IMPROVED INTER MODE DECISION FOR H.264/AVC USING WEIGHTED PREDICTION

Amrita Ganguly and Anil Mahanta
Dept. of Electronics and Communication Engineering, Indian Institute of Technology Guwahati, Guwahati, India

Keywords: H.264/AVC, Inter mode decision, Edge histogram, Video coding.

Abstract: H.264/AVC video coding standard outperforms former standards in terms of coding efficiency but at the expense of higher computation complexity. Of all the encoding elements in H.264, inter prediction is computationally most intensive and thus adds to the computational burden for the encoder. In this paper, we propose a fast inter prediction algorithm for JVT video coding standard H.264/AVC. Prior to performing the motion estimation for inter prediction, characteristics like stationarity and homogeneity of each macroblock is determined. The macroblocks correlation with neighboring macroblocks in respect of predicted motion vectors and encoding modes are studied. Weights are assigned for these parameters and the final mode is selected based upon these weights. The average video encoding time reduction in the proposed method is 70% compared to the JVT benchmark JM12.4 while maintaining similar PSNR and bit rate. Experimental results for various test sequences at different resolutions are presented to show the effectiveness of the proposed method.

1 INTRODUCTION

The H.264/AVC is the latest video coding standard developed by Joint Video Group (JVT) of ITU-T Video Coding Experts Group (VCEG) and ISO/IEC MPEG Video Group (JVT.G050r1, 2003). It offers better compression efficiency and greater flexibility in compressing, transmitting and storing video. There are many advanced techniques that significantly improve the performance of the H.264/AVC video coding standard but at the cost of higher computational overhead. The efficient inter prediction using the variable block size motion estimation (ME) and compensation increase the coding complexity. H.264 permits the use of different block sizes, 16×16 pixels called a macroblock (MB) down to 4×4 pixels.

The H.264/AVC standard supports both intra and inter prediction processes. In the inter prediction process, there are seven block sizes P_{16x16}, P_{16x8}, P_{8x16}, P_{8x8}, P_{8x4}, P_{4x8} and P_{4x4} that are used by H.264 besides the SKIP and the INTRA modes (Richardson, 2003), (Wiegand et al., 2003). The block sizes are shown in Fig.1. For each MB, all the modes are tried and one which gives the least rate distortion (RD) cost is selected for encoding. The computation of the RD cost requires the availability of the reconstructed image and the actual bit count. This necessitates that the encoding and decoding processes are completed for every mode. Thus the computational requirements for the mode selection process is very high.

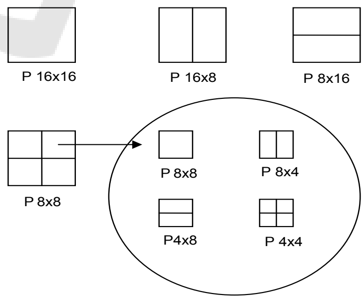


Figure 1: Inter prediction block sizes for a MB.

Several approaches have been proposed to reduce the complexity and time for the inter mode decision (Jing and Chau, 2004; Kim et al., 2006; Kim et al., 2004; Lee et al., 2006; Liu et al., 2009; Park and Capson, 2008; Shen et al., 2008; Wang et al., 2007; Wu et al., 2005; Zeng et al., 2009; Ganguly and Mahanta, 2010; Huang et al., 2010). In (Kim et al., 2006) adaptive MB selection is used for mode decision. In (Lee et al., 2006) the fast motion estimation is based on the successive elimination algorithm (SEA) using sum norms to find the best estimate of the motion vectors and to implement efficient calculations for variable blocks. In (Wu et al., 2005) edge histogram parameters have been used for fast inter mode decision.

