RD-OPTIMIZATION FOR MPEG-2 TO H.264 TRANSCODING

Gerardo Fernandez-Escribano[1], Hari Kalva[2], Pedro Cuenca[1] and Luis Orozco-Barbosa[1]

[1]Instituto de Investigación en Informática de Albacete. Universidad de Castilla-La Mancha.
Campus Universitario, 02071 Albacete, SPAIN.
[2]Department of Computer Science and Engineering, Florida Altantic University, Boca Raton, FL, USA

ABSTRACT

In this paper, we introduce and evaluate a low complexity macroblock partition mode decision algorithm for inter-frame prediction in MPEG-2 to H.264 transcoder. The proposed tools are used to compute an optimal MB coding mode decision with significantly reduced computational complexity. Specifically, we achieve the computational savings by using the following MB information coming from MPEG-2: the MB coding modes, the coded block pattern (CBPC) in MPEG-2, and the mean and variance of the 16 4x4 sub blocks of the MPEG-2 residual MBs. We use data mining algorithms to develop a decision tree for H.264 coding mode decisions. The decision trees are built using RD optimized mode decisions and result in highly efficient mode decisions. The proposed transcoder is 35% faster than the RD optimized H.264 reference transcoder without a significant PSNR degradation. The proposed transcoder performs over 3 dB better than the SAE cost based H.264 transcoding.

1. INTRODUCTION

Transcoding MPEG-2 video to H.264 is important to enable gradual migration to H.264, also known as MPEG-4 AVC [1]. However, given the significant differences between the MPEG-2 and the H.264 coding algorithms, transcoding is much more complex and new approaches to transcoding are necessary. The main problems that need to be addressed in the design of an efficient heterogeneous MPEG-2/H.264 transcoder are: the inter-frame prediction, the transform coding and the intra-frame prediction. These problems are being examined in various research efforts that are underway [2-7]. In this paper, we focus our attention on a part of the inter-frame prediction: the *macroblock partition mode decision*, one of the most computationally intensive tasks involved in the encoding process.

The H.264 video coding standard uses the notion of *macroblock partition* to refer to the group of pixels in a macroblock that share a common prediction. The encoder selects the coding-modes for the macroblock, including the best macroblock partition and mode of prediction for each macroblock partition, such that the video coding performance is optimized. In this paper, we introduce and evaluate a novel macroblock partition mode decision algorithm with emphasis on RD optimized transcoding and complexity reduction in the inter-frame prediction in a MPEG-2 to H.264 transcoder. We achieve the computational savings by reusing the following MB information coming from the MPEG-2 decoding stage: the MB coding modes, the coded block pattern (CBPC) in MPEG-2, and the mean and variance of the 16 4x4 sub blocks of the MPEG-2 residual MBs. From an exhaustive analysis of this information, we derive our tree decision suitable for their integration into our algorithm. Our results show that the proposed transcoder is 35% faster than the RD optimized H.264 reference transcoder without a significant PSNR degradation. The proposed transcoder performs over 3 dB better than the SAE cost based H.264 transcoding.

The rest of the paper is organized as follows. Section 2 describes the data mining tools and the process of building a decision tree for MB mode estimation. Section 3 introduces our macroblock partition mode decision algorithm specifically designed for MPEG-2 to H.264 transcoders. Section 4 presents the results and discussion. We compare the performance of our proposal to the SAE-cost and the RD-optimized methods proposed by the H.264 standard. Finally, Section 5 draws our conclusions and outlines our future research plans.

2. DATA MINING FOR THE MODE DECISION TREE

In the world of machine learning, a decision tree is made by mapping the observations about a set of data in a tree made of arcs and nodes. The nodes are the variables and the arcs the possible values for that variable. The tree can have more than one level; in that case, the nodes (leafs of the tree) represent the decision based on the values of the different variables that drives us from the root to the leaf. These types of trees are used in the data mining processes for discovering the relationship in a set of data, if it exits. The tree leafs are the classifications and the branches are the features that lead to a specific classification.

This work was partly supported by the Ministry of Science and Technology of Spain under CICYT project TIC2003-08154-C06-02 and the Council of Science and Technology of Castilla-La Mancha.

