VLC Table Prediction Algorithm for CAVLC in H.264 Using the Characteristics of Mode Information

Jin Heo and Yo-Sung Ho

Gwangju Institute of Science and Technology (GIST)
261 Cheomdan-gwagiro, Buk-gu, Gwangju, 500-712, Korea
{jinheo,hoyo}@gist.ac.kr

Abstract. The most recent H.264 video coding standard adopted context-based adaptive variable length coding (CAVLC) as the entropy coding tool. By combining adaptive variable length coding (VLC) with context modeling, we can achieve a better coding performance. However, CAVLC in H.264 has a problem that correctness of VLC table prediction is low. In this paper, we propose a new VLC table prediction algorithm using the correlation of coding mode between the current and neighboring blocks and the statistics of mode distribution in both intra and inter frames. Moreover, we can further increase correctness of VLC table prediction considering the structural characteristics of the mode information in inter frame. Experimental results show that the proposed algorithm increases correctness of VLC table prediction by 10.07% and reduces the bit rate by 1.21% on average.

Keywords: H.264, CAVLC, VLC table, mode information.

1 Introduction

The latest international video coding standard, H.264, was developed by the Joint Video Team (JVT) from the ITU-T Video Coding Experts Group and the ISO/IEC Moving Picture Experts Group [1]. For higher compression efficiency, H.264 has adopted several powerful coding techniques, such as variable block-size macroblock modes, multiple reference frames, integer discrete cosine transform (DCT), and efficient entropy coding techniques [2].

In order to further increase compression efficiency, H.264 adopted context-based adaptive variable length coding (CAVLC) as an entropy coding technique, where quantized transform coefficients are coded using variable length coding (VLC) tables that are switched according to the values of previous syntax elements [3] [4]. Therefore, coding efficiency of CAVLC depends on how to estimate an appropriate VLC table accurately. However, since correctness of VLC table prediction is low in the current CAVLC, we need to enhance the accuracy of the VLC table in encoding quantized transform coefficients.

In this paper, we propose a new VLC table prediction algorithm for CAVLC in H.264. Considering the relation between the VLC table and the coding mode, we adopt two features, the correlation of coding mode between the current and neighboring blocks and the statistics of mode distribution to select a proper VLC

Y.-M.R. Huang et al. (Eds.): PCM 2008, LNCS 5353, pp. 119–128, 2008.
© Springer-Verlag Berlin Heidelberg 2008

table. Moreover, by using the mode information such as the mode structure, we can further increase correctness of VLC table prediction to encode the current 4×4 block.

This paper is organized as follows. After we introduce an overview of CAVLC framework including five coding steps and explain how to determine a VLC table for the current 4×4 block in Section 2. We propose a new VLC table prediction algorithm for CAVLC in Section 3. Experimental results are presented in Section 4, and we draw conclusion in Section 5.

2 Overview of CAVLC in H.264

In this section, we briefly describe a CAVLC scheme which is the entropy coding algorithm in H.264 baseline profile used to encode residual data, zig-zag ordered 4×4 blocks of quantized transform coefficients and explain a VLC table prediction algorithm [5]. First, we introduce five major steps of CAVLC.

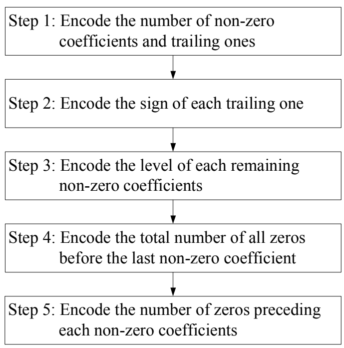


Fig. 1. Flowchart of CAVLC

In Step 1, we encode both the total number of non-zero coefficients and the number of trailing ones as a combination, using a selected VLC table out of the four tables based on the number of non-zero coefficients in the neighboring blocks.

In Step 2, since trailing ones are all equal to ±1, they only need the sign specification. Thus, the sign is encoded with a single bit ('+' = 0, '–' = 1) for each trailing one.

