Motion Estimation and DCT Coding Combined Scheme for H.264/AVC Codec

Wei-Jhe Hsu [1] , Hsueh-Ming Hang [1] , and Yi-Fu Chen [2]

[1] Department of Electronics Engineering, National Chiao-Tung University, Hsinchu, Taiwan
hsu761001@gmail.com, hmhang@mail.nctu.edu.tw
[2] Telecommunication Laboratories, Chunghwa Telecom Co., Ltd., Taoyuan, Taiwan

Abstract. A typical H.264 video encoder (such as JM) selects the best motion vector based on the sum of absolute difference (SAD) and the sum of absolute transformed difference (SATD) in different accuracy layers. In this paper, we propose a jointly optimal approach that selects the best motion vector that minimizes the rate-distortion cost of the quantized transform coefficients. We test the proposed scheme on a number of sequences. The results indicate that our scheme provides a bit-rate gain up to 4% for P pictures.

Keywords: Motion Estimation (ME), H.264/AVC, Rate-Distortion (R-D) Optimization, Hadamard Transform.

1 Introduction

The H.264/AVC video coding standard [1] provides a rather high coding efficiency. A typical H.264 video coder contains a motion vector (MV) selection module and a coding mode selection module. Typically, such as the standard committee reference software JM 18.0 [2], picks up the best motion vectors and the best coding modes in two separate steps. In this paper, we focus on picking the best motion vectors for the best final coding results and thus they improve the overall rate-distortion (R-D) performance. The remaining sections of this paper are organized as follows. Section 2 introduces the typical R-D optimization in the inter procedure for H.264/AVC and the related work. Our proposed algorithm is described in Section 3. Section 4 presents the experimental results and the discussion. Section 5 concludes our work.

2 Motion Estimation for H.264/AVC and Related Work

The encoder of H.264/AVC uses the transform coding technique to encode the motion-compensated prediction errors. A residual block is produced by subtracting the prediction from the current block. Then, the residual block is transformed by the 4x4 separable integer DCT (IDCT) [3] or the 4x4 Hadamard transform (H matrix) as shown in (1) which is an approximation form of IDCT for low complexity. For more details of encoding procedure, please refer to [1], [4].

adfa, p. 1, 2011.

2.1 R-D Optimization in Inter Procedure

We use the reference software JM 18.0 as the platform, which is widely recognized as one of the best H.264 encoder from the coding performance viewpoint. The general R-D cost function for video coding is expressed by (2), the so-called Lagrange cost function. In (2), symbol *D* denotes the distortion, which is often the absolute difference between the processed image block and the original block. Symbol *R* denotes the rate, which is the bit rate needed to send the processed information. How to select the optimal Lagrange multiplier λ is a difficult problem in practice; often, an empirical formula is in use, as described in [5], [6].

$$H = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix} \tag{1}$$

$$J = D + \lambda R \tag{2}$$

A traditional H.264/AVC encoder splits the cost function optimization process into two steps, and the 2-step process is illustrated by Fig. 1. We describe below the R-D optimization scheme in the original setting of JM18.0 [2]. In the first step, we find the MVs with the least residual distortion and the MV coding bits. Based on the motion R-D cost function in (3), the motion estimation step finds the vector with the least cost for various block sizes. Given the current and the reference frames and the Lagrange multiplier λ_{motion}, ME search looks for the best MV for each partition block s_i to minimize (3).

$$J_{motion} = D(s_i, m) + \lambda_{motion} R_{motion}(s_i, m) \tag{3}$$

where m is the set of all possible vectors. In (3), R_{motion} is the number of bits for transmitting MV, and $D(s_i,m)$ is the term of SAD given by

$$SAD = \sum_{x,y} \left| Dblock(x, y) \right| \tag{4}$$

The symbols in (4), x and y, are the pixel location in a block, and *Dblock* is the difference between the referenced candidate block and the original block. It should be noted that the best sub-pixel MV (half and quarter accuracy) is decided according to the Hadamard transform consideration in $D(s_i,m)$ in (3); that is, the term, $D(s_i,m)$, is SATD defined by (5), where H is the Hadamard matrix in (1).

$$SATD = \sum_{x,y} \frac{1}{2} \left| H * Dblock * H \right| \tag{5}$$

In the second step of the inter mode encoding process, the encoder also applies the Hadamard transform to the motion-compensated residual signals of each inter mode, and then we choose the best MB coding mode by minimize cost function (2).

