FAST INTER MODE DECISION USING RESIDUAL HOMOGENEITY IN H.264/AVC

Zhiru Shi, W. A. C. Fernando and A.M. Kondoz

I-Lab multimedia Communications Research, CVSSP, University of Surrey, Guildford, GU2 7XH, UK
E-Mail: {Z.Shi, W.Fernando and A.Kondoz}@surrey.ac.uk

ABSTRACT

The H.264/AVC coding standard has achieved significant coding gain by using flexible block sizes inter prediction. However, the complexity of inter-mode decision is extremely high. In this paper, a fast inter-mode decision algorithm is presented by exploiting residual homogeneity for both macroblock and sub-macroblock levels. In this algorithm, the residual homogeneity indicators calculated in three directions are used to identify the block's homogeneous characteristic. Normalized motion vector activity is also considered to improve the performance. The experimental results illustrate that the proposed algorithm achieves about 49% of computational complexity saving in terms of encoding time, while with negligible quality degradation, compared to H.264/AVC reference software JM.

Index Terms—inter mode decision, residual homogeneous indicator, motion activity, H.264/AVC

1. INTRODUCTION

The H.264/AVC video coding standard improves coding efficiency significantly by employing several advanced features. One of the most important but computational intensive operation is various block sizes for inter-mode prediction, which are allowed to maximize the coding efficiency based on rate-distortion (R-D) Optimization [8]. To achieve the best R-D performance, the exhaustive inter modes decision process calculates R-D cost of each mode in a tree-structured approach. The one with lowest cost is chosen as the optimal mode. The exhaustive mode decision process from large down to small partition-sizes is illustrated in Fig.1. However, the R-D cost is obtained only after a sequence of computational intensive calculations. In the end, only the optimal motion vector belonging to the optimal partition size mode is actually encoded and finally transmitted. The remaining motion vectors searched from different reference frames and the remaining different block size modes are discarded. Although the exhaustive mode checking process normally achieves the best R-D performance, it significantly increases computational complexity and limits the use of H.264/AVC in real-time applications. Therefore, fast mode decision (FMD) algorithms without notable loss of coding efficiency are indispensable for real-time implementations of H.264/AVC.

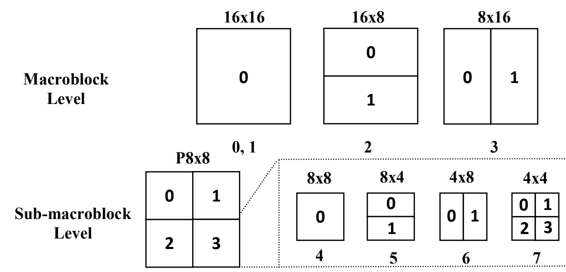


Fig. 1. Macroblock and submacroblock partition sizes in H.264.

In the past few years, many kinds of fast inter-mode decision schemes have been reported. In [1, 2], motion field characteristics are evaluated to reduce the candidate inter prediction modes. However, since both algorithms perform 4x4 block size motion estimation (ME) as the first step, large amount of computational resource are wasted. Liu et al. [3] suggested an efficient early mode decisions by analyzing residual characteristics. Yang et al. [4] proposed an approach by utilizing coded block patterns (CBP) to reduce the complexity of mode decision process. In [5], the CBP based scheme has been improved by using more precise candidate mode classifier in both macroblock and sub-macroblock level. However, the CBP based fast mode decision algorithm would be less effective for the sequences with large amount of irregular motion. Moreover, if the lower Quantization Parameters (QP) values are applied in the encoder, more image details will appear in the decoded pictures. In this case, the CBP that can only signal which sub-block contains zero coefficients is not sensitive enough. To tackle this drawback, a residual homogeneity based fast mode decision algorithm attempting to reduce the influences of QPs is suggested in this paper.

Generally the optimal block size has high correlation with the region's homogeneity characteristic. In H.264/AVC, motion homogeneous MBs are more likely to be divided into larger block sizes, such as 16 x 16, 16 x 8, 8 x 16 and 8 x 8 pixels (denoted as mode 1, 2, 3, 4). These coarse partitions are sufficient to represent the smooth motion and may not result in large residual. While for the regions with complex motion, the uniformed motion vectors of large block divisions are not adequate and may lead to a higher residual energy. Therefore, each 8 x 8 block is further partitioned into 8 x 4, 4 x 8 or 4 x 4 pixel blocks. Although the fine partitions can better represent the irregular motion, the residual energy still remains in higher level compared to homogeneous regions. In this case, the residual information inferring the block's homogeneity characteristic is used to predict candidate modes subset in the proposed algorithm.