In Step 3, we encode the level (sign and magnitude) of each remaining non-zero coefficient in the current 4×4 block in the reverse order. The code for each level is made up of a prefix and a suffix.

In Step 4, we encode the sum of all zeros preceding the highest frequency non-zero coefficient of each 4×4 block.

In Step 5, the number of zeros preceding each non-zero coefficient is encoded in the reverse order.

At the first step, there are four VLC tables (three variable length code tables and one fixed length code table) used to encode both the total number of non-zero coefficients and the number of trailing ones in each 4×4 block. VLC table selection depends on the number of non-zero coefficients in the current 4×4 block which is predicted using the number of non-zero coefficients in the upper and left blocks. Figure 2 shows the upper and left blocks of the current block. The size of each block is 4×4.

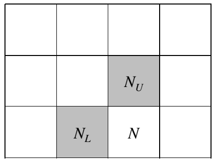


Fig. 2. Upper and Left Blocks of the Current 4×4 Block

When both upper and left blocks are available, the number of predicted non-zero coefficients in the current 4×4 block is calculated by

$$N=round(N_U+N_L)/2\cdot \tag{1}$$

Where N represents the number of predicted non-zero coefficients in the current 4×4 block. N_U and N_L are the number of non-zero coefficient in the upper and left blocks, respectively. If only the upper block is available, $N=N_U$. If only the left block is available, $N=N_L$. If neither is available, N is set to zero. Using the parameter N, we choose an appropriate VLC table from Table 1.

Table 1. Choice of VLC Table

N	VLC Table
0, 1	*Num-VLC0*
2, 3	*Num-VLC1*
4, 5, 6, 7	*Num-VLC2*
8 or above	*Num-FLC*

Unlike VLC algorithms in previous video coding standards, CAVLC selects one VLC table from the four possible VLC tables adaptively according to the values of previous syntax elements. Since VLC tables are context dependent, coding efficiency of CAVLC is better than those of other schemes using a single VLC table in previous video coding standards.

However, H.264 has a drawback that correctness of VLC table prediction is low. As shown in Fig. 3, the correct prediction rate of VLC table for six test sequences

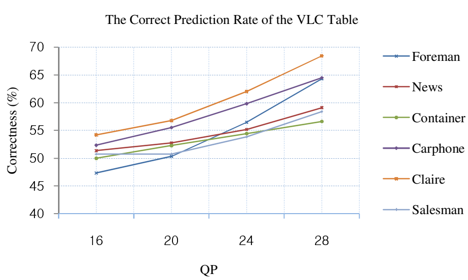


Fig. 3. Correct Prediction Rate of VLC Table in H.264 CAVLC

(Foreman, News, Container, Carphone, Claire, and Salesman) is about 55% on average. Consequently, the optimal VLC table is not used for encoding quantized transform coefficients and it reduces coding efficiency.

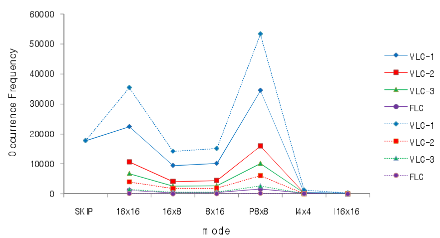
3 New VLC Table Prediction Algorithm

3.1 Observation of VLC Table According to the Mode

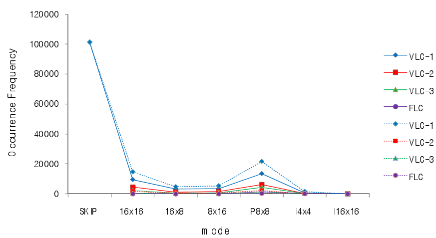
For a given macroblock, H.264 chooses the best coding mode from seven different potential prediction modes: *SKIP*, *16*×*16*, *16*×*8*, *8*×*16*, *P8*×*8*, *I4*×*4*, and *I16*×*16*. It also uses rate-distortion optimization (RDO) algorithm [6] to choose the best coding mode for one macroblock.

