DCT DOMAIN INTRA MB MODE DECISION FOR MPEG-2 TO H.264 TRANSCODING

Branko Petljanski and Hari Kalva
Computer Science and Engineering, Florida Atlantic University, Boca Raton, FL

Abstract

This paper presents a macroblock (MB) mode estimation technique for MPEG-2 to H.264 intra frame video transcoding. This technique can be applied in number of consumer electronic applications such as DVRs and content delivery systems. The DCT coefficients gathered from the MPEG-2 process are used to estimate the MB coding mode. The estimated mode is used to constrain the MB prediction mode computation just to the estimated MB mode. Our approach uses the DCT coefficients obtained during the MPEG-2 decoding stage and classifies an MB as INTRA16x16 or INTRA4x4. The computational cost of the DCT domain operations is negligible but results in substantial savings in the H.264 mode estimation. The results show that the PSNR drop caused by this direct estimation is negligible.

I. INTRODUCTION

The wide use of the MPEG-2 video today and the expected adoption of H.264 creates a need for tools and techniques for MPEG-2 to H.264 video transcoding. Video transcoding discussed in this paper is the process of converting video in the MPEG-2 format to the H.264 format. The obvious way of achieving this is using a reference transcoder, an MPEG-2 decoder followed by an H.264 encoder. This approach represents the upper bound on transcoding complexity as well as the video quality. The key issues in video transcoding are minimizing the complexity while maximizing the quality [1].

There is a relatively small amount of published literature on MPEG-2 to H.264 video transcoding. Chen et. al. present an MPEG-2 to H.264 video transcoder with operations in the transform domain [2]. Lu et. al. present a fast mode decision focusing on MPEG-2 to H.264 transcoding [3]. This approach is based on reducing the mode estimation complexity by reducing the number of prediction modes evaluated based on a selected threshold. INTRA16x16 mode is evaluated fully, and based on the cost of the INTRA16x16 block, the INTRA4x4 mode complexity is reduced by decreasing the number of modes evaluated. Xin et. al. reported a tool for transforming DCT coefficients to H.264 coefficients that can be useful in MPEG-2 to H.264 transcoding [4].

The proposed approach to complexity reduction uses the MPEG-2 DCT coefficients to estimate the MB coding modes and the best MB prediction modes. By adding the MB mode and MB prediction mode stages, the complexity of the reference transcoder can be reduced substantially. A reduced complexity transcoder with DCT domain prediction mode estimation based on the directional textures is described in our earlier work [5]. In this paper we present the details of the DCT variance based technique for MB mode estimation.

II. MB MODE ESTIMATION

The SAD (sum of absolute differences) of the predicted and original blocks is typically used to determine the optimal block size (mode) for intra MB coding. An intra MB is coded as either a group of 4x4 blocks or a single 16x16 block (or 8x8 blocks with recent amendment to the H.264 standard). It can be shown that there is a correlation between activity within a MB, required quantization, and mode estimation. Intensity variance in the spatial domain has been widely used as a metric for measuring activity within a region of an image.

The variance of a 16x16 block is strongly correlated with the variance of its non-overlapping 8x8 sub-blocks. Since the MPEG-2 decoding stage yields the DCT coefficients of 8x8 blocks, DCT variance of four adjacent blocks can be used to estimate the variance of a combined 16x16 block. A variance threshold is determined as a function of the selected H.264 quantization parameter (QP). If the estimated variance of the 16x16 block is less than a pre determined threshold, then the MB is coded in INTRA16x16 mode; otherwise the MB is coded in INTRA4x4 mode. The thresholds were experimentally determined using Football, Foreman, and Akiyo sequences. These three sequences represent the content with high, medium, and low spatial activity.

III. RESULTS AND DISCUSSION

The experiments are conducted with JM8.5 and H.264 baseline profile. The encoder was configured to encode all frames as H.264 intra frames with a constant quantizer. The equivalent MPEG-2 video bitstream at the input of the H.264 encoder uses a constant quantizer of 16 to quantize DCT blocks. The H.264 encoder uses SAD to decide the optimal block size for intra MB coding.

