EFFICIENT INTRA MODE SELECTION USING IMAGE STRUCTURE TENSOR FOR H.264/AVC

Chiuan Hwang, ShinShan Zhuang, and Shang-Hong Lai

Department of Computer Science, National Tsing Hua University, Hsinchu, Taiwan
lai@cs.nthu.edu.tw

ABSTRACT

Intra mode decision and motion estimation for spatial and temporal prediction play important roles for achieving high video compression ratio in the latest video coding standard H.264/AVC. However, both components take most of the computational cost in the video encoding process. In this paper, we propose an efficient intra mode prediction algorithm based on image structure tensor analysis. The image structure tensor can provide the local image structure information for making the intra-mode decision. We show significant reduction of the computation time with negligible video quality degradation for H.264 video encoding by implementation into JM reference program .

Index Terms— *Image coding, image analysis, video coding, video compression, intra prediction, video codec.*

1. INTRODUCTION

To improve the computational efficiency of the intra-mode decision in H.264 video coding [1], there have been many different algorithms proposed recently. The traditional method is to run all intra modes for Luma and Chroma blocks, but it takes too much encoding time. Later, Cheng and Chang [2] used the block content characteristics, such as the block RD-cost correlation [2] between the mode and its neighboring modes, to reduce the search of nine Luma 4x4 modes to no more than six modes. In addition, Zhang et al. [3] computed the block gradient information to determine a rough edge direction in the local block. They compute the horizontal and vertical edge strength [3] to decide which direction may have the lowest RD-cost. Based on this direction, they used the corresponding directional mode and its neighboring two modes plus the DC mode to determine the mode with the lowest cost. In addition, Feng Pan et al.[4] applied a different method to predict the direction in a block. They collect statistics of the gradients for all pixels in a block and then choose the most dominant direction from the statistics as the prediction mode.

In this paper, we propose a new and efficient intra mode decision algorithm based on the image analysis of the local block. This local image analysis is called image structure tensor, which can provide the information of local image structure. We will use the image structure information to determine a set of candidate modes to reduce the search of the intra prediction modes.

2. INTRA PREDICTION IN H.264/AVC

The H.264/AVC standard [1] provides variable block size on motion estimation and more directional modes for intra prediction. For the Luma 16x16 component and Chroma 8x8 component, they support four modes listed in Table 1, including three directional modes and one DC mode.

Table 1. (a) Luma 16x16 prediction modes and (b) Chroma 8x8 prediction modes.

(a)

Mode	Mode description
0	Vertical
1	Horizontal
2	DC
3	Plane

(b)

Mode	Mode description
0	DC
1	Vertical
2	Horizontal
3	Plane

The modes for Luma 16x16 and Chroma 8x8 are different only with their orders. For Luma 4x4 sub-blocks, there are nine modes, including eight directional modes and one DC mode. The Luma 4x4 modes are listed in Figure 1.

Mode	Mode description
0	Vertical
1	Horizontal
2	DC
3	Diagonal_Down_Left
4	Diagonal_Down_Right
5	Vertical_Right
6	Horizontal_Down
7	Vertical_Left
8	Horizontal_Up

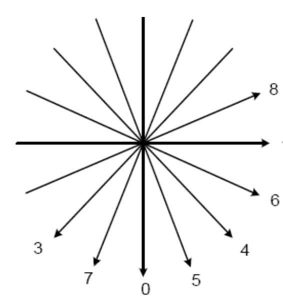

Figure 1. The nine Luma 4x4 intra prediction modes and their corresponding directions

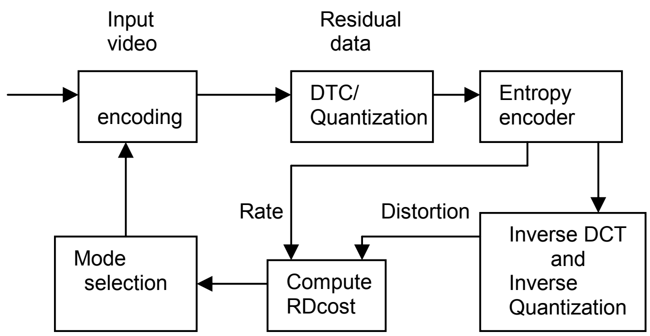

Figure 2. H.264 encoding flowchart

The H.264 encoding process is illustrated in Figure 2. The RD-Cost is computed with the following formula:

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

where s and c are the source video signal and the reconstructed video signal, respectively, QP is the quantization parameter, λ_{MODE} is the Lagrange multiplier, MODE indicates a macroblock mode, like P16x16, P16x8, P8x16, P8x8, I16x16, I4x4, etc., SSD is the sum of the square differences between s and c, R(s,c,MODE|QP) is the number of bits associated with the chosen macroblock MODE and QP.

In H.264 video coding, the main idea for intra prediction is based on the concept of rate distortion optimization (RDO). RDO means that when starting to encode the intra block, as shown in the above flow chart, the encoder needs to try all the mode combinations and choose one with the minimum RD cost in the mode selection procedure. According to the intra coding structure, the number of the mode combinations can be computed as C8*(L4*16+L16), where C8, L4, and L16 denote the numbers of the chroma 8x8 modes, Luma 4x4 modes, and Luma 16x16 modes, respectively. So we can compute the total number of combinations to be 4*(9*16+4)=592. This will take a lot of time to compute the RD cost, hence we need to reduce the total number of modes in the mode decision process to speed up the computation.

3. IMAGE STSRUCTURE TENSOR

To analyze the local image structure, we first apply the Sobel operator to the image as follows:

$$\begin{aligned} D_x(x,y) &= I(x-1, y+1) + 2I(x, y+1) + I(x+1, y+1) \\ &\quad - I(x-1, y-1) - 2I(x, y-1) - I(x+1, y-1) \\ D_y(x,y) &= I(x+1, y-1) + 2I(x+1, y) + I(x+1, y+1) \\ &\quad - I(x-1, y-1) - 2I(x-1, y) - I(x-1, y+1) \end{aligned} \quad (2)$$

where I(x,y) represents the pixel intensity. Then, we obtain the x-directional gradient D_x and y-directional D_y at each pixel location. For instance, if the block is 4x4, we apply the Sobel operator and obtain the horizontal and vertical gradient components at all the 16 pixels in this 4x4 block.

Secondly, the gradient vectors (D_x, D_y) are used to compute the image structure tensor matrix. For each local block, we can compute a 2x2 matrix H given by

$$H = \sum_{window} (\nabla I)(\nabla I)^T \approx \sum_{window} \begin{bmatrix} D_x^2 & D_x D_y \\ D_x D_y & D_y^2 \end{bmatrix} \quad (3)$$

After computing this 2x2 matrix, we use the eigen decomposition and get two eigenvalues and their associated eigenvectors which are orthogonal to each other. The eigenvector corresponding to the smaller eigenvalue can roughly represent the edge direction of the window. In this way, we can obtain a rough idea of the intra mode depending on the direction determined from the image structure tensor.

From the singular value decomposition, we can formulate the eigen decomposition of the 2x2 image structure tensor matrix H in the following.

$$\begin{aligned} H &= \begin{bmatrix} V_1 & V_2 \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} V_1 & V_2 \end{bmatrix}^T \\ &= \lambda_1 V_1 V_1^T + \lambda_2 V_2 V_2^T \end{aligned} \quad (4)$$

where λ_1 and λ_2 are the singular values with $\lambda_1 \geq \lambda_2 \geq 0$, and V_1 and V_2 are the corresponding eigenvectors. Note that V_1 and V_2 are orthogonal to each other. To approximate the mode from the direction V_2, we can substitute V_1 and V_2 by the mode angle unit vector and its corresponding orthogonal mode angle unit vector. Take Luma 4x4 prediction mode for example, mode 3 is 45° and its corresponding orthogonal mode 4 is 135°. So V_1 is replaced by $\begin{bmatrix} \cos 45^\circ \\ \sin 45^\circ \end{bmatrix}$ and V_2 is replaced by $\begin{bmatrix} \cos 135^\circ \\ \sin 135^\circ \end{bmatrix}$. Therefore, there are four pairs of [V1,V2] for the eight modes except the DC mode for a 4x4 block.