Ganguly A. and Mahanta A..
IMPROVED INTER MODE DECISION FOR H.264/AVC USING WEIGHTED PREDICTION.
DOI: 10.5220/0003517300730078
In *Proceedings of the International Conference on Signal Processing and Multimedia Applications* (SIGMAP-2011), pages 73-78
ISBN: 978-989-8425-72-0

In this paper, the characteristics of a MB is first analyzed based on the stationarity, homogeneity, predicted motion vectors (MV) and the encoding modes of neighboring MBs. All these parameters are first determined for each MB. Weights are then assigned to these MBs depending upon the values of these four parameters. The decision on the final encoding mode for the MB is taken based upon the weights of all these parameters. The proposed method gives a simple model for arriving directly at the encoding mode without performing rigorous ME. Results show that the proposed method speeds up the encoding process by 70% when compared to the JM12.4 software with negligible loss in quality and coding efficiency.

The paper is organized as follows. The next section introduces the proposed fast mode decision algorithm. Section 3 illustrate the results. The conclusions are drawn in Section 4.

2 PROPOSED FAST DECISION ALGORITHM

In real video sequences the distribution of modes is not uniform among the MBs (Richardson, 2003). It depends upon the characteristics of the video sequences. Regions with homogenous motion, smooth motion of moving background and static background use larger block size motion compensation. For regions with high detail and complex motion, smaller block sizes to represent motion gives better coding efficiency. Changes between video frames may be caused by object motion, camera motion, uncovered regions and lighting variations. The neighboring video frames have large similarities between them. ME and compensation attempts to reduce the temporal redundancy by exploiting these similarities. Natural video sequences contain many regions with homogeneous motion that result in a large number of MBs being encoded with larger block sizes. The block size is also dependent on the QP. For large value of QP, more MBs tend to be encoded with larger block sizes. Since video sequences have different motion complexity, the sequences have been divided into three different classes: Class A having low and simple motion, Class B with medium to high motion and Class C with high and complex motion as given in Table 1.

2.1 Determination of Stationarity

Stationarity refers to the stillness between consecutive frames in the temporal direction. Regions having similar motion in consecutive frames are also considered temporally stationary. The MBs which are temporally stationary usually get encoded in the SKIP or in the P_{16x16} mode whereas MBs with large motion get encoded with smaller block sizes. The simplest method of temporal prediction is to use the previous frame as the prediction for the current frame.

Table 1: Different Classes of Sequences.

Type	CIF Sequence	QCIF Sequence
Class A	News, Mother and daughter Container, Hallmonitor	Suzie, Claire Missamerica
ClassB	Foreman, Coastguard Harbour, Ice	Foreman, Silent Crew
Class C	Mobile, Flower Tempete, Stefan	Mobile, Football Soccer

Figure 2: Difference image of frame 10 and 11 of Mother and Daughter sequence.

Fig.2 shows the difference frame formed by subtracting frame 10 from frame 11 of the Mother and Daughter (MaD) sequence. Regions which have little or no motion have zero or low frame difference residual values represented by the grey regions and are encoded with larger block sizes. The light and dark grey areas correspond to positive and negative differences representing higher motion activity and hence use smaller block sizes. For each MB in the current frame (MB_C), the residual block (R_{DF}) is obtained from the collocated MB in the previous frame (MB_P) as

$$R_{DF}(i,j) = MB_C(i,j) - MB_P(i,j), \quad i,j = 1,\ldots,16 \tag{1}$$

From full search ME it is observed that at low values of QP, MBs encoded in SKIP mode have very low values of residuals in R_{DF}. As QP increase, MBs in the SKIP mode have larger values of residuals in R_{DF}. Fig.3 shows that when encoded with SKIP mode, a large number of residuals in a MB have values that are below 2. It is also observed that the MBs having large motion have high valued residuals in R_{DF}. Another observation is that the values of residuals for QP values upto 24, the MBs encoded in the SKIP mode have more than 90% residuals with absolute values that are usually below 2. In the proposed method, these prop-

