A Fast and Efficient Inter Mode Decision Algorithm for the H.264/AVC Video Coding Standard

Skoudarli Abdellah[1], Nibouche Mokhtar[2] and Serir Amina[1]
[1]*FEI USTHB, BP 32 El Alia Bab Ezzouar, Alger, Algeria*
[2]*Frenchay Campus, Coldharbour Lane, Bristol BS16 1QY, UWE, U.K.*

Keywords: H.264/AVC, Mode Decision, Inter/Intra Prediction, Homogeneity, Stationarity, Complexity Reduction.

Abstract: The H.264/AVC video coding standard is used in a wide range of applications from video conferencing to high-definition TV. Compared to the previous standard, the H.264/AVC has significantly better performance in terms of PSNR and visual quality at the same bit rate. It uses a complex mode decision technique based on rate-distortion optimization (RDO). Therefore, this technique introduces a high computational complexity. However, the computational complexity is one key challenge for the high efficient compression. In order to reduce the H.264/AVC complexity a new efficient and fast mode decision algorithm, based on the spatial homogeneity and temporal stationary characteristics of the current macroblock, is proposed in this paper. The experimental results show that the proposed algorithm is able to reduce up to 66,90 % of the computational complexity compared to the high complexity algorithm in the JM16.1 reference software with tolerant performance degradation.

1 INTRODUCTION

The H.264/AVC encoder represents the latest video coding standard. Compared to the previous standard, the H.264/AVC has significantly better performance in terms of PSNR and visual quality at the same bit rate (Richardson, 2003).To improve the coding efficiency, H.264/AVC adopts new coding tools such as multiple reference frames, sub-pel accuracy motion estimation (ME), in loop deblocking filter, Variable Block Size (VBS) (Wiegand, 2003). These tools permit a higher coding efficiency in comparison to prior standards. Unfortunately, this comes at the expense of increased complexity. In fact, the encoder complexity increases tremendously. This leads to long encoding processing time and huge power consumption, which make the deployment of the algorithm in real time applications and embedded systems more difficult.

As the improved coding efficiency comes at the expense of added complexity to the coder/decoder, H.264/AVC utilizes some methods to reduce the implementation complexity. One way to speed up the H.264/AVC encoding time is to reduce the complexity of macroblock mode selection.

Several fast mode decision algorithms have been developed to simplify the mode selection by exploiting the features of regions in a video sequence.

In (Bharanitharan, 2010) a classified region algorithm that analyses the spatial and temporal homogeneity of the macroblock is used. The proposed algorithm is based on a computation of the gradient function of the current macroblock (MB). The intensity differences in vertical and horizontal directions are computed. In (Young Lee, 2012) the proposed inter-mode decision scheme determines the best coding mode of a given macroblock by predicting the best mode from neighboring MBs in time and in space and by estimating its rate-distortion cost RD Cost from the MB in the previous frame. In (Ri, 2009) the proposed method reduces the number of candidate modes by detecting spatially and temporally homogeneous regions and analyzing motion costs for inter modes and intra prediction costs for intra modes.

Other relevant approaches for fast algorithms that consider adaptive thresholding method were proposed and demonstrate that can improve performance for a wide range of video sequences.

In (Ren, 2008a) an adaptive threshold for early termination is introduced with fast multiple reference frame motion estimation based on texture and motion information. In (Martinez-Enriquez,

Abdellah S., Mokhtar N. and Amina S..
A Fast and Efficient Inter Mode Decision Algorithm for the H.264/AVC Video Coding Standard.
DOI: 10.5220/0004527800790085
In *Proceedings of the 10th International Conference on Signal Processing and Multimedia Applications and 10th International Conference on Wireless Information Networks and Systems* (SIGMAP-2013), pages 79-85
ISBN: 978-989-8565-74-7

2010) an algorithm based on the rate distortion cost (RDCost) statistics is also proposed. The differences between rate distortion costs statistics for each mode are used to obtain successive adaptive thresholds to establish several early terminations. In (Ren, 2008b) a computationally efficient mode prediction and selection approach is proposed based on the following attributes: both the spatial and temporal information are used to achieve early termination using adaptive thresholds, inclusion of a modulator capable of trading off computational efficiency and accuracy, and a homogeneous region detection procedure for 8x8 blocks based on adaptive thresholds.