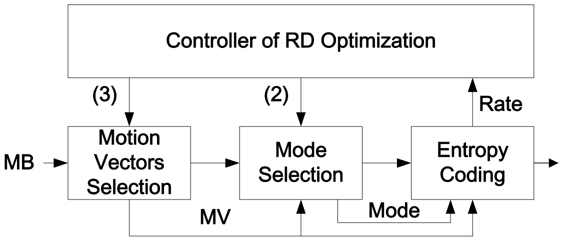

Fig. 1. R-D optimization for selecting MV and mode in JM.

2.2 Related Work

In [7], the effect of SATD on ME at different layers is discussed and tested. The encoder uses SATD for searching for integer MVs, and averagely achieves 1.85% bit rate saving but with 781% encoding time increase when the sub-pixel motion search is disabled. However, the same method leads a small amount of coding loss, about 0.39% BD rate [8], when the sub-pixel MV is enabled. The reason is that SATD tries to match frequency-domain patches rather than the pixel-domain patches. The interpolated pixels at sub-pixel accuracy seem to have negative effect. The report in [7] is interesting but there are 2 undesirable points. First, $(2 \times \text{search range} + 1)^2$ searching points with SATD require high complexity. Second, the experiment indicates that SAD is a better criterion in finding MV at integer pixel level. Therefore, our algorithm is designed to improve the above problems.

3 Combined Motion Estimation and DCT Algorithm

In this section, we describe the principle behind the proposed combined ME and DCT algorithm and its implementation step by step. In the traditional H.264/AVC encoder, the ME procedure chooses the integer pixel vector that minimizes (3) with SAD criterion. However, (3) does not truly reflect the final distortion and the bit rate of the encoded block. Therefore, we include (2) into the ME procedure in selecting MVs to improve overall coding performance. That is, we combine (2) and (3) in the integer ME procedure.

The motivation is as follows. Although a selected MV is not the best candidate in the MV decision at the integer pixel level, its residual DCT may have fewer large transform coefficients and thus produces fewer bits in entropy coding in the final stage. Figs. 2–4 show an image example. Fig. 2 shows the difference between the JM-coded frame and the coded frame using the proposed method. In Figs. 3–4, we compare the residual MBs produced by two MVs on the second frame of the FOREMAN sequence. The comparison is done in both the spatial domain and the frequency domain. Our proposed algorithm chooses a different MV in the final stage (called Motion RDcost#2 means the 2nd best MV in the integer ME step). The resultant residual block has a more clustered frequency domain distribution; that is, the large magnitude coefficients are fewer and are close to each other as shown in Fig. 4 (right). Therefore, these coefficients are easier to compress.

The core concept of the combined ME and DCT algorithm is illustrated by Fig. 5. In the integer-pixel level of ME procedure, our proposed method chooses 5 top candidate MVs with the integer-pixel accuracy based on SAD, and then it finds their corresponding half and quarter-pixel MVs using the Hadamard SAD. At the end, we use the modified function from the mode decision function to calculate the distortion based on the Hadamard transform again and estimate the bit rate. Therefore, in addition to the integer MV search, we increase the sub-pixel MV searches (SATD) for about 5 × [2 × (sub search points) +1] times. After our proposed scheme, we get the best integer vector of each partitioned block, and then use it to the following steps as the original JM, such as sub-pixel ME and mode decision.

Fig. 2. The left decoded picture is coded by our proposed algorithm, and the right residual picture is the difference between the left picture and the coded picture by JM at QP = 22. The differences are 10 times magnified.

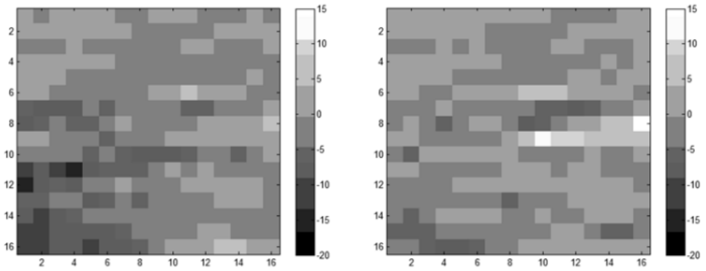

Fig. 3. Spatial domain: The residual MBs of Inter-16x16 mode on the second frame. The MB location (upper-left corner) is (80,160). Gray values are adjusted to show a range from 15 to -20 (the maximum and minimum pixel values). (left) The residual block produced by the MV with Motion RDcost#1. (right) The residual block produced by the MV with Motion RDcost#2.

