FAST MODE DECISION ALGORITHM FOR INTRA PREDICTION IN H.264/AVC*

Rui Su, Guizhong Liu, Tongyu Zhang

School of Electronics and Information Engineering, Xi'an Jiaotong University, 710049, China

ABSTRACT

In H.264/AVC, a novel criterion named the rate distortion optimization (RDO) is employed to select the optimal coding modes for each macroblock (MB) within the intra prediction coding pictures, which can achieve a high compression ratio while leading to a great increase in the complexity and computational load unfortunately. In this paper, a fast mode decision algorithm based on integer transform and adaptive threshold is proposed. Before the intra prediction, integer transform operations on the original image are executed to find the directions of local textures. According to this direction, only a small part of the possible intra prediction modes are tested at the first step. If the summation of differences (SAD) of the reconstructed block corresponding to the best mode is smaller than an adaptive threshold, calculation is terminated. Otherwise, more possible modes are needed. Simulations show that the fast mode decision algorithm proposed in this paper can accelerate the speed of intra picture coding significantly only with a negligible PSNR loss or a bit rate increment.

1. INTRODUCTION

Compared with the other existing video coding standards, H.264/AVC can achieve a significant improvement in compression performances [1]. The common elements used in other existing video coding standards are also adopted by H.264/AVC, so the improvement is mainly brought about by the new techniques, one of which is the intra prediction for intra coded blocks. To acquire the optimal mode in intra prediction, the rate distortion optimization (RDO) technique is adopted for each macroblock (MB) [2]. In order to select the best encoding mode for each MB, video encoder calculates the rate distortion cost (RDcost) for every possible mode. Consequently, the computational load is much more intensive than that of any existing video coding standards.

Recently, several fast algorithms have been developed to cut down the heavy computational burden of RDO calculations of the intra encoding. The optimal intra prediction mode is closely related to the texture direction. By virtue of this feature, Feng Pan et al [3] have proposed a successful fast algorithm. In their scheme, the Sobel edge operator is adopted to calculate the edge direction and amplitude of each pixel. For a 4×4 subblock, the edge direction histogram was acquired and the edge direction of the block is selected. Only 4 modes are chosen to execute RDO. Similarly, for the 16×16 luma and 8×8 chroma block, the possible modes is cut down to 2. This algorithm can shorten the encoding time by about 56% if all the frames are the intra coding modes while keeping a similar PSNR [3]. But in their algorithm, a large number of division and float addition operations are involved. And in [4], I-Ming Pao et al. deduced the relationship between the minimum mean absolute error (MMAE) and the DCT coefficients.

After the investigation of the above papers, a novel fast intra prediction algorithm based on integer transform and adaptive threshold is presented in this paper. Our method does the edge direction calculation for a whole block. In this way, the large number of Sobel edge detection operations for a 4×4 block is replaced by executing a simpler integer transform only once. Hence, a large part of the massive division (in Sobel operator) and float addition operations (in acquiring the edge direction histogram) is also avoided. Compared to Sobel, accuracy of the integer transform has some degradation. Hence, more possible modes should be tested and an adaptive threshold is employed to decide whether the extra RDO are needed. So the distribution of the residual is observed, the data prove that a Laplacian distribution can also serve well. The relationship between the MMAE and the integer transform coefficients is established and the adaptive threshold is also developed in this paper. The simulation results show that our fast algorithm can increase the encoding speed of intra picture coding significantly (up to 50% in comparison with the speed of the default algorithm in Jm6.1 if all frames are employed in the intra coding modes) with a negligible PSNR loss or an insignificant bit rate increment.

* This work is partially supported by China National Natural Science Foundation (CNSF, Project No.60272072), Ministry of Education of China Trans-Century Elitists Project (TCEP, 2002), and Ministry of Education of China 'the Tenth Five Years Plan'-'211' project at Xi'an Jiaotong University.

The organization of the rest paper is as follows. Section 2 will introduce RDO calculation for intra-mode decision in H.264/AVC. The two techniques (determination of texture direction with integer transform, the adaptive threshold for 4×4 blocks) will be discussed in Section 3. Then our proposed fast algorithm will be illustrated in Section 4. The simulation and comparisons will be presented in Section 5. And finally some conclusions will be drawn in Section 6.

2. RDO CALCULATION FOR MODE DECISION

For the luma component of a given MB, its predictor may be composed of sixteen individual 4×4 blocks (denoted as I4MB) or a whole 16×16 MB (denoted as I16MB). When the I4MB is adopted, there are totally nine possible prediction modes available. If the I16MB is selected, there are four possible modes available. Compared with the I4MB modes, I16MB modes are suitable to those smooth image areas [3]. Similarly, the chroma components have four possible modes like the luma ones in I16MB.