Taking into consideration the spatial homogeneity and the temporal stationarity of the current macroblock and in order to reduce the computational complexity, a fast and efficient mode decision algorithm is proposed in this paper.

This paper is organized as follows: Section 2 describes the mode decision algorithm for the H.264/AVC video coding standard. Section 3 highlights the motivation behind the proposed work. Section 4 is dedicated to the explanation of the algorithm of the proposed mode decision. The performance and the experimental results are shown in section 5. Finally, a conclusion sums up the findings of paper.

2 MODE DECISION FOR THE H.264/AVC

The H.264 standard supports various intra prediction modes and inter prediction modes techniques, where most of them contribute to the coding efficiently.

2.1 Intra Prediction

The intra-prediction exploits the spatial redundancy between adjacent macroblocks in a frame. There are three different intra prediction modes in the H.264/AVC standard. Intra4x4 (I4_MB), Intra8x8 (I8_MB) and Intra 16x16 (I16_MB), respectively.

The intra16×16 mode has four directional predictions:

- Intra_16×16_Vertical,
- Intra_16×16_Horizontal,
- Intra_ 16×16_Plane
- and Intra_ 16 × 16_DC.

The intra4×4 mode has nine different directional predictions:

- Intra_4×4_Vertical,
- Intra_4×4_Horizontal,
- Intra_4×4_Diagonal_Down_Left,
- Intra_4×4_Diagonal_Down_Right,
- Intra_4×4_Vertical_Right,
- Intra_4×4_Horizontal_Down,
- Intra_4×4_Vertical_Left,
- Intra_4×4_Horizontal_Up,and Intra_4×4_DC.

The intra8×8 mode has four directional predictions:

- Intra_8×8_Vertical,
- Intra_8×8_Horizontal,
- Intra_ 8×8_Plane
- and Intra_8 ×8_DC.

In inter-frame coding, intra-modes are also taken into consideration for seeking the best coding mode in order to maintain higher encoding efficiency.

2.2 Inter Prediction

The inter prediction exploits temporal redundancy between macroblocks in different frames. There are in total seven different block sizes that can be used in inter prediction (16x16,16x8, 8x16, 8x8, 8x4, 4x8 and 4x4). These different block sizes form two level of hierarchy inside a macroblock (MB). The first level includes block size of 16x16, 16x8, 8x16. In the second level, in which a MB is specified as P8x8, each 8x8 block can be one of the SubMB 8x8, 8x4, 4x8, 4x4, respectively. The relationship between these different block sizes is illustrated in Figure 1.

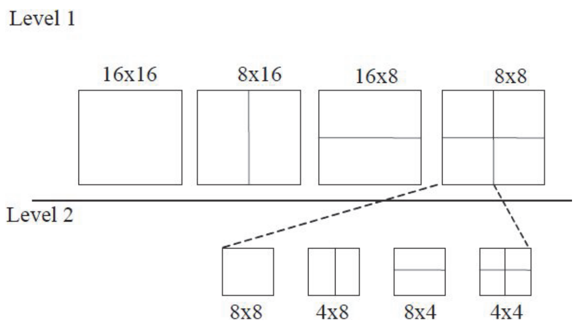


Figure1: Different MB partitions and MB sub-partitions.

The whole procedure of inter-modes and intra-modes in inter-frame coding consists of three parts:

- Calculate the minimum cost of inter-prediction modes.
- Calculate the minimum cost of intra-prediction modes.
- Compare the minimum cost of inter-modes with,

at least, one of intra-modes to decide on the final coding mode. If minimum cost of intra-modes is less than the inter-modes one, then the final coding mode will be intra and vice versa.