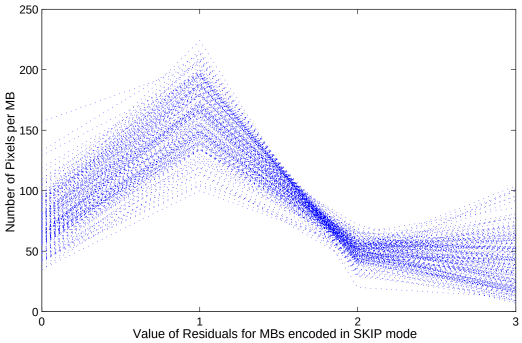


Figure 3: Residuals in R_{DF} for News sequence (CIF).

erties are exploited. For each MB a weight is assigned depending upon the QP and the values of the residuals in R_{DF}. MBs with low values of residuals are given lower weights indicating a region with little motion whereas MBs with large values of residuals are assigned higher weights indicating high to complex motion. Thus a threshold TH_{24} for QP values upto 24 is defined and is taken to be equal to 2. For higher values of QP, the threshold is defined as

$$TH_{QP} = \begin{cases} TH_{24} & \text{if} \quad QP \leq 24 \\ \text{floor}[\frac{1}{4}(QP - 24) + TH_{24}] & \text{if} \quad QP > 24 \end{cases}$$

Hence with the increase of QP the threshold also increases. In the proposed method, for each MB the fraction of the residues N in R_{DF} that are below the TH_{QP} is determined where N is given by

$$N = \frac{\text{Number of residues in } R_{DF} \text{ below } TH_{QP}}{\text{Total number of residues in } R_{DF},}$$

A difference frame weight DF_{wt} is assigned to each MB based on the value of N and is given in Table 2 below.

Table 2: Difference Frame Weights DF_{wt}.

N range	> 0.9	0.8-0.9	0.7-0.8	0.6-0.7	< 0.6
DF_{wt}	0	1	2	3	4

2.2 Determination of Homogenous MB

Natural videos have many homogeneous regions. Homogeneous regions have similar spatial properties. These regions in most cases get encoded with larger block size. Regions with more complex texture get encoded with smaller block sizes. If the homogeneity of a MB is detected early, then decision on the encoding mode can be taken. There exist many techniques for detecting homogenous regions. In this paper the homogenous region detection is based on the edge information as video object have strong edges. The edge based homogeneous region detection has been proposed in (Wu et al., 2005).

The edge map is created for each frame using the Sobel operator. Each pixel in the block will be associated with an edge vector containing edge direction and amplitude. The edge vector is defined as $D_{i,j} = \{dx_{i,j}, dy_{i,j}\}$ where

$$dx_{i,j} = p_{i-1,j+1} + 2 \times p_{i,j+1} + p_{i+1,j+1} - p_{i-1,j-1} - 2 \times p_{i,j-1} - p_{i+1,j-1}, \quad (2)$$

$$dy_{i,j} = p_{i+1,j-1} + 2 \times p_{i+1,j} + p_{i+1,j+1} - p_{i-1,j-1} - 2 \times p_{i-1,j} - p_{i-1,j+1}, \quad (3)$$

Here $dx_{i,j}$ and $dy_{i,j}$ are the degrees of differences in vertical and horizontal directions. The amplitude of the edge vector and the sum of the amplitudes are computed as