The decision tree was made using the WEKA data mining tool. The files that are used for the WEKA data mining program are known as ARFF [8] (Attribute-Relation File Format) files. An ARFF file is written in ASCII text and shows the relationship between a set of attributes. Basically, this file has two different sections; the first section is the header with the information about the name of the relation, the attributes that are used and their types; and the second data section contains the data. In the header section, we have the attribute declaration. For each macroblock, the proposed algorithm uses sixteen means and sixteen variances of 4x4 residual sub-blocks, the MB mode in MPEG-2 (skip, intra, and three non-intra modes), the coded block pattern (CBPC) in MPEG-2, and the corresponding H.264 MB coding mode decision for that MB as determined by the standard reference software. The following code shows our declaration for the ARFF files:

```
@RELATION mean-variance_4x4

@ATTRIBUTE mean0 Numeric
@ATTRIBUTE variance0 Numeric
@ATTRIBUTE mean1 Numeric
@ATTRIBUTE variance1 Numeric
.................................................................
@ATTRIBUTE mean15 Numeric
@ATTRIBUTE variance15 Numeric
@ATTRIBUTE mode_mpeg2 {0,1,2,4,8}
@ATTRIBUTE CBPC0 {0,1}
.................................................................
@ATTRIBUTE CBPC6 {0,1}
@ATTRIBUTE class {0,1,8,9}
```

The supposed dependent variable, namely **class** in the example, is the variable that we are trying to understand, classify, or generalize. The other variables are the variables that are going to help us to make the classification. The ARFF data section has the instance lines, which are the samples used to train our model. Each macroblock sample is represented on a single line.

The decision tree, that is proposed to solve the inter-prediction problem, is a model of the data that encodes the distribution of the class label (namely **class** in the example) in terms of the attributes (the others variables in the example). The final goal of this decision tree is to help us to find a simple structure to show the possible dependences between the attributes and the class.

3. LOW COMPLEXITY MODE DECISION ALGORITHM

This section discusses the proposed macroblock partition mode decision algorithm (the decision tree, figure 1) aiming to accelerate the inter-frame prediction. This goal is achieved by making use of the MPEG-2 MB coding mode, the coded block pattern (CBPC), and the mean and variance of the residual information for this MB calculated for its 4x4 sub-blocks (figure 2a., 2b. and 2c.). MPEG-2 uses 16x16 motion compensation (MC) and does not temporally decorrelate an image fully. The MC residual can thus be exploited to understand the temporal correlation of variable block sizes in H.264.

The open source WEKA [8] data mining tool is used to discover a pattern of mean, variance, MPEG-2 coding modes, and the coded block pattern in MPEG-2 (CBPC) for H.264 coding mode decisions. The exit of the tree is the H.264 MB coding mode (intra, skip, 8x8 and 16x16). The training set was made with the flower MPEG-2 sequence encoded at 8 Mbps, with P frames. The H.264 decisions in the training set were from encoding the MPEG-2 decoded sequence with a quantization parameter of 25 and RD optimization enabled. We found that this one sequence is sufficient to capture the mean and the variance distribution of the residual in MPEG-2 coded video sequences and the decision tree made with this training set was able to make accurate mode decisions for all tested sequences, as we will show in the performance section. Based on this training file, the WEKA data mining tool was used to create a decision tree (Figure 1, node 1) with a set of rules for using the mean and variance of the 4x4 sub-blocks. We used the J.48 algorithm to build the tree. The J4.8 algorithm is based in the C4.5 algorithm proposed by Ross Quinlan [9].

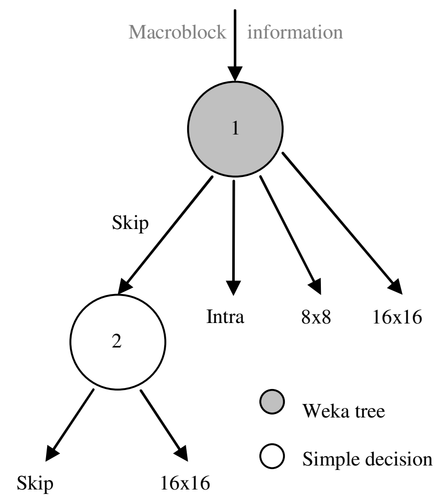


Figure 1. The Decision Tree

We use one WEKA decision tree, shown in the picture with a grey ball. The Weka tree is used to check for the skip, Intra, 8x8 and 16x16 MBs modes. If a MB is skipped, a second decision tree is used for selecting the mode of the MB. The decision tree works as follows:

Node 1. The inputs for this node are all the MPEG-2 coded MBs. In this node we use a tree decision generated with WEKA to decide whether the MB should be coded in H.264. This tree examines whether the MB has a very high residual or a medium residual. The output of this node is the mode that will be use for coding the MB: skip, Intra, 8x8 or 16x16. For example, the following rules was given by WEKA:

- **a**. If the MPEG-2 MB was coded in skip mode, then the decision should be coding as skip (see node 2).
- **b**. If the MPEG-2 MB was coded in intra mode, then the MB will be coding as intra in H.264, or 8x8. This decision depends of the residual.
- **c.** If the MPEG-2 MB was not coded in MC, (non-zero MV present, none of the 8x8 block has coded coefficients), then the MB will be coding as 16x16 in H.264.

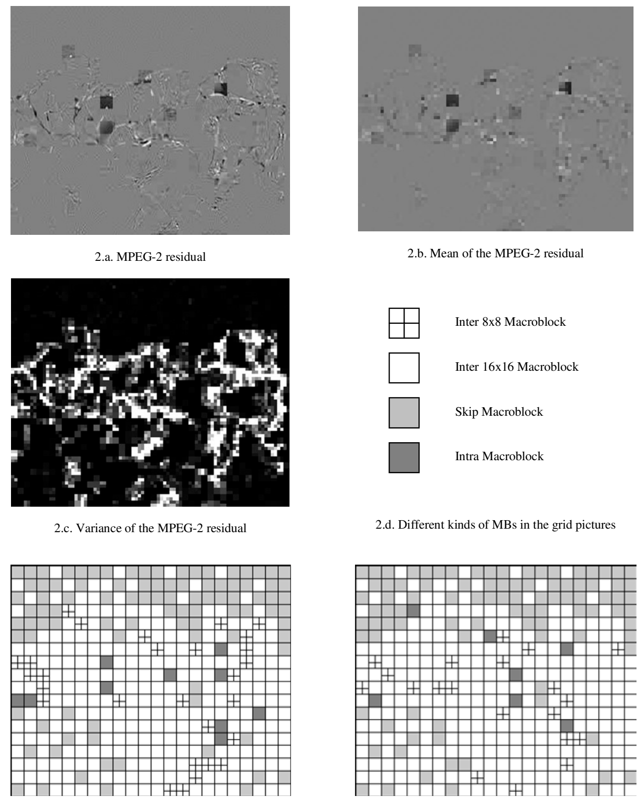
Node 2. The inputs for this node are skip-mode MBs in the MPEG-2 bitstream proposed by the node 1. This node evaluates only the H.264 16x16 mode (without the modes 16x8 or 8x16). Then, the node selects the best option, skip or inter 16x16.

The WEKA tool was used to determine one set of mean and variance threshold for the MPEG-2 residual, for skip, Intra, 8x8 or 16x16 decision in node 1. Due to space constraints we cannot show all the rules that are evaluated in the WEKA decision nodes. This decision node implements a decision tree that examines the mean and variance of MPEG-2 MC residual in 4x4 sub-blocks to arrive at the decision. Since the MB mode decision, and hence the thresholds, depend on the quantization parameter (QP), the decision tree thresholds are computed for a mid-QP of 25 and then adjusted for other QPs. Since the quantization step size in H.264 doubles when QP increases by 6, the thresholds are adjusted by 12.5% for a change in QP of 1.

The proposed algorithm only works with the first level of the decision in P frames. The algorithm determines whether the macroblock is coded as skip, intra, inter 8x8 or inter 16x16. For the second level decision, the proposed mechanism uses the reference software method: running all the possibilities and selecting the best sub-partition (with the Rd-optimized option enable). Work is ongoing to develop decision trees for sub-partition decisions.

4. RESULTS AND DISCUSSION

In order to evaluate the proposed macroblock partition mode decision algorithm, we have implemented the proposed approach based on the H.264 reference software [10] (version 10.2). The metrics we used to evaluate the algorithm are the computational cost reduction and rate distortion function. Throughout our experiments, we have used various video sequences exhibiting different spatial characteristics and different size formats (CCIR, CIF and QCIF). We use Q factors from QP=0 to QP=50 (corresponding to the full H.264 QP range). The size of the GOP is 12 frames; where the first frame of every GOP was encoded as I-frame, and the rest of the frames of the GOP were encoded as a P-frames. The rate control was disabled for all the simulations.