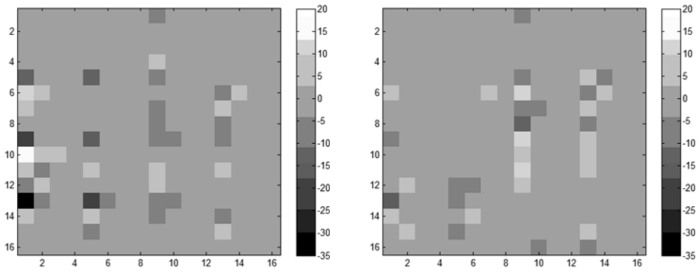

Fig. 4. Frequency domain: The transformed and quantized residual MBs of Fig. 3. Coefficients are produced by 4x4 integer DCT with QP 22. Gray values are adjusted to show a range from 20 to -35. (left) A residual transform block produced by the MV with Motion RDcost#1. (right) A residual transform block produced by the MV with Motion RDcost#2.

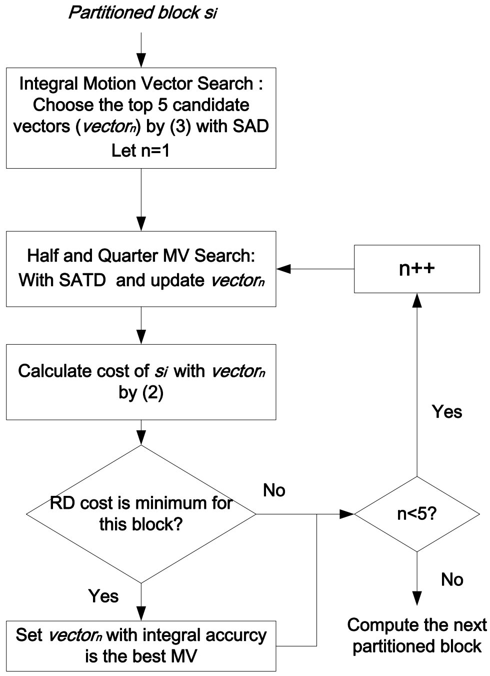

Fig. 5. For each sub-block, select 5 candidate MVs. Compute and compare their R-D costs to decide the best integer MV.

4 Simulation Results

To verify the effectiveness of our proposed motion estimation and DCT combined algorithm, we implement it on the software JM 18.0 [2], which is the reference software of the H.264/AVC encoder. We compare its performance with that of the original JM encoder. The experimental conditions are listed in Table 1, and the test sequences are from [9] (MPEG test video).

Table 1. Experiment conditions

Profile: Baseline
Used QP values : 22, 27, 32, and 37
Encoded Frames : 32
Sequence type : IPPP
Intra Period : 16
Search mode: Fast Full Search
Search range: ±32
Reference frame: The previous frame
Entropy Coding: CAVLC
RD-Optimization: High complexity
ME-Distortion-FPel: SAD
ME-Distortion-HPel: Hadamard SAD
ME-Distortion-QPel: Hadamard SAD
MD-Distortion: Hadamard SAD

Table 2 shows the PSNR and rate comparison at different QP for the FOREMAN sequence and Fig. 6 shows their RD curve at different QPs. We find that the curve has a larger gain in the high rate region because the 8x8 modes are used more often. In this case, because more MVs may be altered and because different MVs may result in different quantized residuals when QP is small, our coding gain becomes more obvious. This phenomenon happens also in the other sequences.