2.3 Mode Decision

The rate distortion cost (RDCost), which helps in deciding the best prediction mode, is computed using the Lagrangian function *J* as follow:

$$J(s,c,MODE \mid QP, \lambda_{MODE}) = SSD(s,c,MODE \mid QP) + \lambda_{MODE} . R(s,c,MODE \mid QP) \quad (1)$$

Where s is the original block and c is its associated reconstruction, respectively. The sum of the squared difference (SSD) between s and c is given by:

$$SSD(s,c,MODE \mid QP) = \sum_{x=1,y=1}^{16,16} (s_Y[x,y] - c_Y[x,y,MODE \mid QP])^2 + \sum_{x=1,y=1}^{8,8} (s_U[x,y] - c_U[x,y,MODE \mid QP])^2 + \sum^{8,8} (s_V[x,y] - c_V[x,y,MODE \mid QP])^2 \quad (2)$$

The Lagrangian multiplier λ is defined by:

$$\lambda_{MODE,P} = \begin{cases} 0.85 \times 2^{QP/3} \\ \max(2, \min(4, \frac{QP}{6})) \times \lambda_{MODE,P} \end{cases} \quad (3)$$

Where R(s,c,MODE|QP) represents the number of bits (MODE& QP),

MODE={SKIP, 16x16,16x8, 8x16, P8x8}
and QP is the quantization parameter:
QP={0,1, 2, 3, … ,50,51}

The rate distortion cost of all modes in the coding mode is calculated. The mode decision is made by selecting the modes having minimum rate distortion costs, limiting thus the candidate prediction modes to a small subset. The approach leads to a reduction of the computational complexity. This concept constitutes the main idea exploited in this paper.

3 MOTIVATION

To make an algorithm faster is an important aspect in real time video applications and high performance low power embedded systems. These challenges, which are the main motivation behind this proposed research, are dealt with by exploiting spatial homogeneity and Temporal stationarity in video to reduce complexity in encoders.

One of the reasons for adopting different modes with variable block sizes in H.264/AVC is to represent the motion more accurately. In general, homogeneous and/or stationary regions with motionless are more likely to be coded using large block sizes, and non-homogeneous or non stationary regions with motion are to be coded using smaller block sizes as illustrated in Figure 2. However, It is observed in natural video sequences, that there exist lots of homogeneous and stationary regions and when objects move, most parts of these objects move in the similar direction. If we could detect these homogeneous and stationary regions in the early stages, a significant time could be saved for the motion estimation (ME) search and for the rate distortion optimization (RDO) computations.

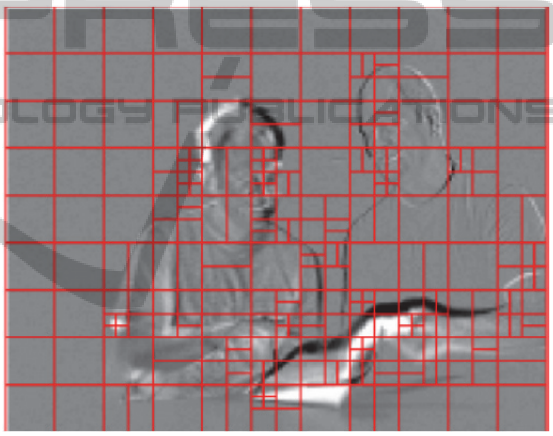


Figure 2: Homogeneity in a frame.

4 PROPOSED ALGORITHM

The proposed fast mode decision algorithm is based on some characteristics of the MB and its collocated MB in the sequence. Before processing each macro-block, low level features including spatial homogeneity and temporal stationarity are extracted. These features are preferred due to their lower complexity towards having a real time implementation.

If a macroblock is homogeneous or stationary, only large block sizes can be used and small block sizes can be skipped, which is very useful to reduce the computation complexity.

4.1 Skip Mode

In the proposed algorithm, the SKIP mode, where no motion and no residual information are encoded, is differentiated from other MB types. Thus, in the first time, a highest priority is given to this mode.

If a macro-block is encoded in SKIP mode, the following conditions should be all satisfied :

- The best motion compensated block size is 16x16.(inter mode 16x16).
- The reference frame is the previous frame.
- The best motion vector is the predicted motion vector.
- The transform coefficients of 16x16 block size are all quantized to zero.