2.a. MPEG-2 residual

2.b. Mean of the MPEG-2 residual

2.c. Variance of the MPEG-2 residual

2.d. Different kinds of MBs in the grid pictures

2.e. H.264 Rd opt, first frame P, Football (CIF) QP= 25. Inter mode selected by H.264

2.f. H.264 Rd opt, first frame P, Football (CIF) QP= 25. Inter mode selected by our proposed

Figure 2. Macroblock partitions generated by the proposed algorithm for the first P-frame in the football sequence.

Figure 2.e and 2.f shows the differences between the inter mode selection made by the H.264 standard (with the RD-optimized option enabled), and the proposed algorithm, with a value of 25 for QP.

Figure 3 shows the RD results of applying the H.264 full estimation algorithm (with and without the RD-optimized option) and the proposed macroblock partition mode decision algorithm (with Rd-optimized option enabled). For all the experiments, the Intra mode decision for P pictures was disabled since we are evaluating the performance of our algorithm for inter MB mode decisions. The ProfileIDC was set to High for all the simulations, with the FRExt options enabled. As seen from the figure, the PSNR obtained when applying our algorithm deviates slightly from the results obtained when applying the considerably more complex full motion estimation procedure. The average PSNR gain over the standard implementation with SAE cost and RD disabled is 3 dB with similar computational cost. We have better results than the standard with the Rd-optimized option disable, namely H.264 (SAE) in the figure.

Table 1 shows the computational cost (with respect to H.264 full estimation procedure, with and without the Rd-optimized option) for a P frame (using a QP factor between

20 and 40), for different video sequences and size formats. Our results show over 35% reduction in the computational complexity for the first level MB decisions alone. The simulations were running on a P4 at 3.0 GHz Intel machine with 512 MB RAM. Our results show that the proposed algorithm is able to maintain a good picture quality while considerably reducing the number of operations to be performed.

Table 1. Time results.

	Total encoding time for a P frame (milliseconds)			Time reduction from Rd opt (%)	
SQ, QP	H.264 (Rd opt)	H.264 (SAE)	Proposed (Rd opt)	H.264 (SAE)	Proposed (Rd opt)
Ayersroc, 20	2902	1892	1973	34.8	32
Ayersroc, 25	2793	1901	1842	31.9	34
Ayersroc, 30	2726	1962	1797	28	34.1
Ayersroc, 35	2743	2018	1777	26.4	35.2
Ayersroc, 40	2784	2091	1767	24.9	36.5
Football, 20	889	577	654	35.1	26.4
Football, 25	866	587	585	32.2	32.4
Football, 30	846	610	589	27.9	30.4
Football, 35	849	628	576	26	32.2
Football, 40	843	636	569	24.6	32.5
Akiyo, 20	188	134	110	28.7	41.5
Akiyo, 25	190	137	102	27.9	46.3
Akiyo, 30	189	139	101	26.5	46.6
Akiyo, 35	194	143	101	26.3	47.9
Akiyo, 40	192	149	100	22.4	47.9

5. CONCLUSIONS

In this paper, we proposed a novel macroblock partition mode decision algorithm for inter-frame transcoding in an MPEG-2 to H.264 transcoder. The proposed algorithm used datamining techniques to exploit the temporal correlation in the MPEG-2 MC residual. The WEKA tool was used to develop decision trees for H.264 coding mode decision. The proposed algorithm has very low complexity as it only requires the mean and variance of the MPEG-2 residual, and the MB information coming from MPEG-2; the MB coding modes and the coded block pattern (CBPC). Our results show that the proposed algorithm is able to maintain a good picture quality while considerably reducing the computational complexity by 35% on average. We are currently working on algorithms, based on the proposed datamining approach, for sub-partition decisions in the H.264 inter MB coding and for the Intra frame prediction in H.264 when the Rd optimized mode decision is enable.

6. REFERENCES

[1] ITU-T RECOMMENDATION H.264 “Advanced Video Coding for Generic Audiovisual Services”. May 2003.