The RDO technique is applied to improve the compression performance when the best mode is decided among the possible modes. RDcost is computed according to the Equation (1).

$$RD\text{cost}=\text{D}+\lambda\text{R} \tag{1}$$

where D is the distortion value, generally represented by the sum of the squared differences (SSD), R is the real bits for encoding the residual of the MB, calculated by the entropy coding module, and λ is the Lagrangian multiplier, which is specified by Equation (2), with Qp being the quantization parameter.

$$\lambda = 0.85 \times 2^{(Qp-12)/3} \tag{2}$$

In order to acquire the RDcost value of a possible mode, the residual block should be encoded once. Therefore, the MB need to be encoded many times during the mode decision. According to [3], the default number of RDO calculation is 4 × (9 × 16 + 4) = 592, which means that a MB would be encoded 592 times before a best mode is selected.

3. DETECTING TEXTURE DIRECTION

To avoid too many divisions and float addition operations, the integer transform is used to replace the Sobel edge detection operator in our algorithm. The rest of this section will illustrate the principle and procedure of this technique.

3.1. Acquire the Direction with Integer Transform

In H.264/AVC, 4×4 integer transform is deduced form the DCT. Hence, we can derive the formulations from the real DCT, its reverse equation is as follows:

$$f(i,j)=\sum_{i=0}^{3}\sum_{j=0}^{3}C(u)C(v)F(u,v)\cos[(2i+1)u\pi/8]\cos[(2j+1)v\pi/8] \tag{3}$$

where f and F are original pixel and transform coefficients respectively. It is obvious that some direction information of local texture is also hidden in the transformed coefficients. Therefore, the texture direction can also be estimated by these coefficients. Equation (4) is used to estimate the tangent value η of the angle θ between the local direction and horizontal axis. This Equation can be derived from Equation (3) with the partial derivative and the property of the DCT within several steps. In view of the limits of the paper, the deductions are omits here. The the tangent value η of a 4×4 block can be estimated by:

$$\eta=\frac{\tilde{F}(0,1)+\tilde{F}(0,2)+\tilde{F}(0,3)}{\tilde{F}(1,0)+\tilde{F}(2,0)+\tilde{F}(3,0)}, \tag{4}$$

where $\tilde{F}=CfC^T$ is the 2D 'core' transform, in which only the integer addition and shift operations are involved.

3.2. Establishment of Modes Assembly

Among the I4MB modes, eight modes reflect the different texture directions of the block. So they can be identified by η by the Equation (4) in literature [3]. After acquiring the local texture direction, the same way is used to establish the modes assembly for a 4×4 subblock. The assembly includs the predicting mode specified by the local texture direction, its two neighboring modes and DC mode.

To the assembly of I16MB modes, the downsampling is need prior to the integer transform. At first, a pixel is extracted from each 4×4 subblock, the new 4×4 block is compose of these extracted pixels. Now the same integer transform is used to decide the local texture direction of the whole MB. With the Equation (5) in literature [3], the mode specified by texture direction can be selected. Finally, this mode and DC mode consist of the assembly.

The two chroma block belonging to one MB should be encoded in the same mode. To make use of the integer transform described above, a 4:1 downsampling (abstracting one sample from four samples) is done to each of the chroma blocks. After the sizes of the chroma blocks are reduced to 4×4, the two values η_1 and η_2 (for the U and V components respectively) are computed according to the Equation (4). Then η is defined by Equation (5):

$$\eta=(\eta_1+\eta_2)/2 \tag{5}$$

And then the primary mode specified by η can be obtained. Note for 8×8 chroma blocks, the similar equation (Equation (5) in literature [3]) is applied, except that the order of mode numbers is different. Finally, this mode and DC mode consist of the assembly.

3.3. Adaptive Threshold for I4MB Luma Blocks

The distribution of the residual values after inter-frame prediction (ME and MC) can be modeled by a zero-mean Laplacian distribution [4]. Several video sequences are encoded in the all I frames structure to investigate the distribution of the residual values after I4MB intra prediction. The data suggest that the distribution could also be modeled by a zero-mean Laplacian distribution. The correlation among the grey level values of the residual pixels after I4MB intra prediction is also investigated. The experiments show that the residual values after I4MB intra prediction may be characterized by a zero-mean Laplacian distribution with a separable covariance, and the correlation parameter $\rho = 0.7$ here.