To decide the skip mode in terms of the above conditions as quickly as possible becomes the key of our algorithm. This is achieved by testing the homogeneity of 16x16 MBs at level one.

4.2 Homogeneity Detection

In general, homogeneous regions have similar spatial properties and refers to texture similarities inside a single video frame. There are many techniques to detect spatial homogeneity in an image. One method to detect spatial homogeneity is to exploit edge information. In (Ganguly, 2010) and (Wu, 2005) edge detection is used to find spatially homogeneous blocks. In this approach an edge map is created for each frame using 3x3 Sobel operator. each pixel in the block will be associated with an edge vector containing edge amplitude and edge direction. Another method is proposed in (Rungta, 2010) to evaluate spatial homogeneity by using the variance of the macro-block. These two approaches introduce a lot of additional complexity in the form of pre-calculation cost.

An alternative approach is based on testing pixels values of the current MB to decide if the MB is homogeneous or not. The testing homogeneity procedure is summarized as follows:

i. The mean value of the pixels in a NxN block is calculated:

$$\mathrm{Mean} = \frac{1}{\mathrm{NxN}} \sum_{i=1}^{N} \sum_{j=1}^{N} p(i,j) \tag{4}$$

Where, p(i , j) is a pixel intensity at position (i, j) in the N x N block.

ii. The absolute difference between each pixel in the block and the mean value of block is calculated.

$$\mathrm{ADPM} = |p(i,j) - \mathrm{Mean}| \tag{5}$$

iii. The number of pixels in the block satisfying test condition 1 (Num_Pix_Less_Th1) is computed

Test condition 1:

$$\mathrm{ADPM} < \mathrm{Th1} \tag{6}$$

Where Th1 is a predefined threshold.

iv. Test the homogeneity of the block as follows:

Test condition 2:

If :Num_Pix_Less_Th1 < Th2,

- then the block is homogeneous,
- otherwise the block is non-homogeneous.

Where, Num_Pix_Less_Th1 represent the number of pixels in the block NxN less than Th1.

Th2 is the threshold depending on both block size and a predefined coefficient α(%).

$$\mathrm{Th2} = \mathrm{NxNx}\alpha \tag{7}$$

In our experiments the coefficient α is set to 5%. The threshold Th1 is set equal to 14 in the case of a macroblock MB and is set equal to 4 in the case of subMB.

In the first stage, the homogeneity of the 16x16 MB at level 1 is tested. depending on the test condition the SKIP mode or 16x16, 16x8, 8x16 modes are selected.

In the second stage, the homogeneity of 8x8 subMBs at level 2 is tested. Early sub-partition termination is considered and the best mode, as mode 4, is selected.

4.3 Stationarity Detection

Temporal stationarity refers to the stillness between consecutive frames in the temporal direction. Our proposed stationarity method detection is based on the SAD sum of absolute difference between the current macroblock in the current frame and the collocated macroblock in the previous frame. The SAD is computed by:

$$\mathrm{SAD} = \sum_{i=1}^{N} \sum_{j=1}^{N} |p_{cur}(i,j) - p_{col}(i,j)| \tag{8}$$

where: $p_{cur}(i,j)$ is the pixel in the current MB and: $p_{col}(i,j)$ is the pixel in the co-located MB.

The temporal stationarity is tested by comparing The SAD with an appropriate threshold Th_S. Then, if the SAD is less than a certain threshold Th_S, the macroblock will be encoded in the SKIP mode or in P_16x16 mode, thus all the other modes can be skipped.

Several experiments were done for different types of video sequences at different QP values and the different threshold were analyzed by the different degradation of video quality to empirically determine the reliable threshold. These experiments shown that these Th_S values achieves a good and consistent performance. Then the average threshold Th_S can be set according to the following tabulated values:

Table 1: Threshold Th_S according to QP values.

QP	24	28	32	36
Th_S	750	950	1100	1250

4.4 Overall Algorithm

The proposed algorithm to encode a MB is developed as follows:

Step 1: Test the homogeneity at level one of the current MB 16x16.