[2] B. Petljanski and H. Kalva, "DCT domain intra MB mode decision for MPEG-2 to H.264 transcoding" Proceedings of the ICCE 2006. January 2006. pp. 419-420.

[3] C. Chen, P-H.Wu, and H. Chen, "MPEG-2 to H.264 transcoding," Picture Coding Symposium, 15-17 Dec, 2004.

[4] Z. Zhou, S. Sun, S. Lei and M-T. Sun, "Motion Information and Coding Mode Reuse for MPEG-2 to H.264 Transcoding", Proceedings of the ISCAS 2005, May 2005. pp 1230-1233.

[5] Y. Su, J. Xin, A. Vetro, and H. Sun, "Efficient MPEG-2 to H.264/AVC Intra Transcoding in Transform-domain", Proceedings of the ISCAS 2005, May 2005. pp 1234-1237.

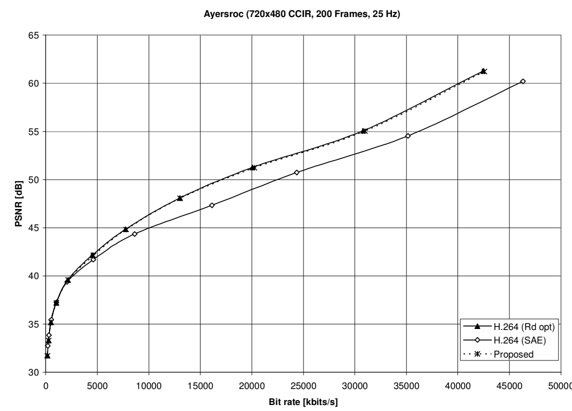
[6] G. Chen, Y. Zhang, S. Lin, and F. Dai, "Efficient block size selection for MPEG-2 to H.264 transcoding," Proceedings of the 12th annual ACM Multimedia, 2004, pp. 300-303.

[7] G. Fernandez-Escribano, P. Cuenca, L. Orozco-Barbosa and A. Garrido. “Computacional complexity reduction of Intra-Frame prediction in MPEG-2/H.264 video transcoders”. Proceeding of ICME 2005, Amsterdam, The Netherlands, July 2005.

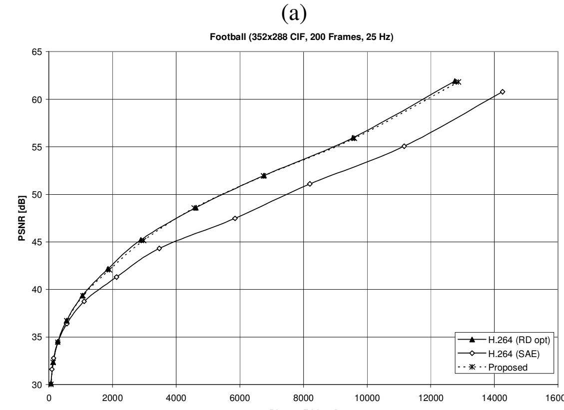
[8] Ian H. Witten and Eibe Frank, “Data Mining: Practical machine learning tools and techniques”, 2nd Edition, Morgan Kaufmann, San Francisco, 2005.

[9] J.R. Quinlan, “C4.5: programs for machine learning”, Morgan Kaufmann, 1993.

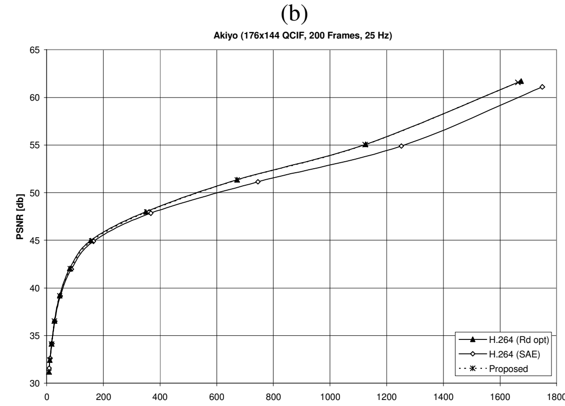
[10] Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, Reference Software to Committee Draft. JVT-F100 JM10.2. 2006.



(a)



(b)



(c)

Figure 3. Rate Distortion Results. (a) CCIR. (b) CIF. (c) QCIF.