For Chroma 8x8, Luma 16x16 and Luma 4x4, we have two pairs of modes, two pairs of modes and four pairs of modes respectively. Now we have substituted the V_1 and V_2, the left unknowns are λ_1 and λ_2 .We rewrite the formula as following form.

$$\begin{bmatrix} V_{1,1} & V_{2,1} \\ V_{1,2} & V_{2,2} \\ V_{1,3} & V_{2,3} \\ V_{1,4} & V_{2,4} \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} H_{1,1} \\ H_{1,2} \\ H_{1,3} \\ H_{1,4} \end{bmatrix} \quad (5)$$

By using the following notations,

$$A = \begin{bmatrix} V_{1,1} & V_{2,1} \\ V_{1,2} & V_{2,2} \\ V_{1,3} & V_{2,3} \\ V_{1,4} & V_{2,4} \end{bmatrix}, B = \begin{bmatrix} H_{1,1} \\ H_{1,2} \\ H_{1,3} \\ H_{1,4} \end{bmatrix}, \lambda = \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} \quad (6)$$

we can write the least square estimation for solving the equation for solving the vector λ that contains λ_1 and λ_2 as follows:

$$\lambda = (A^T A)^{-1} A^T B \quad (7)$$

Then, we determine the pair of modes that provides the best approximation to the matrix H to be the selected mode.

$$\mathrm{mod}\, e = \arg\min_{i \in M} \| B - A_i \lambda \| \quad (8)$$

Note that M denotes the set of modes and A_i is the A matrix with the eigenvectors corresponding to the direction of the *i*-th mode. The mode with the smallest approximation error provides a good estimation for deciding the direction of intra-predicting this block. We take the eigenvector corresponding to the smallest eigenvalue as the rough edge direction of the block. The vector represents one directional mode, which is used to determine the list of candidate modes.

4. PROPOSED MODE DECISION ALGORITHM

The flowchart of the proposed mode decision algorithm is depicted in Figure 3. It provides a candidate mode for the block. Based on the mode distribution analysis, we know the DC mode is the most frequently used mode. Since our structure tensor analysis did not consider the DC mode, we also add DC mode to the candidate mode list for each block. Therefore, we have two candidate modes including the predicted direction mode and DC mode.

We apply our algorithm to Chroma8x8 U and V components, Luma16x16 blocks, and Luma4x4 blocks. Luma4x4 blocks are decomposed from Luma16x16 block. So we just reuse the image gradient computed in the Luma16x16 block for the corresponding Luma4x4 blocks.

For Luma4x4 blocks, according to the analysis in [2], we can find that the neighboring modes to the best mode usually also have small RD costs. Because of this characteristic, we add 2 neighboring mode of the directional modes to the candidate mode list for the block to achieve more accurate mode decision. Thus, there are four candidate modes for the intra 4x4 mode decision.

In Chroma8x8, we use our algorithm on its U and V component individually. If we obtain the same predicted direction mode from both them, we have two candidate modes. One is the predicted direction mode, and the other one is DC mode. On the other hand, if we obtain two different predicted direction modes from U and V, we will have three candidate modes. This idea is referred to [4].