It is noted that when AC coefficients of the residual are all zero, the corresponding prediction mode can be thought the optimal because the intra-prediction aims to eliminate the influence of the texture. Therefore, when Equation (6) is satisfied, the optimal mode is already found:

$$QP_{step} > 3\sigma_F(0,1) \tag{6}$$

where QP_{step} is the quantization step size in H.264/AVC, σ_F is the matrix which is composed of the variances of the integer transform coefficients. With the same deduction in [4], an adaptive threshold is derived:

$$TH = 2.6536 \times QP_{step} \tag{7}$$

Thus, if the summation of the differences (SAD) value is smaller than TH, the mode decision can be stopped.

4. FAST MODE DECISION ALGORITHM

The flowchart of mode decision for a MB is shown in Fig. 1. At first, the primary directions of the chroma blocks are detected and the possible assembly containing two modes is established. Secondly, for each of the 4×4 luma blocks (total number is sixteen for a MB), its primary direction is determined and an assembly containing four I4MB modes is set up. For all these chroma and I4MB modes, RDO calculations are carried out to find the best. Calculate the SAD of the luma block and compare it with the threshold TH. If SAD is large, more extra I4MB modes should be carried out the RDO calculation until SAD is smaller than TH or all the I4MB modes are tested. Otherwise, this I4MB mode is regarded acceptable and the RDO calculations are terminated. This procedure is repeated for all the sixteen luma blocks and $RDcost_{4X4}$ (summation of RDcosts of the sixteen luma blocks) is recorded. In the next step, the best mode of the I16MB modes assembly is found by RDO method and the minimum $RDcost_{16X16}$ is recorded. Finally, the value $RDcost_{4X4}$ is compared with $RDcost_{16X16}$ and the mode with the smaller value is selected as the optimal mode.

5. SIMULATING RESULTS

Feng Pan et al's algorithm [3] and our proposed algorithm are implemented into the software JM6.1 [5] provided by JVT. Totally 12 video sequences are used to acquire the performances of the different algorithms. The

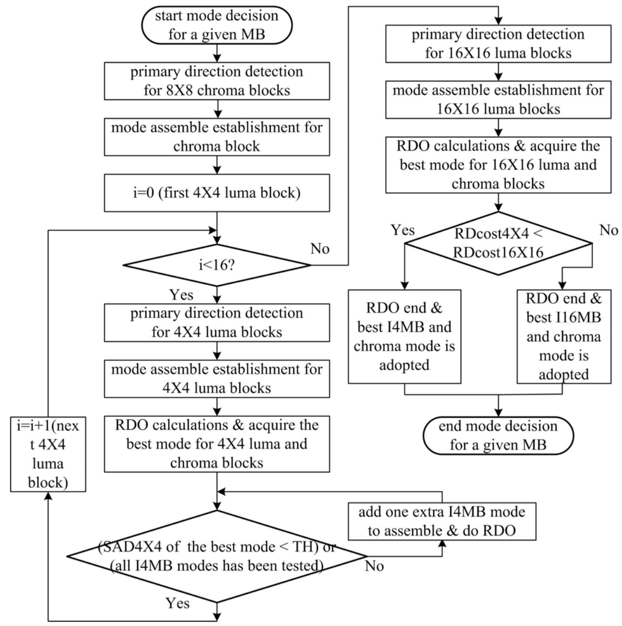

Fig. 1. Flowchart of fast mode decision algorithm for intra prediction in H.264/AVC.

quantization parameters (QP) are set to 28, 32, 36, and 40, as specified in [6]. The results (both Feng Pan et al's and our algorithms) are tabulated in terms of the difference of coding time ($\Delta Time$), the average PSNR difference ($\Delta\overline{PSNR}$) and the bit-rate difference ($\Delta Bitrate$). The values of three items are acquired in the same way as that of literature [3]. The coding time data is generated from JM6.1 software encoder. The platform is Pentium IV 2.8GHz with 1Gbytes RAM.The data of performances on PSNR and Bitrate are calculated according to the numerical average between the fitting rate-distortion curves produced by the fast algorithms and JM6.1 default algorithm. The curve-fitting method and the detail procedures in calculating the differences on performance of PSNR and Bitrate are described in a JVT document [7]. Hence, the PSNR and Bitrate differences should be regarded as equivalent or alternative, i.e., either the degradation in PSNR or the increase in Bitrate come forth (not both at the same time) [3].