$$Amp(D_{i,j}) = |dx_{i,j}| + |dy_{i,j}|, \quad (4)$$

$$S_{Amp} = \sum_{i=1,j=1}^{16,16} Amp(i,j), \quad (5)$$

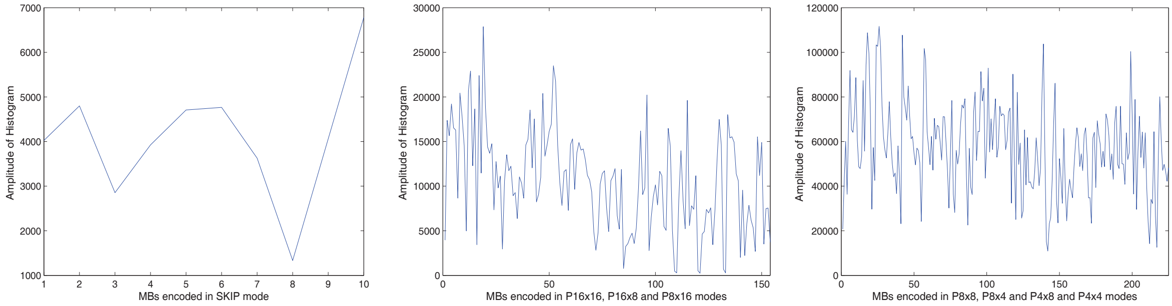
Homogenous regions will have a small value of S_{Amp} whereas this value increases as the MB becomes less homogeneous. The value of S_{Amp} is plotted and is shown in Fig. 4 where the edge histogram statistics are shown for the Mobile sequence. It clearly shows that for the SKIP mode the value of S_{Amp} is below 5000. For large block sizes, S_{Amp} is generally below 30000. For smaller block sizes, this value is significantly large. In the proposed algorithm, weights are assigned based on the values of S_{Amp} which are obtained from the study of the histogram characteristics for different sequences. Some ranges are defined. Lower weight is given for the MB if S_{Amp} is low and higher weight for larger values of S_{Amp}. The weights assigned are given in Table 3.

Table 3: Homogeneous MB Weights Hom_{wt}.

S_{Amp} Range	< 5000	5001-15000	15001-20000	20001-25000	> 25000
Hom_{wt}	0	1	2	3	4

2.3 Determination of Predicted MV

For a MB, the predicted MVs (pmv) are obtained from the MVs of the neighboring MBs (JVT.G050r1, 2003). Motion vectors for neighboring partitions are often highly correlated. Thus the calculated pmvs give a good indication of the amount of possible motion in the MB. Higher pmvs signal the probability of higher motion for the MB and vice versa. In the proposed paper, a pmv weight (pmv_{wt}) is introduced where weights are assigned for an MB depending



(a) SKIP mode. (b) P_{16x16},P_{16x8},P_{8x16} mode. (c) P_{8x8},P_{8x4},P_{4x8},P_{4x4} mode.

Figure 4: Histogram Amplitude of Frame 2 of Mobile sequences for different encoding modes.

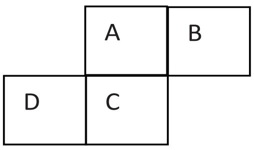
upon the pmv for the MB. For low values of pmvs smaller weight is assigned and for higher values larger weight is assigned and is given in Table 4 where pmv_x and pmv_y are predicted MVs in in the horizontal and vertical directions respectively.

Table 4: Predicted MV Weights pmv_{wt}.

$\max\lvert pmv_x\rvert$ or $\max\lvert pmv_y\rvert$ or both	0	1-2	3-4	5	>5
pmv_{wt}	0	1	2	3	4

2.4 Determination of Neighboring Mode for MB

It is observed that the modes of the neighboring MBs are often correlated.



Mode	Value Assigned
SKIP	0
P16x16	1
P16x8	2
P8x16	3
P8x8	4
P8x4	5
P4x8	6
P4x4	7
INTRA	8

Figure 5: Neighboring MBs: C is the current MB.

Referring to Fig.5, let C be the current MB and A, B and D be the neighboring MBs that have already been encoded. In the proposed algorithm, the likely mode for C is predicted from the modes of A, B and D and the following relation is used

$$Neigh_{MODE} = \text{median}\{A_{MODE}, B_{MODE}, D_{MODE}\}$$

Depending upon the value of $Neigh_{MODE}$ for a MB, a neighboring mode weight (NM_{wt}) is defined. If $Neigh_{MODE}$ for the MB indicate large block size partition then smaller weight is assigned to NM_{wt} and higher weight is assigned if block partition is small. The NM_{wt} for the MB is as given in the Table 5.