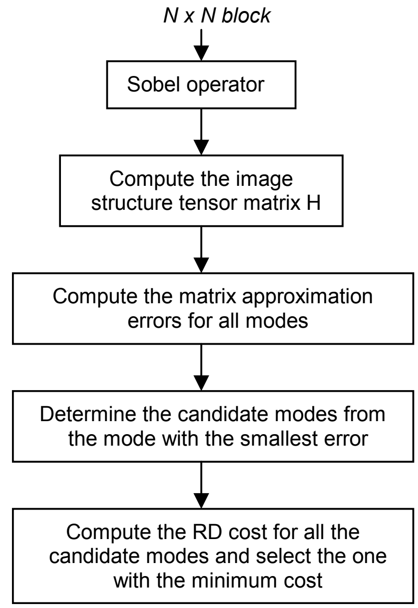

Figure 3. Flow chart of the proposed intra mode decision algorithm

5. EXPEREMENTAL RESULTS

In our implementation, we integrate the proposed intra mode decision function into the JM 10.2 reference program. We compare the performance by using the proposed algorithm with the standard method in the reference program and the algorithm by Pan et al. [4]. The RDO is open and the 8x8 transform is turned off in our program. We implement two versions of our algorithm. In the first version, we define the candidate list to contain the best mode determined by the image structure tensor analysis, its two neighboring modes, the DC mode, and the decided modes in the upper and left neighboring blocks. For the second version, the candidate list only contains the mode determined from the image structure tensor analysis, its two neighboring modes, and the DC mode. In our experiments, we use all I frames in the H.264 encoding. The experimental results by using both versions of the proposed algorithm on different video sequences are summarized in Table 2 and 3.

Table 2. The performance of the first version of the proposed algorithm on different sequences compared to JM 10.2 (QP=28)

Sequence Name	△PSNR(dB)	△BR(%)	△Time(%)
Foreman_300_352x288.yuv	-0.0386	0.64%	-58.60%
Football_125_352x240.yuv	-0.0658	0.00%	-59.37%
Carphone_96_176x144.yuv	-0.0305	0.56%	-58.01%
Coastguard_300_352x288.yuv	-0.0515	0.12%	-61.88%
Container_300_352x288.yuv	-0.0397	0.61%	-61.91%
Hall_monitor_300_352x288.yuv	-0.0266	1.03%	-59.37%
Tennis_112_352x240.yuv	-0.0502	0.14%	-58.30%

Table 3. The performance of the second version of the proposed algorithm on different sequences compared to JM 10.2(QP=28)

Sequence Name	△PSNR(dB)	△BR(%)	△Time(%)
Foreman_300_352x288.yuv	-0.0399	1.77%	-60.88%
Football_125_352x240.yuv	-0.0764	0.37%	-62.24%
Carphone_96_176x144.yuv	-0.0359	1.53%	-60.51%
Coastguard_300_352x288.yuv	-0.0566	1.06%	-63.49%
Container_300_352x288.yuv	-0.0438	1.36%	-63.03%
Hall_monitor_300_352x288.yuv	-0.0266	2.73%	-60.28%
Tennis_112_352x240.yuv	-0.0512	0.67%	-60.14%

We also test both versions of our algorithm compared to the H.264 reference program on different sequence with four different QP values (28, 32, 36, 40). The results are summarized in Table 4 and 5. We also implemented the intra mode decision method by Pan et al. [4] and give the bitrate and PSNR degradation compared to the reference program in Table 6. The RD curves for these different intra mode selection methods on two video sequences are shown in Figure 4 and 5. It is obvious that the RD curves of the proposed algorithm are very close to the exhaustive search in the reference code. In contrast, the intra mode decision method by Pan et al. [4] has significant degradation in the RD curve.

Table 4. The bitrate and PSNR degradation for the first proposed algorithm compared to JM 10.2 with different QP values.

Sequence Name		**QP=28**	**QP=32**	**QP=36**	**QP=40**
Carphone_96_ 176x144	△BR	0.55%	0.70%	1.01%	1.21%
	△PSNR	-0.031	-0.016	-0.022	-0.028
Foreman_300_ 352x288	△BR	0.64%	0.82%	0.93%	1.19%
	△PSNR	-0.040	-0.037	-0.024	-0.02
Container_300 _352x288	△BR	0.55%	0.70%	1.01%	1.21%
	△PSNR	-0.031	-0.016	-0.022	-0.028

Table 5. The bitrate and PSNR degradation for the second proposed algorithm compared to JM 10.2 with different QP values.

Sequence		**QP=28**	**QP=32**	**QP=36**	**QP=40**
Carphone_96_ 176x144	△BR	1.53%	2.26%	3.55%	4.80%
	△PSNR	-0.036	-0.012	-0.005	-0.0013
Foreman_300_ 352x288	△BR	1.77%	2.40%	2.90%	3.13%
	△PSNR	-0.039	-0.029	-0.017	-0.003
Container_300 _352x288	△BR	1.36%	1.96%	2.91%	3.34%
	△PSNR	-0.044	-0.038	-0.030	-0.003

Table 6. The bitrate and PSNR degradation for the method by Pan et al. [4] compared to JM 10.2 with different QP values.