Figure 4 shows the occurrence frequencies of each VLC table for the seven modes. We perform experiments with 100 frames of Foreman and News sequences in the QCIF format (176×144). The coding structure is IPPP…P. Solid lines indicate the actual VLC tables and dotted lines represents the predicted VLC tables. From Fig. 4, we can observe that the occurrence frequencies of each VLC table are different according to the best coding mode. Moreover, there is a difference between the actual VLC table and the predicted VLC table at each mode.

When the current block mode is *SKIP*, no motion or residual information is encoded in the current 4×4 block. Since we do not need to encode the current block, the number of non-zero coefficients is set to zero. This is why the actual VLC table and the predicted VLC table are the same at *SKIP* in Fig. 4.



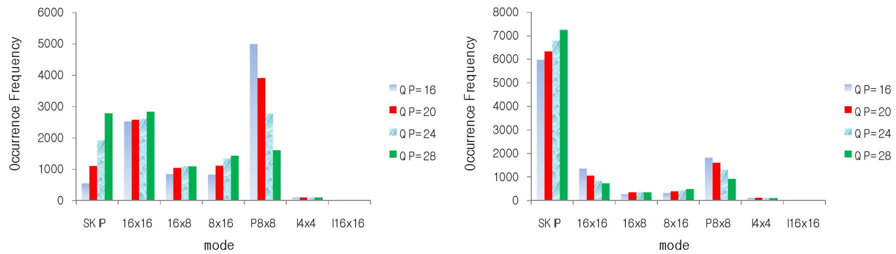
(a) Foreman



(b) News

Fig. 4. Comparison of Occurrence Frequency of each VLC Table at each Mode for QP=20

Figure 5 shows the distribution of seven coding modes according to the four quantization parameters (QP). We perform experiments with 100 frames. The coding structure is IPPP…P. In Foreman sequence, the most popular mode is *P8*×*8* in the low QP value. As the QP value is increased, the occurrence frequencies of both *SKIP* and *16*×*16* are increased. In News sequence, the most popular mode is *SKIP* in all QP values. It is found that, in general, except for *SKIP*, the occurrence frequencies of *16*×*16* and *P8*×*8* are higher than those of other modes. Therefore, if *16*×*16* or *P8*×*8* is occurred in the neighboring blocks, the VLC table of the current block is likely to be predicted to be one of the VLC tables of the neighboring blocks. Using the statistics of mode distribution, we can determine a proper VLC table.



(a) Foreman (b) News

Fig. 5. Distribution of Coding Mode according to QP

3.2 VLC Table Prediction Algorithm in Intra Frame

The available coding modes for a given macroblock in I-slice include $I4\times4$ and $I16\times16$. In I-slice coding, we develop some conditions for proper VLC table prediction. In Table 2, M_N, M_N_U, and M_N_L indicate the best coding mode of the current, upper, and left blocks. N indicates the number of non-zero coefficients in the current 4×4 block. $N_{I4\times4}$ and $N_{I16\times16}$ represent the number of non-zero coefficients in the 4×4 block with $I4\times4$ and $I16\times16$, respectively.

In the first condition, we predict N using Eq. (1). In the second condition, the occurrence frequency of $I4\times4$ is higher than that of $I16\times16$ as shown in Fig. 5. From this observation, if M_N is $I4\times4$, the probability that N is similar to $N_{I4\times4}$ is high. On the contrary, if M_N is $I16\times16$, the probability that the current 4×4 block is included in one macroblock with $I16\times16$ is high. This means that the characteristics of all 4×4 blocks in one macroblock with $I16\times16$ are similar. Therefore, in this case, the probability that N is similar to $N_{I16\times16}$ is high.