2. ANALYSIS ON PREDICTED RESIDUAL VALUES

A similar problem confronted in the CBP based method [5] is that the optimal residual information is not available until the whole mode decision process is completed. It can be observed that 16 x 16 mode has the largest possibility to be chosen as the best mode and it needs smallest computation comparing to other modes. It is desired to use the residual of the16 x 16 block size mode to predict the optimal residual value without investigating all modes. To verify this assumption, experiments and statistical analysis are carried out. The correlation coefficient of the residual values between the 16 x 16 partition size res_{16} and the final residual value res_{final} for best partition size is employed,

$$\rho_k = \frac{cov_k(res_{16}, res_{final})}{\sigma_k(res_{16}) \cdot \sigma_k(res_{final})} \tag{1}$$

where ρ_k is the residual correlation coefficient of the frame k, $cov_k(res_{16}, res_{final})$ is the covariance between res_{16} and res_{final}, $\sigma_k(res_{16})$ is the standard deviation of the 16 x 16 residual values res_{16} and $\sigma_k(res_{final})$ is the standard deviation of the final residual value res_{final}. Similarly at sub-macroblock level, it would be desirable to predict the optimal residual values by 8 x 8 partition size motion compensations.

Experiments were performed to calculate the correlation coefficient for every P frame on CIF sequences, Akiyo and Stefan and HD sequences Cafe and Sunflower. For each sequence, the first 100 frames at four different QPs (24, 28, 32 and 36) are tested. The average values of correlation coefficient are listed in Table 1. It can be seen from the table that the correlation is very strong with an average coefficients ranging from 0.925 to 0.993 at all QP conditions. Noting that for the homogeneous sequences like Akiyo and Café, the average correlations are above 0.95 which is slightly higher than the sequences with complex motion, i.e. Stefan and Sunflower.

Table 1 Average correlation coefficients of macroblock level at different QP values

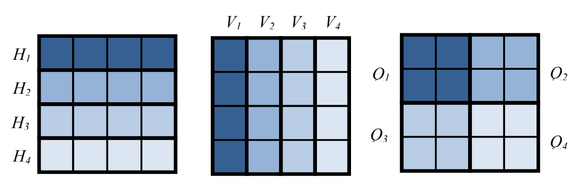
Correlation coefficients		QP 24	QP 28	QP 32	QP 36
Macroblock level	Akiyo	0.955	0.963	0.975	0.974
	Stefan	0.936	0.938	0.925	0.930
	Cafe	0.957	0.974	0.982	0.993
	Sunflower	0.973	0.979	0.969	0.972

Based on the analysis above, it can be observed that correlation is indeed strong between res_{16} and res_{final} in macroblock level. Similar results can be obtained in sub-macroblock levels. This leads to the conclusion that the residual obtained from 16 x 16 or 8 x 8 mode investigation can be used to predict the optimal residual without going through the entire mode decision process. In the next section, a fast mode decision algorithm is proposed based on the predicted residual homogeneity characteristic

3. PROPOSED ALGORITHM

A novel approach of homogeneity detection method mainly based on residual homogeneity (RH) is proposed. In this proposed RH based approach, the same view was shared from the CBP based fast mode decision method that the block's homogeneous characteristic has its direction. Take advantage of this, more computational loads can be reduced. For example, if the residual is vertical homogeneity, then it is less likely to be horizontally divided.

According to the analysis in the previous section, the proposed algorithm extracts residual information by 16x16 pixels ME and generates a 16x16 residual block. Then the basic residual unit (BRU) concept is applied to partition the 16x16 residual block. At macroblock level, BRU size is 4x4 pixels and the mean residual value of each BRU $\overline{r_{i,j}},\ i,j \in (1,2,3,4)$ is calculated.