Step 2: If the 16x16 MB homogeneity is less than Th_H1 then terminate partition and choose SKIP mode. (Th_H1=Th2)

Step 3: If the 16x16 MB homogeneity is between Th_H1 and Th_H2 then perform RD optimization on the 16x16, 16x8, 8x16 blocks. (Th_H2=Th2*0.85)

Step 4: If the 16x16 MB is non-homogeneous, then test stationarity of the MB.

Step 5: If the 16x16 MB is stationary then perform RD optimization on the 16x16, 16x8, 8x16 blocks.

Step 6: If the 16x16 MB is not stationary then test homogeneity at level 2 of each 8x8 block in the current MB.

Step 7: If one of the 8x8 block is homogeneous, skip the partition of this block and choose the 8x8 as best mode. After, test other 8x8 subpartitions if not homogeneous perform RD optimization on 8x8, 8x4, 4x8 and 4x4 blocks.

Step 8: Otherwise, perform a complete RD optimization on the MB and choose the best mode among all the modes.

5 EXPERIMENTAL RESULTS

The proposed complexity reduction algorithm was applied to encode test sequences (Foreman, Carphone, Salesman, Hall, Container and Akiyo). For the purpose of evaluation, the reference software JM16.1 (JVT Reference software) has been used. Based on the proposed approach and for testing purposes, a modified version of the JM16.1 software has been developed. The original and modified JM16.1 was executed on an Intel Core 2Duo based computer with 4 Go RAM under windows XP Professional operating system.

The test conditions are as follows:

- GOP structure is IPPP;
- The number of frames in a sequence is 100;
- The Hadamard transform is adopted;
- The Fast Full Search algorithm is adopted;
- Reference frame number equals 5;
- MV resolution is ¼ pel;
- RD optimization is enabled;
- CABAC is adopted;

The encoding efficiency of the proposed algorithm is evaluated according to these three parameters:

The encoding time saving rate: ΔTime(%)

$$\Delta\text{Time}(\%) = \frac{\text{Time}_{\text{proposed}} - \text{Time}_{\text{Ref}}}{\text{Time}_{\text{Ref}}} \quad (9)$$

The variation of video quality ΔPSNR(dB):

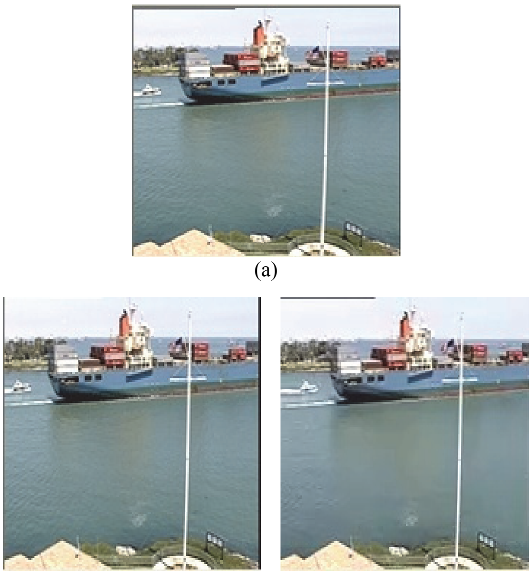
$$\Delta PSNR = PSNR_{Proposed} - PSNR_{Ref} \quad (10)$$

Where the Average PSNR of the sequence is defined as:

$$\text{PSNR} = \frac{4.\text{PSNR}_Y + \text{PSNR}_{Cb} + \text{PSNR}_{Cr}}{6} \quad (11)$$

The undulation rate of bits: ΔBit(%)

$$\Delta\text{Bit}(\%) = \frac{Bit_{Proposed} - Bit_{Ref}}{Bit_{Ref}} \quad (12)$$



(a) (b) (c)

Figure 3: Frame N° 50 of the container sequence (a)Source (b)reconstructed (c) decoded.

Table 2: Results for "IPPP Sequences (100 frames)" with QP=24.