Table 5: Neighboring Mode Weights NM_{wt} .

$Neigh_{MODE}$	SKIP	P_{16x16}	P_{16x8} P_{8x16}	P_{8x8} P_{8x4},P_{4x8}	P_{4x4} INTRA
NM_{wt}	0	1	2	3	4

2.5 Overall Algorithm

For each MB, the different weights are first determined and the $Total_{wt}$ is obtained for each MB.

$$Total_{wt} = [\ DF_{wt} \quad Hom_{wt} \quad pmv_{wt} \quad NM_{wt},\ ]$$

From the weights in $Total_{wt}$, a final weight $Final_{wt}$ is obtained for each MB as follows:

- If at least three weights in $Total_{wt}$ are equal to say x where x=0,1,2,3,4 then $Final_{wt}$=x;
- If any two weights in $Total_{wt}$ are equal then $Final_{wt}$=ceil (median ($Total_{wt}$));
- If all the weights in $Total_{wt}$ are unequal then $Final_{wt}$=max($Total_{wt}$);

Based on this value of $Final_{wt}$, the decision on the encoding mode is taken. A low value of $Final_{wt}$ for a MB suggest that the MB is homogeneous with little motion and will be encoded with larger block size. A high low value of $Final_{wt}$ indicate higher motion and complexity and will be encoded using smaller block sizes. The mode selection for each MB is given in Table 6.

Table 6: Final Encoding Modes for MBs.

$Final_{wt}$	0	1	2	3	4
Mode	SKIP	P_{16x16}	P_{16x8} P_{8x16}	P_{8x8} P_{8x4},P_{4x8}	P_{4x4} INTRA

3 RESULTS AND DISCUSSIONS

This section compares the results of the proposed algorithm with the previously reported Wu. *et al.*'s algorithm (Wu et al., 2005). Results are presented as improvements over the standard H.264/AVC benchmark JM12.4. The experiments were carried out using some common video sequences of different classes at the CIF and QCIF resolution. The configuration used is the baseline profile, motion search range of ± 16, sequence type IPPP and one reference frame. Only the first frame is intra coded and QP used are 24, 28, 32 and 36 as per the recommended simulation conditions in (Sullivan and Bjontegaard, 2001). Comparisons are made in terms of distortion and percentage differences in rate and time taken for encoding. To evaluate the average encoding performance over a range of QPs, the differences in PSNR (ΔPSNR) in dB and bitrate (ΔRate (%)) are calculated according to numerical averages between RD curves as given by Bjontegaard (Bjontegaard, 2001).

3.1 Distortion and Compression Ratio Comparisons

Table 7 lists the performance of the proposed algorithm in comparison to JM12.4 implementation and Wu *et al.*'s algorithm (Wu et al., 2005). The results are arranged for different classes of sequences. Class A sequences have simple motion, Class B sequences have medium to high motion and Class C sequences have high to complex motion. The trend in the results shows that for the sequences, there is only a marginal loss in the PSNR performance. There is an average 0.03 dB loss in PSNR in the proposed method. Also the average bitrate increase is 1.5%. The results demonstrate the effectiveness of the proposed algorithm.

3.2 Comparison with JM12.4 Modes

Table 8 shows the percentage of MBs that are encoded (using the proposed method) in the same mode as the JM12.4 for different QP. The results show that the proposed method is effective as it has been able to maintain the same final encoding modes as the JM12.4 to a large extent.

3.3 Computational Speedup

Table 7 shows the percentage reduction in encoding time ΔT(%) for sequences of different classes. The time saving obtained depends upon the type of sequence. An increased saving is noted for Class A sequences where for some sequences 90% time saving is noted whereas time saving obtained for Class C sequences is comparatively low. This is due to the fact that Class A sequences have low motion complexity and hence a large number of MBs get encoded with larger block sizes. The saving in time is achieved as the decision on the final mode for encoding is taken prior to the ME and for each MB at the most only three modes are searched. However, for all sequences, the proposed algorithm exhibits a good computational saving regardless of the QP setting.