Table 2. VLC Table Prediction Condition for $M\_N = M\_N_U = M\_N_L$ and $M\_N = (M\_N_U$ or $M\_N_L)$ in Intra Frame

Condition	Current 4×4 Block Mode	N
$M\_N = M\_N_U = M\_N_L$	$I4\times4$	$round(N_U+N_L)/2$
	$I16\times16$	
$M\_N = (M\_N_U$ or $M\_N_L)$	$I4\times4$	$N_{I4\times4}$
	$I16\times16$	$N_{I16\times16}$

3.3 VLC Table Prediction Algorithm in Inter Frame

H.264 supports all seven modes for a given macroblock in P-slice. However, the occurrence frequencies of $I4\times4$ and $I16\times16$ in inter frame are relatively small against the occurrence frequencies of all other five modes as shown in Fig. 5. Therefore, we do not consider these two modes in inter frame.

Figure 6 shows two inter prediction modes, 16×8 and 8×16. Label A-P and a-p represent a 4x4 block within one macroblock in 16×8 and 8×16 modes, respectively. We define two same regions and one boundary region in 16×8 and 8×16. One same

A	B	C	D
E	F	G	H
I	J	K	L
M	N	O	P

(a) *16×8* Mode

a	b	c	d
e	f	g	h
i	j	k	l
m	n	o	p

(b) *8×16* Mode

Fig. 6. Two Prediction Modes

region is A, B, C, D, E, F, G, H and the other same region is I, J, K, L, M, N, O, P in *16×8* and one same region is a, b, e, f, i, j, m, n and the other same region is c, d, g, h, k, l, o, p in *8×16*. One boundary region is E, F, G, H, I, J, K, L in *16×8* and b, c, f, g, j, k, n, o in *8×16*.

In order to evaluate the influence of vertical and horizontal boundaries on the selection of a correct VLC table in *16×8* and *8×16*, we compare the occurrence probabilities of the same VLC tables in the same and the boundary regions. In order to calculate the occurrence probability of the same VLC table in *16×8*, we compare the VLC tables of (A, E), (B, F), (C, G), (D, H) and the VLC tables of (I, M), (J, N), (K, O), (L, P) in the same regions. We compare the VLC tables of (E, I), (F, J), (G, K), (H, L) in the boundary region. In *8×16*, we compare the VLC tables of (a, b), (e, f), (i, j), (m, n) and the VLC tables of (c, d), (g, h), (k, l), (o, p) in the same regions. We compare the VLC tables of (b, c), (f, g), (j, k), (n, o) in the boundary region.

In Table 3, we can observe that the probability that the same VLC table is selected in the same regions is higher than the probability that the same VLC table is selected in the boundary region in both *16×8* and *8×16*.

Table 3. Comparison of Probability of the Same VLC Table Selection in *16×8* and *8×16*

Sequence	QP	*16×8*		*8×16*	
		Same Region (%)	Boundary Region (%)	Same Region (%)	Boundary Region (%)
Foreman	20	41.10	36.83	47.36	44.10
	28	57.60	55.09	57.52	56.41
News	20	54.58	44.66	45.81	37.79
	28	47.96	45.30	54.30	53.05

Table 4 shows the VLC table prediction condition for the case of $M\_N = M\_N_U = M\_N_L$, $M\_N = M\_N_U$, and $M\_N = M\_N_L$. In Table 4, $N_{Same\ Region}$ represents the number of non-zero coefficients in the 4×4 block in the same region. In these conditions, we classify the current 4×4 block into two different types. One includes *16×8* or *8×16* and the other includes *16×16* or *P8×8*.