Sequence	ΔPSNR(dB)	ΔBit(%)	ΔTime(%)
Akiyo	-0.31	+1.81	-64.56
Container	-1.61	+1.13	-48.43
Hall	-0.61	+1.15	-39.56
Carphone	-1.02	+2.09	-35.59
Salesman	-0.28	+2.10	-26.30
Foreman	-1.64	+2.34	-23.15
Average	-0,91	+1,77	-39,60

Table 3: Results for "IPPP Sequences (100 frames)" with QP=28.

Sequence	ΔPSNR(dB)	ΔBit(%)	ΔTime(%)
Akiyo	-0.13	+1.02	-65.11
Container	-0.82	+0.71	-48.22
Hall	-0.17	+1.6	-40.57
Carphone	-1.09	+1.42	-37.82
Salesman	-0.12	+1.81	-26.13
Foreman	-1.12	+2.21	-25.88
Average	-0,58	+1,46	-40,62

Table 4: Results for "IPPP Sequences (100 frames)" with QP=32.

Sequence	ΔPSNR(dB)	ΔBit(%)	ΔTime(%)
Akiyo	-0.10	-0.72	-65.09
Container	-0.72	+1.18	-49.32
Hall	-0.13	+1.81	-40.98
Carphone	-1.09	+1.51	-37.82
Salesman	-0.07	+1.12	-26.59
Foreman	-0.86	+1.83	-26.13
Average	-0,50	+1,12	-40,99

Table 5: Results for "IPPP Sequences (100 frames)" with QP=36.

Sequence	ΔPSNR(dB)	ΔBit(%)	ΔTime(%)
Akiyo	-0.09	-0.76	-66.90
Container	-0.45	+1.15	-51.24
Hall	-0.19	+1.82	-39.64
Carphone	-0.61	+1.41	-38.37
Salesman	-0.10	+0.89	-25.25
Foreman	-0.51	+1.91	-26.69
Average	-0,33	1,07	-41,35

The above experimental results indicate an efficient algorithm design that consider both computation complexity reduction and coding performance degradation. The proposed method is very close to JM Reference Software in low bit rate with less PSNR loss and less bit rate increase.

The rate distortion performance of the proposed method is shown in the following figures in the form of R-D curves.

From tables (2, 3, 4 and 5), our experimental results show that the proposed algorithm achieves 40,64% time saving on average. We can see that the bit rate increment and PSNR loss depend on the quantization parameter QP.

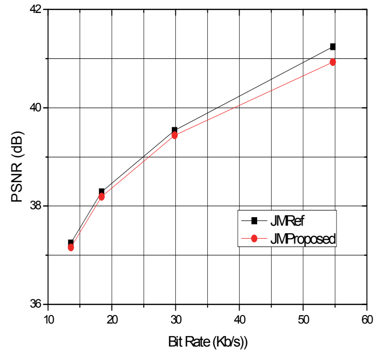


Figure 4: Rate distortion curves for JM Ref and proposed method. QCIF sequence: Akiyo.

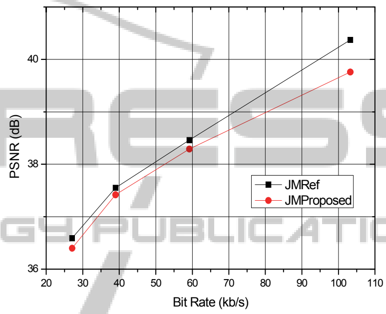


Figure 5: Rate distortion curves for JM Ref and proposed method. QCIF sequence: Hall.

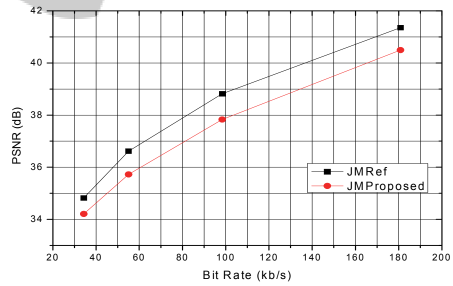


Figure 6: Rate distortion curves JM Ref and proposed method. QCIF sequence: Carphone.