Sequence Name		**QP=28**	**QP=32**	**QP=36**	**QP=40**
Carphone_96_ 176x144	△BR	14.87%	15.66%	14.57%	11.72%
	△PSNR	-0.0935	-0.1254	-1.139	-0.1254
Foreman_300_ 352x288	△BR	10.90%	12.02%	11.78%	10.57%
	△PSNR	-0.0982	-0.115	-0.1237	-0.1284
Container_300 _352x288	△BR	11.52%	14.09%	15.53%	14.16%
	△PSNR	-0.1093	-0.1087	-0.1454	-0.1338

6. CONCLUSION

This paper presented a fast intra mode decision algorithm based on the image structure tensor analysis. Based on the block image structure analysis, we can significantly reduce the RD-cost computation from 4x(4+16x9) times to 2x(2+16x4) times. Our experimental results show that we can achieve about 60% time saving with negligible bitrate and PSNR degradation. We also show the superior video compression performance of the proposed algorithm through RD curve comparison with the previous method.

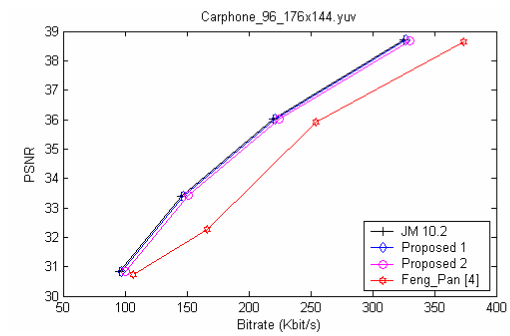

Figure 4. RD curve comparison of different methods on the Carphone sequence

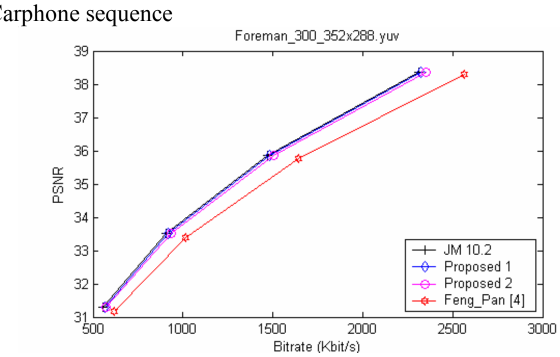

Figure 5. RD curve comparison of different methods on the Foreman sequence

ACKNOWLEDGEMENTS

This research work was supported in part by MOEA project under grant 95-EC-17-A-01-S1-034 and National Science Council, under grant 95-2220-E-007-028.

REFERENCES

[1] ITU-T Rec. H.264/SO/IEC 11496-10, "'Advanced Video Coding", Final Committee Draft, Document IVT F100, December 2002

[2] C.-C. Cheng and T.-S. Chang, "Fast Three Step Intra Prediction Algorithm for 4x4 blocks in H.264," Proc. IEEE ISCAS, Vol. 2, pp. 1509 – 1512, May 2005

[3] Y.-D. Zhang, D. Feng, and S.-X Lin, "Fast 4x4 intra-prediction mode selection for H.264," Proc. IEEE ICME., Vol. 2, pp. 1151–1154, 2004

[4] F. Pan, X. Lin, S. Rahardja, K. P. Lim, Z. G. Li, D. Wu, and S. Wu, "Fast mode decision algorithm for intraprediction in H.264/AVC video coding," IEEE Trans. Circuits Systems for Video Technology, Vol. 15, No. 7, pp. 813-822, July 2005.

[5] JVT Reference Software JM10.2

[6] G. Bjontegaard, "Calculation of average PSNR differences betweenRD-curves," presented at the 13th VCEG-M33 Meeting, Austin, TX, 2001.

[7] F. Pan, F., L.S. Rahardja, K.P. Lim, L.D. Wu, W.S. Wu, C. Zhu, W. Ye, and Z. Liang, "Fast intra mode decision algorithm for H.264-AVC video coding," Proc. IEEE ICIP, Vol. 2, pp. 781-784, Oct. 2004.