(a) (b) (c)

Fig. 2. Merged blocks for residual homogeneity measures in Horizontal (a), Vertical (b) and Quartered (c)

According to this observation, every four BRUs are then merged in horizontal, vertical and quartered, as illustrated in Fig.2 Thus the 16x16 residual block is divided into 4 merged blocks $M_k, k \in (1,2,3,4)$, which represent each row block H_k in Fig.2 (a), each column block V_k in Fig.2 (b) or each quartered block Q_k in Fig.2 (c). The residual's mean absolute deviation (MAD) for each merged block M_k is defined as

$$MAD(M_k) = \frac{1}{4}\sum_{M_k}\left|\overline{r_{i,j}} - \bar{r}_k\right| \tag{2}$$

where, $\bar{r}_k$ is the mean residual of all merged blocks. As explained in (2), the Residual Homogeneity Indicator (RHI) in horizontal, vertical and quartered directions are defined as,

$$RHI(H) = \frac{1}{4}\sum_{k=1}^{4} MAD(H_k), \tag{3}$$

$$RHI(V) = \frac{1}{4}\sum_{k=1}^{4} MAD(V_k),$$

$$RHI(Q) = \frac{1}{4}\sum_{k=1}^{4} MAD(Q_k)$$

Based on the RHI*s* listed in (3), the MB's homogeneity characteristics are classified into one of the five classes as defined below.

Class A: The residual is completely homogeneous in each direction

$$RHI(H) < T\ ,\ RHI(V) < T \text{ and } RHI(Q) < T \tag{4}$$

In this case, the MB is detected as homogeneity block and smaller block partition mode can be skipped, so that the complexity can be maximally saved. T is a threshold playing an important role in class detection. The threshold T at MB level is set 0.4 by performing extensive experiments.

Class B: The residual homogeneity indicator is smaller in quartered partition and more than three quartered merged blocks Q_k are completely homogeneous, and is defined by

$$RHI(H) < T\ ,\ RHI(V) < T \text{ and } RHI(Q) < T \tag{5}$$

As no more than one merged 8x8 block contains complex residual, it can be inferred that the MB has high probability to be homogeneous. But it is not enough to make sufficient early mode decision. To avoid image quality degradation, normalised motion activity (NMA) is also employed in the proposed algorithm. Compared to finding the maximum city-block length from neighbour Motion Vectors as in [6], NMA in the proposed algorithm is directly obtained from the ME process performed on current 16x16 block. Hence there is no extra computational burden.

Class C: The residual is more likely to be homogeneous in the horizontal direction. So that vertical partition mode 8x16 can be skipped in further modes investigation.

$$RHI(H) < RHI(V)\ and\, RHI(H) < RHI(Q) \tag{6}$$

Class D: The residual is more likely to be homogeneous in the vertical direction

$$RHI(V) < RHI(H)\ and\, RHI(V) < RHI(Q) \tag{7}$$

Class E: The residual is not homogeneous in any direction. It would be quite difficult to exclude any inter mode from the candidate set.

Combining residual homogeneity and NMA feature, the hybrid fast mode decision algorithm at macroblock (MB) level is illustrated in Fig. 3. In Class B, when NMA is smaller than a threshold, which is set as 1, mode 8 x 16 and mode 16 x 8 are skipped. Otherwise, the exhaustive mode searching will be conducted. Furthermore, if a MB is detected to be homogeneous in terms of residual and motion, the sub-macroblock mode checking will be skipped by changing parameter value "subMB_flag" in Fig.3.

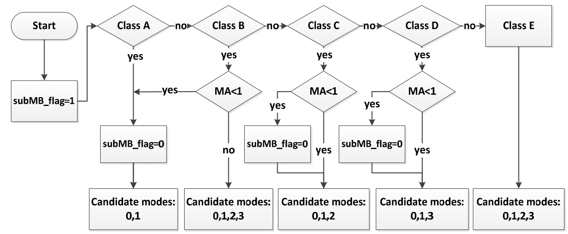


Fig. 3. Determination of candidate modes set at macroblock level in RH based fast mode decision scheme

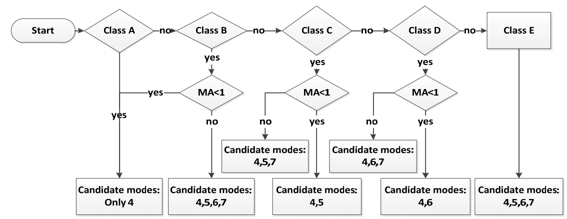


Fig. 4. Determination of candidate modes set at sub-macroblock level in RH based mode decision scheme

Otherwise, the sub-macroblock level modes investigation is performed.