From the rate distortion curves shown in figures (4, 5 and 6), we can also see that the rate distortion degradation is less in low bit rate than in high bit rate.

6 CONCLUSIONS

A simple and effective scheme for fast mode decision in the H.264/AVC video coding standard has been proposed in this research paper. The scheme exploits the spatial homogeneity and temporal stationary features in the macroblocks (MBs) to avoid unnecessary computation. The proposed method, as indicated by the experiments, provides the best trade-off between coding efficiency and speed. This simple and effective reduction of the encoder complexity will be very useful for real time implementations of the H.264/AVC standard.

However, the problem with this approach lies in the fixed thresholds. For the delivery of fast mode decision performance, the threshold values play a crucial role on the entire inter mode decision process. Adaptive thresholds based on spatial and temporal information can be obtained by analyzing the texture of the video signal and by analyzing motion information. So we can improve our method by adopting an adaptive thresholds used to detect spatial homogeneity and temporal stationarity of the macroblocks.

REFERENCES

Richardson I. E. G., 2003. "*H.264 and MPEG-4 Video Compression, Video Coding for Next Generation Multimedia*", John Wiley & Sons Publications, first Edition, England.

Wiegand, T., Sullivan, G. J., Bjontegaard, G., and Luthra, A., 2003. "Overview of the H.264/AVC video coding standard", *IEEE Transactions on Circuits and Systems for Video Technology*. July. Vol. 13, N°7, pp 560-576.

Bharanitharan, K., Liu, B., Yang, J., Vol. 2010, pp 1-10. "Classified Region Algorithm for Fast Intermode Decision in H.264/AVC Encoder". *EURASIP Journal on Advances in Signal Processing*.

Young Lee, J., Wook Park, H., Vol. 22, N°3,pp 393-402, March 2012. "Fast mode Decision Method Based on Motion Cost and Intra Prediction Cost for H.264/AVC", *IEEE Transactions on Circuits and Systems for Video Technology*.

Ri, S., Vatis, Y., Ostermann, J., Vol. 19, N°2,pp 302-306, February 2009. "Fast Inter-Mode Decision in an H.264/AVC Encoder Using Mode and Lagrangian Cost Correlation", *IEEE Transactions on Circuits and Systems for Video Technology*.

Ren, J., Kehtarnavaz, N., Budagavi, M., 2008a."A Fast Featured-Assisted Adaptive Early Termination Approach for Multiple Reference Frames Motion Estimation in H.264" *Springer Journal of Real Time Image Processing* ,vol 3, pp 77-88.

Martinez-Enriquez, E., Jiménez-Moreno, A., Diaz-de Maria, F., 2010. "An Adaptive Algorithm for Fast Inter Mode Decision in the H.264/AVC Video Coding Standard", *IEEE Transactions on Consumer Electronics*, May. Vol 56, N° 2,pp 826-834.

Ren, J., Kehtarnavaz, N., Budagavi, M., 2008b."Computationally Efficient Mode Selection in H.264/AVC Video Coding" *IEEE Transactions on Consumer Electronics*, May. Vol 54, N° 2,pp 877-886.

Wu, D., Pan, F., Lim, K. P., Wu, S. Z., Li, G., Lin, X., Rahardja, S., Ko, C. C., 2005. "Fast Intermode Decision in H.264/AVC Video Coding", *IEEE Transactions on Circuits and Systems for Video Technology*, July. Vol. 15, N°6, pp 953-958.

Ganguly A., Mahanta, A., 2010. "Fast Mode Decision Algorithm for H.264/AVC using Edge Characteristics of Residue Images", *ICVGIP'10*, 12-15 DECEMBER, Chenai, India.

Rungta, S., Verma, K., Shukla, A., 2010. "A Fast Mode Selection Algorithm Using Texture Analysis for H.264/AVC", *International Journal of Computer Sciences Issues,* July. Vol.7, Issue 4, N°9, pp. 40-44.

JVT Reference Software Version 16.1, available at http://iphome.hhi.de/suehring/tml/download/.