded and re-encoded with QP equal to 28, 32, 36 and 40. The saved mode percentage ($\Delta Save$) and the change of the mode decision time ($\Delta Time$) are used to show the speed of this algorithm. $\Delta PSNR$[10] is also used to show the rate distortion performance of this algorithm.

Table 3 and table 4 show the performance of this algorithm. The rate distortion curve of some QCIF and CIF video sequences are shown in Fig. 4 and Fig. 5. It can be seen that this algorithm is very robust. About half of the intra 4x4 mode can be saved with little loss in the rate distortion performance. On some sequences such as bridge and football, the performance is even better than the original JM encoder because the original mode decision method is not the *really* optimal method. The conditions that this algorithm will be better than the original method still needs to be investigated. Nevertheless, the performance loss of this fast algorithm is very little.

5. CONCLUSION

In this paper, a fast intra mode decision algorithm for the down rate transcoding is proposed. The disadvantages of the fast intra mode decision algorithms are avoided in this algorithm. Because the γ value is not needed, this make the algorithm practical. Extensive experimental results show that this algorithm can save about half of the computations in the intra mode decision process with negligible loss in the rate distortion performance.

ACKNOWLEDGMENTS

This work is supported in part by National Science Foundation of China (60803068), National Science Foundation of China (60832004), and National Basic Research Program of China (973 Program, 2009CB320904).

Table 3. QCIF Video Sequence Test Results

QP / sequence	$\Delta Save$				$\Delta Time(\%)$				$\Delta PSNR(dB)$
	28	32	36	40	28	32	36	40	
foreman	54.19%	53.67%	52.99%	51.93%	56%	55%	53%	52%	-0.059
akiyo	57.65%	55.62%	55.06%	53.14%	57%	54%	54%	52%	-0.047
coastguard	45.05%	44.81%	44.70%	44.61%	45%	44%	44%	44%	0.034
container	46.36%	46.84%	47.47%	47.22%	46%	46%	46%	47%	-0.005
hall	54.92%	53.32%	52.49%	51.63%	55%	53%	51%	51%	-0.078
silent	44.93%	44.44%	43.89%	43.10%	44%	43%	43%	42%	-0.060
soccer	45.94%	45.63%	45.85%	45.42%	46%	46%	45%	45%	0.011
stefan	44.82%	43.94%	42.94%	41.68%	39%	38%	38%	38%	-0.029
flower	43.65%	42.82%	42.07%	41.03%	36%	36%	36%	36%	-0.053
football	43.42%	43.43%	43.65%	43.60%	43%	42%	42%	42%	-0.008
news	54.60%	53.47%	53.59%	51.93%	53%	52%	52%	50%	-0.038
bus	42.22%	42.13%	41.77%	41.52%	41%	41%	40%	41%	-0.058
bridge-close	42.08%	41.93%	41.78%	42.41%	40%	39%	40%	41%	-0.015
bridge-far	42.20%	42.86%	43.46%	44.74%	42%	42%	42%	44%	0.623
average	47.29%	46.48%	46.55%	46.00%	46%	45%	45%	44.6%	0.015

Table 4. CIF Video Sequence Test Results

QP / sequence	$\Delta Save$				$\Delta Time(\%)$				$\Delta PSNR(dB)$
	28	32	36	40	28	32	36	40	
foreman	50.95%	50.67%	50.22%	49.61%	51%	51%	50%	49%	-0.064
akiyo	54.75%	54.17%	52.82%	51.40%	54%	53%	52%	50%	0.015
hall	50.48%	50.31%	49.98%	49.49%	52%	51%	50%	49%	-0.068
soccer	46.44%	45.70%	45.61%	45.50%	44%	44%	44%	44%	0.028
stefan	47.27%	46.41%	45.44%	44.21%	43%	42%	41%	40%	-0.010
flower	44.80%	43.74%	43.19%	42.45%	36%	35%	36%	36%	-0.068
football	47.45%	46.76%	46.31%	46.30%	45%	44%	45%	45%	0.088
news	53.53%	52.78%	51.14%	48.89%	53%	52%	50%	49%	0.026
bus	44.27%	43.69%	43.27%	42.78%	43%	42%	42%	42%	-0.055
mobile	45.54%	45.16%	44.70%	43.72%	41%	41%	40%	40%	-0.076
paris	54.07%	53.17%	52.57%	51.69%	53%	51%	50%	50%	-0.026
tempete	45.26%	44.49%	43.71%	42.73%	41%	39%	39%	39%	-0.061
bridge-close	41.93%	42.33%	42.55%	43.03%	39%	39%	39%	41%	-0.003
bridge-far	35.95%	36.91%	38.22%	39.44%	35%	34%	36%	38%	0.222
average	46.99%	46.61%	46.21%	45.73%	45%	44%	44%	43.7%	-0.007