Table 7: Performance Comparison For Different Sequences.

Class	Sequence	Performance Comparison					
		Proposed			Wu *et al.*'s (Wu et al., 2005)		
		ΔPSNR	ΔRate (%)	ΔT (%)	ΔPSNR	ΔRate (%)	ΔT (%)
CIF Class A	News	-0.05	2.29	88.42	0.02	2.01	39.23
	MaD	0.02	2.86	81.38	0.01	0.85	43.21
	Container	0.01	1.32	60.33	0.05	1.55	46.18
	Hall	0.09	0.40	76.84	0.06	0.90	34.67
Class B	Foreman	0.08	0.82	72.79	0.07	1.22	34.90
	Coastguard	-0.05	0.66	65.07	0.05	0.54	26.30
	Ice	0.05	3.14	85.17	0.07	1.29	45.68
	Harbour	0.06	1.88	65.37	0.05	1.10	21.56
Class C	Flower	0.08	0.19	76.43	0.05	2.98	36.85
	Stefan	0.08	1.59	70.59	0.02	1.46	32.25
	Tempete	0.02	0.82	71.88	0.09	1.02	27.21
	Mobile	0.04	1.88	62.41	0.09	1.69	12.35
	Average	**0.03**	**1.48**	**73.05**	**0.05**	**1.38**	**33.36**
QCIF Class A	Claire	-0.01	-2.81	90.69	-0.01	-0.95	47.35
	MissAmerica	-0.07	-1.40	89.87	0.02	0.91	48.23
	Suzie	0.04	1.21	77.73	0.05	0.49	42.91
Class B	Foreman	0.08	1.04	67.73	0.04	1.29	30.25
	Silent	0.03	2.40	82.46	0.09	0.79	42.62
	Crew	0.08	2.17	63.86	0.05	1.65	19.64
Class C	Football	0.10	1.50	65.30	0.05	1.86	32.42
	Mobile	0.14	1.62	28.09	0.07	1.50	15.32
	Soccer	0.04	1.17	63.67	0.03	3.03	20.19
	Average	**0.04**	**0.76**	**69.93**	**0.04**	**1.17**	**33.21**

MaD: Mother and Daughter
ΔPSNR(+/-): picture quality loss/gain measured in dB
ΔRate(+/-): bitrate increase/decrease measured as a %
ΔT(+/-): encoding time saving/loss measured as a %

Table 8: % of MBs encoded in the same mode w.r.t FME.

Sequence (CIF)	Quantization Parameter			
	24	28	32	36
News	82.10	75.12	84.12	86.01
MaD	66.12	87.86	89.6	90.66
Container	76.50	86.01	89.30	89.15
Hall	59.11	68.69	72.45	83.33
Foreman	46.11	58.71	64.90	76.23
Coastguard	62.33	67.83	78.84	79.47
Ice	67.42	73.99	71.97	72.98
Harbour	46.21	52.66	55.08	63.33
Flower	69.70	70.71	69.44	72.14
Stefan	22.73	57.32	66.31	68.99
Tempete	52.27	59.75	68.23	70.43
Mobile	29.80	34.34	49.39	66.19

4 CONCLUSIONS

In this paper, an improved mode decision algorithm for H.264/AVC video coding standard has been proposed based on the weights assigned for different characteristics of the MB. Stationarity based weights are obtained from frame difference residuals. Homogeneity based weights are obtained from edge histograms parameters. The pmv weight and the neighboring mode weights are obtained from the correlation the the MB with neighboring MBs. Results of simulations carried out on different sequences demonstrate that there is very little degradation of the PSNR and the bitrate performance in the proposed algorithm despite a large saving in encoding time and computation. The average encoding time saving is around 70%. The proposed method achieves almost the same coding performance in terms of picture quality and compression ratio as that of the H.264/AVC standard and improves on Wu *et al.*'s (Wu et al., 2005) algorithm. Hence, for a variety of sequences with varying motion activities, the proposed algorithm gives a consistent performance on encoding time reduction, computational saving and coding efficiency.