At sub-macroblock level, the basic residual unit size is 2 x 2 pixels. Then, the average residual values for BRUs are updated by applying ME to each 8x8 block. Similar block merging and residual homogeneity detection processes given in (2)-(7) are implemented. The proposed mode decision algorithm at sub-macroblock level is shown in Fig. 4. The threshold T for sub-MB level is set 2.8, which achieves a good performance on different video sequences. Combining the candidate mode determination in macroblock and sub-macroblock levels, the proposed residual homogeneity based fast motion decision algorithm is illustrated by the flowchart in Fig. 5.

4. EXPERIMENT RESULT

The proposed fast mode decision algorithm was implemented into H.264/AVC reference software JM16.1 [7]. To verify the performance of fast mode decision algorithms, comparisons were made against the exhaustive mode decision scheme used in JM and Zhi's [2] motion homogeneity based fast mode decision algorithm in terms of computational complexity (speed measured by total encoding time saving (TS)) and Rate-distortion performance (measured by PSNR and bit rate).

Following the suggestion by JVT, the test conditions are set as follows: motion estimation searching range is ±32 pels; fast motion estimation is UMHexagonS; reference frame number is 5; GOP structure is IPPP; I frame interval is 15; the number of frames to be encoded is 200 for CIF sequence and 100 for the HD sequences; CABAC is used in Main profile. A group of experiments with five QP values (24, 28, 32, 36 and 40) were performed on a PC with 2.44 GHz CPU and 4GB RAM. Ten video sequences were chosen to cover a variety of video contents. These sequences

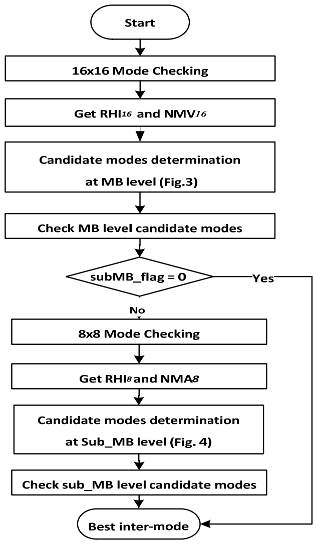


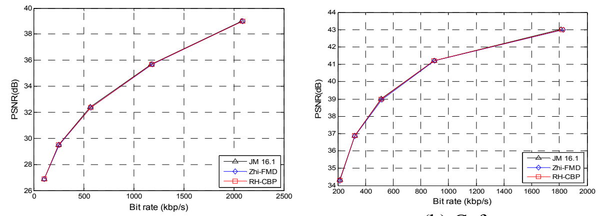
Fig.5. Flow chart of the proposed fast inter-mode decision based on RH algorithm

contain five CIF sequences including *Akiyo* (AK), *Coastguard* (CG), *Football* (FB), *silent (*SL*)* and *Soccer* (SC) and five HD (1920 x 1080) sequences including *Café* (CF), *Rush Hour* (RH), *Hall* (HL), *Sun Flower* (SF) and *Street* (ST). To evaluate the average encoding performance over a range of QPs, the Bjontegaard delta peak signal-to-noise ratio (BD-PSNR) and Bjontegaard delta bit-rate (BD-BR) specified in [8] are used.

The simulation results are listed in Table 2. It indicates that the proposed algorithm achieved better performance in terms of encoding time saving. In particular, the proposed algorithm performed well for not only slow-motion or simple context video sequences (such as Akiyo and Silent) with about 54% time reduction, but also complex video sequences (such as Football and Soccer) with average savings 43.5% and 51.1%. While the complexity reduction of Zhi's algorithm in these complex motion sequences (Football and Soccer) are only 20.7% and 31.1%. Even for

Table 2. Performance comparison of the proposed algorithm

Seq.	Zhi's algorithm [2]			Proposed algorithm		
	TS (%)	BDBR (%)	BDPSNR (dB)	TS (%)	BDBR (%)	BDPSNR (dB)
AK	46.53	0.65	-0.06	53.63	-0.15	-0.01
CG	35.03	0.65	-0.02	45.68	0.22	-0.02
FB	20.73	1.33	-0.02	43.52	1.00	-0.05
SL	44.31	3.95	-0.08	54.96	0.31	-0.02
SC	31.09	1.81	-0.01	51.05	0.97	-0.01
CF	44.01	0.45	-0.03	50.25	0.09	-0.02
RH	33.80	-0.23	-0.09	44.94	-0.10	-0.04
HL	41.33	2.17	-0.08	48.47	0.79	-0.03
SF	31.42	0.20	-0.03	49.51	-0.12	-0.03
ST	42.33	0.87	-0.04	50.10	0.22	-0.02
Average	37.06	1.19	-0.05	49.21	0.32	-0.03



(a) Coastguard (b) Cafe

Fig. 6 Rate-Distortion performance comparison of JM 16.1, Zhi's and proposed FMD at various QPs on Coastguard and Cafe

the HD sequences, the proposed algorithm could also get better performance with average 48.65% of time reducing.