Encoding a video sequence in the all Intra frames structure is necessary to some applications. Therefore, the fast algorithms for the all intra frames structure encoding are firstly tested. In this experiment, there are 150 frames for each video sequence, and the period of I-frames is set to 1, i.e., all the frames are intra encoded. Table I lists the simulation results for the two algorithms, one is Feng Pan et al's labeled by 'SOBEL' and the other is our method labeled by 'DCT'. Note that, in the table, positive values represent increments and negative values mean decrements. From these data, it is observed that our algorithm achieves a consistent timesaving (in average 50%), which means that

TABLE I RESULTS FOR IIIII STRCTURE

Format	Sequence	ΔTime(%)		ΔPSNR(dB)		ΔBitrate(%)	
		SOBEL	DCT	SOBEL	DCT	SOBEL	DCT
CIF	Foreman	-57.3712	-51.9941	-0.2975	-0.2343	5.2163	3.9352
	Coastguard	-56.8120	-46.1231	-0.2271	-0.1557	3.4930	2.4551
	Mother	-56.4348	-54.1071	-0.4486	-0.4117	8.5137	7.5593
	News	-57.7041	-52.7792	-0.4507	-0.4793	5.9716	6.2256
	Hall	-56.6255	-52.3307	-0.4191	-0.3954	6.0167	5.6184
	Flower	-58.3835	-48.3996	-0.3358	-0.1022	3.4611	1.0190
QCIF	Carphone	-55.2367	-48.7847	-0.3377	-0.3013	4.5180	3.8867
	Coastguard	-54.8700	-42.1725	-0.1744	-0.0950	2.6337	1.3508
	Container	-56.1531	-47.3424	-0.2612	-0.2085	3.8038	2.8490
	Foreman	-55.7099	-47.0781	-0.2486	-0.1907	3.7381	2.6848
	Funfair	-56.5835	-41.4179	-0.2338	-0.0834	2.8944	0.9485
	Hall	-54.4948	-47.4551	-0.3215	-0.2018	3.9540	2.4415

our algorithm only takes about half of the time that is needed by the default algorithm in JM6.1. The average loss of PSNR is about 0.2383dB and the average increment in bitrate is about 3.4145%. Though Feng Pan et al's algorithm has a higher timesaving (in average 56%), the average degradation in PSNR and the average increment in bitrate are also more than ours, about 0.3130dB and 4.5179% respectively.

6. CONCLUSION

A fast mode decision algorithm for intra prediction in H.264/AVC video coding is proposed in this paper. In order to decrease the computational load during the intracoding procedure, two techniques are adopted in our fast algorithm. One is the integer transform upon the original picture blocks, by which the texture directions of blocks can be decided with less computation (especially for division and float addition operations) than that of Feng Pan et al's algorithm. The other is the adaptive threshold in the mode decisions for the 4×4 luma blocks, by which the accuracy is improved. Simulation results show that the proposed algorithm has a negligible loss of PSNR or an insignificant increment of bitrate in comparison with the original algorithm while achieving a considerable timesaving. In addition, in view of the capability of cutting down the division and float addition operations significantly, our algorithm should be more suitable than Feng Pan et al's to be adopted in some applications (such as the embedded systems) in which CPUs are sensitive to the cost of multiplication/division and float addition operations.

7. REFERENCES

[1] G. J. Sullivan and T. Wiegnad, "Video Compression—From Concepts to the H.264/AVC Standard", *Proceedings of IEEE*, Vol. 93, pp. 18-31, Jan. 2005.

[2] G. Sullivan, T. Wiegand, and K. P. Lim, "Joint model reference encoding methods and decoding concealment methods, " presented at the 9th JVT Meeting (JVT-I049d0), San Diego, CA, Sep. 2003.

[3] F. Pan, X. Lin, S. Rahardja, K. P. Lim, Z. G. Li, D. J. Wu, and S. Wu, "Fast mode decision algorithm for intraprediction in H.264/AVC video coding," *IEEE Tans. Circuits Syst. Video Technol.*, vol. 15, no. 7, pp. 813-822, July. 2005.

[4] I-Ming Pao, Ming-Ting Sun, "Modeling DCT coefficients for fast video encoding", *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 4, pp. 608–616, Jul. 1999.

[5] JM6.1, Reference Software of JVT. http://bs.hhi.de/~suehring/tml/download/ jm61d.zip.

[6] JVT Test Model Ad Hoc Group, "Evalution sheet for motion estimation," Draft version 4, Feb. 19, 2003.

[7] G. Bjontegarrd, "Calculation of Average PSNR Differences between RD-curves", presented at 13th VCEG-M33 Meeting, Austin, TX, Apr. 2001.