REFERENCES

Bjontegaard, G. (2001). Calculation of average PSNR difference between RD-curves. *ITU-T VCEG, Doc. VCEG-M33*, pages 1–2.

Ganguly, A. and Mahanta, A. (2010). Fast Mode Decision Algorithm for H.264/AVC using Edge Characteristics of Residue Images. In *Proceedings of ACM , ICVGIP*, pages 512–517.

Huang, Y.-H., Ou, T.-S., and Chen, H. H. (2010). Fast Decision of Block Size, Prediction Mode and Intra Block for H.264 Intra Prediction. *IEEE Trans. on Circuits and System for Video Technology*, 20(8):1122–1132.

Jing, X. and Chau, L. P. (2004). Fast Approach for H.264 Inter Mode Decision. *Electronic Letters*, 40(17):1050–1052.

JVT.G050r1 (2003). 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), document JVT-G050r1 2003. Technical Report document JVT-G050r1, Joint Video Team of ITU-T and ISO/IEC JTC 1.

Kim, D., Kim, J., and Jeong, J. (2006). Adaptive Macroblock Mode Selection for Reducing the Encoder Complexity in H.264. In *ACVIS*, number 4179 in LNCS, pages 396 – 405.

Kim, G. Y., Moon, Y. H., and Kim, J. H. (2004). An Early Detection of All-Zero DCT Block in H.264. In *Int. Conf. Image Process (ICIP)*, volume 1, pages 453–456.

Lee, Y. L., Lee, Y. K., and Park, H. (2006). A Fast Motion Vector Search Algorithm for Variable Blocks. In *ACVIS*, number 4179 in LNCS, page 311 322.

Liu, Z., L.Shen, and Zhang, Z. (2009). An Efficient Intermode Decision Algorithm Based on Motion Homogeneity for H.264/AVC. *IEEE Trans. on Circuits and System for Video Technology*, 19(1):128–132.

Park, I. and Capson, D. W. (2008). Improved Inter Mode Decision Based on Residue in H.264/AVC. In *International Conference on Multimedia & Expo*, pages 709–712.

Richardson, I. E. G. (2003). *H.264 and MPEG-4 Video Compression: Video Coding for Next-generation Multimedia*. John Wiley and Sons.

Shen, L., Liu, Z., Zhang, Z., and Shi, X. (2008). Fast Inter Mode Decisions Using Spatial Property of Motion Field. *IEEE Trans. on Multimedia*, 10(6):1208–1214.

Sullivan, G. and Bjontegaard, G. (2001). Recommended Simulation Common Conditions for H.26L Coding Efficiency Experiments on Low-Resolution Progressive-Scan Source Material. *ITU-T VCEG, Doc. VCEG-N81*, pages 1–3.

Wang, H., Kwong, S., and Kok, C.-W. (2007). An Efficient Mode Decision Algorithm for H.264/AVC Encoding Optimization. *IEEE Trans. on Multimedia*, 9(4):882–888.

Wiegand, T., Sullivan, G. J., Bjontegaard, G., and Luthra, A. (2003). Overview of the H.264/AVC Video Coding Standard. *IEEE Trans. on Circuits and System for Video Technology*, 13(7):560–576.

Wu, D., Pan, F., Lim, K. P., Wu, S., Li, Z. G., Lin, X., Rahardja, S., and Ko, C. C. (2005). Fast Intermode Decision in H.264/AVC Video Coding. *IEEE Trans. on Circuits and System for Video Technology*, 15(6):953–958.

Zeng, H., Cai, C., and Ma, K. (2009). Fast Mode Decision for H.264/AVC Based on Macroblock Motion Activity. *IEEE Trans. on Circuits and System for Video Technology*, 19(1):1–11.