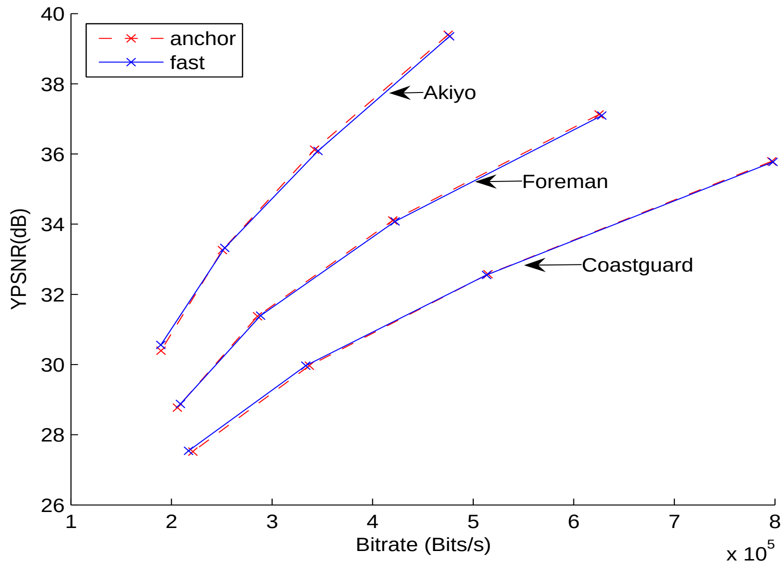

Figure 4. The R-D Curve of QCIF video sequence

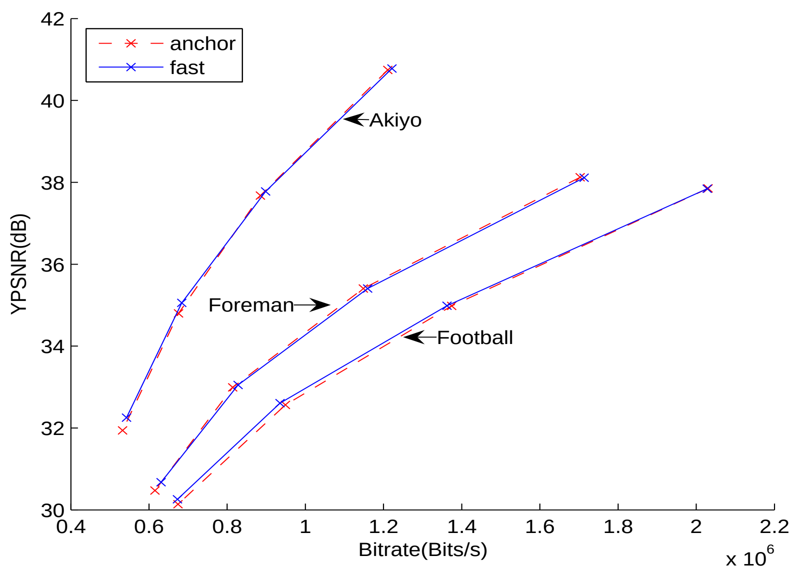

Figure 5. The R-D Curve of CIF video sequence

REFERENCES

[1] Wiegand, T., Sullivan, G. J., Bjontegaard, G., and Luthra, A., "Overview of the H.264/AVC video coding standard," *IEEE Transactions on circuits and systems for video technology* **13**, 560–576 (July 2003).

[2] Marpe, D., Wiegand, T., and Gordon, S., "H.264/MPEG4-AVC fidelity range extensions: tools, profiles, performance, and application areas," in [*Proc. IEEE International Conference on Image Processing ICIP 2005*], **1**, I–593–6 (2005).

[3] Kim, C., Shih, H., and Kuo, C., "Fast H. 264 intra-prediction mode selection using joint spatial and

transform domain features," *Journal of Visual Communication and Image Representation* **17**(2), 291–310 (2006).

[4] Pan, F., Lin, X., Rahardja, S., Lim, K., Li, Z., Wu, D., and Wu, S., "Fast mode decision algorithm for intra prediction in H.264/AVC video coding," *IEEE Transactions on Circuits and Systems for Video Technology* **15**(7), 813 (2005).

[5] Sim, D. and Kim, Y., "Context-adaptive mode selection for intra-block coding in H. 264/MPEG-4 Part 10," *Real-Time Imaging* **11**(1), 1–6 (2005).

[6] Zhang, K., Wang, Q., Huang, Q., Zhao, D., and Gao, W., "A Context-based Adaptive Fast Intra_4x4 Prediction Mode Decision Algorithm for H.264/AVC Video Coding," *PCS, Nov* (2007).

[7] Zhang, K., Ji, X., Huang, Q., Zhao, D., and Gao, W., "An efficient coding method for intra prediction mode information," in [*Circuits and Systems, 2009. ISCAS 2009. IEEE International Symposium on*], 2814–2817 (May 2009).

[8] Wiegand, T., Lightstone, M., Campbell, T. G., and Mitra, S. K., "Efficient mode selection for block-based motion compensated video coding," in [*ICIP*], 2559–2562 (1995).

[9] Berger, T., [*Rate distortion theory: A mathematical basis for data compression*], Prentice-Hall, Englewood Cliffs, NJ (1971).

[10] Bjontegaard, G., "Calculation of average PSNR differences between RD-curves," *Doc. VCEG-33M* (2001).