To illustrate the R-D performance, the RD curves sequences are given in Fig. 6. It can be seen that the proposed fast mode decision algorithm can retain almost the same RD performance as that of the JM. More details can also be referred to Table 2. The result shows that the proposed algorithm has overall better RD performance than Zhi's algorithm. Further comparison has been done between the proposed algorithm and the CBP-FMD [5] at lower QP condition. The results are shown in Table 3. It can be seen that the RH based algorithm obtains higher efficiency from QP24 to QP32. It infers that the proposed algorithm has successfully solved the drawback of CBP-FMD algorithm that the CBP mechanism loses the sensitive to detect blocks' homogeneity at lower QP conditions.

5. CONCLUSION

In this paper a hybrid fast motion decision algorithm is proposed by detecting residual and motion homogeneity. Three residual homogeneity indicators were generated to determine the homogeneity for both macroblock and sub-macroblock levels. Thus, the asymmetrical modes, such as 8x16 and 16x8, are more easily to be excluded early from the candidate subset. Motion vector activity was also considered to improve the performance of the proposed method. The proposed algorithm has achieved significant complexity reduction with negligible loss in RD performance. Especially the proposed algorithm can achieve better performance at low QP condition.

Table 3. Performance comparison of the proposed algorithm to the CBP based algorithm [5] at low QP condition

Seq.	CBP based algorithm TS			Proposed algorithm TS		
	QP 24	QP 28	QP 32	QP 24	QP 28	QP 32
CG	20.4	32.3	45.0	41.1	45.1	48.3
FB	14.8	5.4	39.9	39.3	42.3	45.7
SC	31.3	44.26	57.4	49.3	52.4	54.8
RH	45.6	55.2	62.0	46.6	56.1	64.4
ST	47.4	48.1	53.3	52.9	55.3	50.8

Acknowledgement: This work has been supported by the MUSCADE Integrating Project (www.muscade.eu), funded under the European Commission ICT 7th Framework Programme

6. REFERENCES

[1] S. Liquan, L. Zhi, Z. Zhaoyang, and S. Xuli, "Fast Inter Mode Decision Using Spatial Property of Motion Field," Multimedia, IEEE Transactions on, vol. 10, pp. 1208-1214, 2008.

[2] L. Zhi, S. Liquan, and Z. Zhaoyang, "An Efficient Intermode Decision Algorithm Based on Motion Homogeneity for H.264/AVC," Circuits and Systems for Video Technology, IEEE Transactions on, vol. 19, pp. 128-132, 2009.

[3] L. Chen, Z. Tianruo, J. Xin, W. Minghui, and S. Goto, "Residual analysis based fast inter mode decision for H.264/AVC," in Intelligent Signal Processing and Communication Systems, 2009. ISPACS 2009. International Symposium on, 2009, pp. 339-342.

[4] Y. Shih-Hsuan, C. Bo-Yuan, and W. Kuo-Hsin, "H.264 Fast Inter-Mode Selection Based on Coded Block Patterns," Information and Systems, IEICE Transactions on vol. E92.D, pp. 1324-1327 2009.

[5] S. Zhiru, W. A. C. Fernando, and A. M. Kondoz, "A Hybrid Coded Block Patterns Based Fast Mode Decision in H.264/AVC," in Multimedia and Expo Workshops (ICMEW), 2012 IEEE International Conference on, 2012, pp. 13-18.

[6] Z. Huanqiang, C. Canhui, and M. Kai-Kuang, "Fast Mode Decision for H.264/AVC Based on Macroblock Motion Activity," Circuits and Systems for Video Technology, IEEE Transactions on, vol. 19, pp. 491-499, 2009.

[7] JM 16.1. http://iphome.hhi.de/suehring/tml/download/

[8] Gisle Bjontegaard, "Calculation of average PSNR differences between RD-curves," Video Coding Experts Group (VCEG), 2001.