Table 4. VLC Table Prediction Condition for $M\_N = M\_N_U = M\_N_L$, $M\_N = M\_N_U$, and $M\_N = M\_N_L$ in Inter Frame

Condition	Current 4×4 Block Mode	Boundary Block	N
$M\_N = M\_N_U = M\_N_L$	16×8 or 8×16	Yes	$N_{Same\ Region}$
		No	$round(N_U+N_L)/2$
	16×16 or $P8\times8$	-	$round(N_U+N_L)/2$
$M\_N = M\_N_U$	16×8	Yes	*Refer to Table 5*
	16×8 or 8×16	No	N_U
	16×16 or $P8\times8$	-	N_U
$M\_N = M\_N_L$	8×16	Yes	*Refer to Table 5*
	16×8 or 8×16	No	N_L
	16×16 or $P8\times8$	-	N_L

In the first condition, if M_N is 16×8 or 8×16, we check whether the neighboring block is a boundary block or not. If either neighboring block is a boundary block, we can select $N_{Same\ Region}$ as N. Otherwise N is determined using Eq. (1). In the second condition, if M_N is 16×8, we check whether the upper block is a boundary block or not. If the upper block is a boundary block, three blocks, the current, upper and left have all different characteristics. Therefore, we determine N using Table 5. Otherwise N is N_U. Since a boundary is vertical in 8×16, we do not need to check whether the upper block is a boundary block or not in this condition. If M_N is 16×16 or $P8\times8$, we can directly select N_U as N. The third condition method is very similar to the second condition method. The difference between them is that the block that has the same mode with the current block is not the upper block but the left block.

In Table 5, $C_{16\times16}$ and $C_{P8\times8}$ indicate the cumulative occurrence frequencies of 16×16 and $P8\times8$, respectively. $N_{16\times16}$ and $N_{P8\times8}$ represent the numbers of non-zero coefficients in the 4×4 block with 16×16 and $P8\times8$, respectively. As I already mentioned before, if the current block mode is *SKIP*, there is not an encoded data in the current 4×4 block and the number of non-zero coefficients is set to zero. Therefore, in M_N M_N_U M_N_L condition, we do not need to consider the cumulative occurrence frequency of *SKIP*.

Table 5. VLC Table Prediction Condition for $M\_N \neq M\_N_U \neq M\_N_L$ in Inter Frame

Current and Neighboring Blocks Mode	Comparison of an Occurrence Frequency	N
SKIP, 16×16, $P8\times8$	$C_{16\times16} < C_{P8\times8}$	$N_{P8\times8}$
	$C_{16\times16} > C_{P8\times8}$	$N_{16\times16}$
SKIP, 16×16, (16×8 or 8×16)	-	$N_{16\times16}$
16×16, 16×8, 8×16	-	$N_{16\times16}$
16×16, (16×8 or 8×16), $P8\times8$	$C_{16\times16} > C_{P8\times8}$	$N_{16\times16}$
	$C_{16\times16} < C_{P8\times8}$	$N_{P8\times8}$

In the first condition, we compare $C_{16\times16}$ with $C_{P8\times8}$. If $C_{16\times16} < C_{P8\times8}$, we can directly select $N_{P8\times8}$ as N. Otherwise N is $N_{16\times16}$. In the second condition, since the characteristics of 16×8 and 8×16 are similar to each other as shown in Fig. 4 and Fig. 5, we can consider these two different modes as the same mode. We know that $C_{16\times16}$ is higher than $C_{16\times8}$ or $C_{8\times16}$ from Fig. 5. Therefore, we can directly select $N_{16\times16}$ as N. In the third and the last conditions, we know that $C_{16\times8}$ or $C_{8\times16}$ is lower than $C_{16\times16}$ or $C_{P8\times8}$. Therefore, we can directly determine N as $N_{16\times16}$ in the third condition. In the last condition, first we check $C_{16\times16}$ and $C_{P8\times8}$ and then determine N.

In this section, we propose a new VLC table prediction algorithm. The proposed algorithm depends on the correlation of coding mode, the statistics of mode distribution, and the structural characteristics of mode information. Figure 7 shows the flowchart of the proposed VLC table prediction algorithm.