Table 2. R-D Comparison for FOREMAN in P slices

FOREMAN	JM18.0		Proposed method		Y BD rate
	Y_PSNR (dB)	Bitrate (kbps)	Y_PSNR (dB)	Bitrate (kbps)	
QP=22	41.078	1121.89	41.115	1091.63	-3.4%
QP=27	37.648	423.31	37.679	409.61	
QP=32	34.651	183.02	34.668	179.77	
QP=37	31.911	97.47	31.924	94.57	

Table 3 shows the BD rate [8] gain for all sequences. There are two sequences, MOTHER_DAUGHTER and SILENT, which have smaller gains at about 1% because these two videos have very little motion and thus the encoder frequently chooses the skip modes. Our MV selection scheme is applied only to the motion-compensated blocks, whose number is now small. Another factor affects the performance is image contents (patterns). In some sequences, such as CITY and MOBILE, our method provides more gain because they contain a number of fine edges, and thus our method has more chances to manipulate the residual distribution patterns. In summary, two factors seem to have major impact on our algorithm performance. One is the percentage of motion-compensated modes in P-slices, and the other one is the texture pattern of the residual blocks.

We collect the final MV choices in our method in Table 4. It shows that the best motion R-D cost vector is chosen with higher probability when QP is large. In this case, because the number of transform coefficients is small, it thus makes little difference on the residual blocks produced by different MVs. On the average, the probability of choosing the fifth candidate MV is less than 5%. Thus, retaining more than 5 candidate MVs does not seem to offer much improvement.

Finally, we like to know how many "different" MVs at the integer-pixel level are chosen at the end using this approach (versus JM 18.0). We examine both the numbers of sub-blocks and their area. Table 5 shows the sub-block numbers and the area ratio of the changed MVs due to the adaptation of our algorithm.

Table 3. BD rate improvement in P slice of all sequences

Test sequences		Y BD rate	Encoding Time
CIF	FOREMAN	-3.4%	+43.2%
	BUS	-2.6%	+46.6%
	FOOTBALL	-1.9%	+49.6%
	MOBILE	-2.4%	+48.9%
	NEWS	-2.7%	+43.0%
	ICE	-4.2%	+39.8%
	PARIS	-1.6%	+45.3%
	MOTHER_DAUGHTER	-1.3%	+41.2%
	SILENT	-1.1%	+43.2%
4CIF	HARBOUR	-2.2%	+47.0%
	CITY	-2.9%	+45.9%
	SOCCER	-1.8%	+46.1%
	CREW	-1.7%	+45.2%
Average		-2.3%	+45.0%

Table 4. Final MV selected from candidate MVs and percentages

FOREMAN	**QP=22**	**QP=27**	**QP=32**	**QP=37**
Motion RDcost1	53.4%	56.7%	61.4%	67.3%
Motion RDcost2	21.8%	21.3%	19.9%	17.4%
Motion RDcost3	11.8%	10.5%	9.1%	7.5%
Motion RDcost4	7.6%	6.7%	5.7%	4.5%
Motion RDcost5	5.5%	4.7%	3.9%	3.3%
ICE	**QP=22**	**QP=27**	**QP=32**	**QP=37**
Motion RDcost1	70.3%	73.9%	75.7%	77.2%
Motion RDcost2	14.2%	13.1%	12.4%	11.8%
Motion RDcost3	7.2%	6.1%	5.6%	5.3%
Motion RDcost4	4.7%	4.0%	3.6%	3.3%
Motion RDcost5	3.6%	3.0%	2.8%	2.5%
SILENT	**QP=22**	**QP=27**	**QP=32**	**QP=37**
Motion RDcost1	83.4%	84.0%	85.8%	89.1%
Motion RDcost2	7.8%	7.8%	7.0%	5.6%

Motion RDcost3	4.2%	4.0%	3.4%	2.5%
Motion RDcost4	2.7%	2.5%	2.2%	1.6%
Motion RDcost5	2.0%	1.7%	1.5%	1.2%

Table 5. Changed MV partitioned sub-blocks and the area ratio used changed vector

FOREMAN	**Changed MV Blocks**	**Partitioned Blocks**	**Changed Area Ratio**
QP=22	16223	36196	35.38%
QP=27	9739	23969	31.60%
QP=32	5780	17047	26.52%
QP=37	3559	14158	20.05%
ICE	**Changed MV Blocks**	**Partitioned Blocks**	**Changed Area Ratio**
QP=22	11279	28617	23.91%
QP=27	9182	24598	21.35%
QP=32	6936	20481	20.23%
QP=37	4787	16719	18.53%
SILENT	**Changed MV Blocks**	**Partitioned Blocks**	**Changed Area Ratio**
QP=22	4534	21115	9.27%
QP=27	3016	16681	8.93%
QP=32	1932	14066	7.92%
QP=37	1214	12743	6.62%

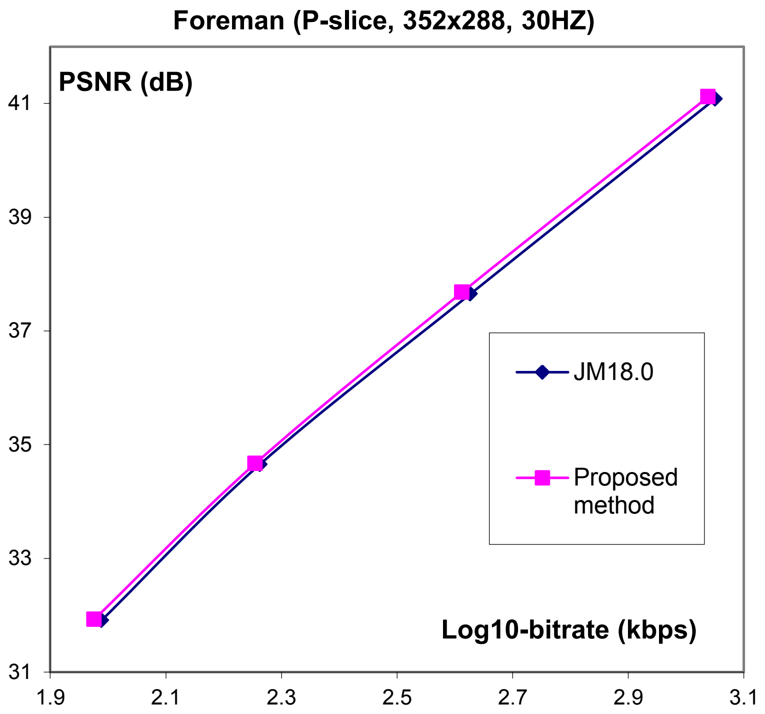

Fig. 6. RD curve of Foreman for P slice.

5 Conclusion

In this paper, we propose a 2-pass motion estimation method to enhance the coding R-D performance by combining motion estimation and DCT for the H.264/AVC encoders. The algorithm considers the transform coding effect in choosing the best motion vectors from integer to quarter pixel accuracy. Based on the multiple sequences tests, we demonstrate that the proposed algorithm can achieve 2.3% average bit-saving without changing the syntax of the AVC/H.264 standard. There is a trade-off between coding efficiency and computational complexity. Although we reduce the SATD operations significantly comparing to [7], the encoding time is still increased by about 45%. Acceleration of our scheme is one of possible future work items.

Acknowledgments. This work was supported in part by the Chung-Hwa Telecomm, Taiwan under Grants TL-100-G113 and by the NSC, Taiwan under Grants 98-2221-E-009 -076.

References

1. T. Wiegand et al. "Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification (ITU-T Rec. H.264 | ISO/IEC 14496-10 AVC)," ISO/IEC JTC/SC29/WG11 and ITU-T SG16 Q.6, JVT-Go50r1, Mar.2003.
2. H.264/AVC Codec [Online] http://iphome.hhi.de/suehring/tml/download/.
3. H. S. Malvar, et al., "Low complexity transform and quantization in H.264/AVC," *IEEE Trans. Circuit Syst. Video Technol.*, vol. 13, no. 7, pp. 598–603, Jul. 2003.
4. I. E. G. Richardson, H.264 and MPEG-4 Video Compression: Video Coding for Next-Generation Multimedia, Wiley.
5. T. Wiegand, et. al., "Rate-constrained coder control and comparison of video coding standards," *IEEE Trans. Circuit Syst. Video Technol.*, vol. 13, no. 7, pp. 688–703, Jul. 2003.
6. T.Wiegand and B. Girod, "Lagrange multiplier selection in hybrid video coder control," *in Proc. Int. Conf. Image Proc.*, Oct. 2001, pp542–545.
7. A. Abdelazim, et al. "Effect of the Hadamard transform on motion estimation of different layers in video coding," *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXVIII, Part5 Commission V Symposium, Newcastle upon Tyne, UK. 2010.
8. G. Bjontegaard, "Calculation of Average PSNR Differences between RD-curves," Document VCEG-M33, Apr.2001.
9. Test sequences [Online] http://media.xiph.org/video/derf/.