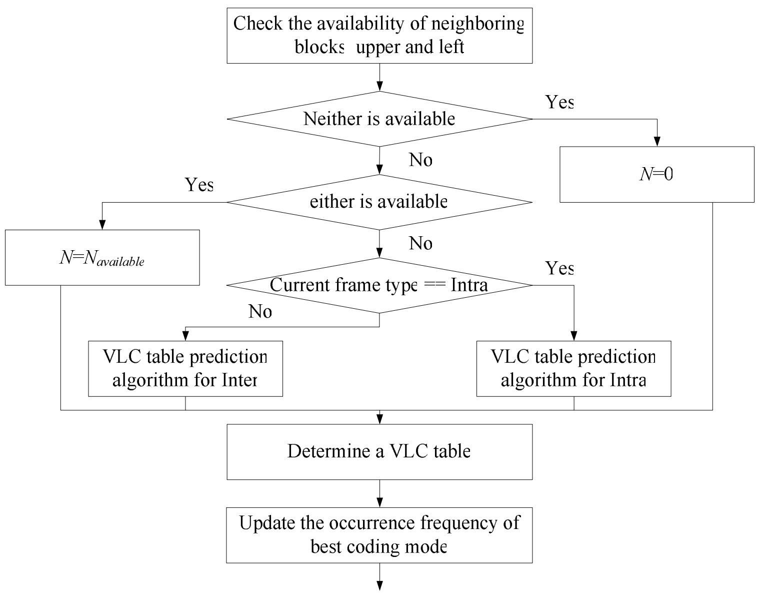


Fig. 7. Flowchart of the Proposed Algorithm

4 Experimental Results and Analysis

In order to evaluate the performance of the proposed algorithm, we encoded first 100 frames from six test video sequences in the QCIF format. JM 11.0 [7] was used to conduct experiments. We used the baseline profile. In motion estimation, one reference frame is enabled with the maximum search range ±16. The coding structure is IPPP…P. We tested for various QPs (16, 20, 24, and 28).

For the performance comparison between H.264 CAVLC and our proposed algorithm, we used delta VLC table prediction (Δ VLCTP) and bit saving (BS) measure as shown in Eq. (2) and Eq. (3).

$$\Delta\ \mathrm{VLCTP} = \mathrm{VLCTP}_{\mathrm{Proposed}} - \mathrm{VLCTP}_{\mathrm{H.264}}\ (\%). \tag{2}$$

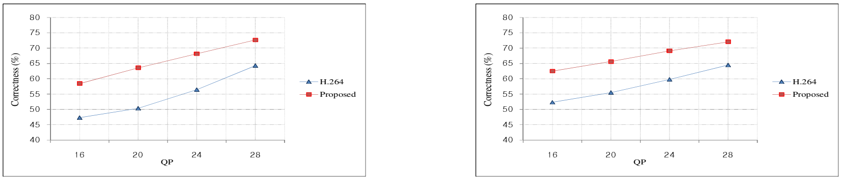
$$\mathrm{BS} = \frac{\mathrm{Bitrate}_{\mathrm{H.264}} - \mathrm{Bitrate}_{\mathrm{Proposed}}}{\mathrm{Bitrate}_{\mathrm{H.264}}} \times 100\ (\%). \tag{3}$$

Table 6 shows the performance of the proposed algorithm. The proposed algorithm achieved 7.81% ~ 13.36% correctness of VLC table prediction and 0.64% ~ 1.61% bit saving. From Table 6, we found that the proposed algorithm works more effectively on high +Δ VLCTP sequences, such as Foreman and Container. Table 6 also shows that correctness of VLC table prediction depends on QP.

Table 6. Comparison in the Performance Measures

Test Sequence	QP	H.264		Proposed		ΔVLCTP (%)	BS (%)
		Correctness of VLC Table (%)	Bit Rate (kbps)	Correctness of VLC Table (%)	Bit Rate (kbps)		
Foreman	16	47.32	698.42	58.45	688.38	+11.13	1.44
	20	50.32	402.89	63.57	396.40	+13.36	1.61
	24	56.47	233.38	68.15	230.35	+11.88	1.30
	28	64.31	136.95	72.67	135.80	+8.58	0.84
News	16	51.37	312.44	60.90	308.72	+9.53	1.19
	20	52.76	199.45	62.48	196.98	+9.72	1.24
	24	55.19	125.34	65.40	123.74	+10.21	1.28
	28	59.12	76.05	67.71	75.36	+8.59	0.91
Container	16	49.99	342.80	60.25	338.04	+10.60	1.39
	20	52.27	174.57	64.64	171.78	+12.58	1.59
	24	54.45	83.74	64.96	82.56	+10.88	1.41
	28	56.63	40.10	65.09	39.68	+8.76	1.06
Carphone	16	51.58	590.59	60.79	583.38	+9.21	1.22
	20	54.93	340.15	65.85	335.35	+10.92	1.41
	24	59.62	195.15	69.61	192.71	+9.99	1.25
	28	63.12	106.05	71.33	105.13	+8.21	0.91
Claire	16	52.34	182.41	62.50	180.24	+10.16	1.19
	20	55.51	102.60	65.60	101.33	+10.45	1.24
	24	59.81	58.35	69.11	57.89	+9.64	0.82
	28	64.48	32.21	72.08	32.00	+7.81	0.64
Salesman	16	49.41	287.02	58.44	283.35	+9.03	1.28
	20	52.79	164.31	63.15	162.09	+10.36	1.35
	24	57.83	96.99	68.49	95.65	+10.66	1.38
	28	63.51	56.88	72.85	56.41	+9.34	0.83

Figure 8 illustrates correctness curves for Foreman and Claire sequences. Correctness curves for Foreman and Claire sequences represent the best case and the worst case among experimental results, respectively. From Fig. 8, we can observe that correctness curves of the proposed algorithm are better than those of H.264 CAVLC. This achieves coding gain of CAVLC.



(a) Foreman (b) Claire

Fig. 8. Correctness Curves

5 Conclusions

In this paper, we proposed a new VLC table prediction algorithm for CAVLC in H.264. Considering the correlation of coding mode between the current and neighboring blocks and the statistics of mode distribution based on the relation between the VLC table and the coding mode, we developed conditions for proper VLC table prediction in both intra and inter frames. Moreover, we can further increase the correct VLC table prediction rate using the structural characteristics of the mode. Experimental results show that the proposed VLC table prediction algorithm increases correctness of VLC table prediction by 10.07% and reduces the bit rates by 1.21% on average, compared to CAVLC in H.264.

Acknowledgements

This work was supported in part by ITRC through RBRC at GIST (IITA-2008-C1090-0801-0017).

References

1. Joint Video Team of ITU-T and ISO/IEC JTC 1, Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification (ITU-T Rec. H.264 | ISO/IEC 14496-10 AVC), Doc. JVT-G050 (March 2003)
2. Wiegand, T., Sullivan, G.J., Bjontegaard, G., Luthra, A.: Overview of the H.264/AVC video coding standard. IEEE Transactions on Circuits and Systems for Video Technology 13(7), 560–576 (2003)
3. Richardson, I.E.G.: H.264 and MPEG-4 Video Compression_Video Coding for Next-generation Multimedia. Wiley, Chichester (2003)
4. Sullivan, G.J., Wiegand, T.: Video Compression-From Concept to the H.264/AVC Standard. Proceedings of the IEEE 93(1), 18–31 (2005)
5. Bjontegaard, G., Lillevold, K.: Context-adaptive VLC (CVLC) coding of coefficients. JVT Document JVT-C028 (May 2002)
6. Sullivan, G.J., Wiegand, T.: Rate-distortion optimization for video compression. Signal Processing Magazine 15, 74–90 (1998)
7. JVT Reference Software Version 11.0, `http://iphome.hhi.de//suehring/tml/download/jm_old/jm11.0.zip`