Figure 1 shows the variance threshold curves for the three test sequences for the full QP range. The threshold curves are constructed to minimize the number of MB mode mismatches when compared with the reference transcoder. The curves clearly indicate that for the three test sequences, the variance threshold below which a MB is coded as INTRA16x16 are very similar. A generalized threshold curve is thus formed by averaging the threshold

0-7803-9459-3/06/$20.00

values of all three sequences for each QP. This generalized curve is then used to determine the intra MB mode with minimal computational cost.

Table 1 shows the effect of using the optimized and generalized threshold on the PSNR. The table shows the average PSNR for the first 30 frames as a function of QP. Reference transcoder columns represent average PSNR for the first 30 frames using the reference transcoder. Optimized DCT threshold represents PSNR when the proposed transcoder uses the optimized threshold computed for that specific sequence. Generalized DCT threshold uses the same threshold curve for all three sequences. It can be seen from table that using both optimized and general thresholds closely follows the reference PSNR. Maximum PSNR loss is 0.38 dB for QP 45 in the foreman sequence (not shown in table because of space considerations). These results clearly establish that the use of the generalized threshold results in a negligible drop in the PSNR while eliminating all the computations for the MB mode estimation in the H.264 encoding stage.

Figure 2 shows the MB mode decisions for frame 30 of the Football sequence made using the reference transcoder and the proposed transcoder. The overlaid 4x4 grid corresponds to INTRA4x4 MB and a 16x16 square corresponds to an INTRA16x16 MB. As shown in the figure, the proposed approach closely matches the optimal MB modes as determined by the reference transcoder. The small number of mismatches does not have any significant impact on the quality of the transcoded video. The reduction in computational cost, however, is substantial.

IV CONCLUSION

This paper presents a DCT domain MB mode decision technique for MPEG-2 to H.264 video transcoding. The MB mode is estimated based on the pre-determined variance threshold. The results show that the variance thresholds used are largely content independent and reduce the computational complexity. The estimated mode is then used to constrain the MB mode prediction just to the estimated MB mode. The results show that the PSNR drop caused by this variance based estimation is negligible. This technique can also be applied to transcoding DCT based video such as MPEG-4 to the H.264 format.

REFERENCES

[1] H. Kalva, "Issues in H.264/MPEG–2 video transcoding," *Proceedings of the IEEE Consumer Communications and Networking Conference*, Jan. 2004, pp. 657–659.

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

[3] X. Lu, A. M. Tourapis, P. Yin, and J. Boyce, "Fast mode decision and motion estimation for H.264 with a focus on MPEG-2/H.264 transcoding," *Proceedings of 2005 IEEE International Symposium on Circuits and Systems (ISCAS)*, Kobe, Japan, May 2005.

[4] J. Xin, A. Vetro, and H. Sun, "Converting DCT coefficients to H.264/AVC transform coefficients," *Proceedings of the IEEE Pacific-Rim Conference on Multimedia (PCM)*, IEEE PCM 2004, November 2004

[5] H. Kalva, B. Petljanski, and B. Furht, "Complexity Reduction Tools for MPEG-2 to H.264 Video Transcoding," WSEAS Transactions on Information Science and Applications, Vol. 2, Issue 3, March 2005, pp. 295-300.

Table 1 PSNR Comparison with 1. Reference Transcoder 2. Optimized DCT Threshold 3. Generalized DCT Threshold

QP	Football			Foreman			Akiyo		
	1	2	3	1	2	3	1	2	3
10	51.07	51.07	51.07	51.08	51.08	51.08	51.37	51.37	51.37
20	42.48	42.48	42.48	42.93	42.93	42.93	42.9	42.9	42.9
30	34.51	34.51	34.49	36.35	36.34	36.35	38.99	38.99	39.02
40	28.25	28.24	28.1	30.63	30.48	30.43	32.71	32.71	32.72
50	22.84	22.75	22.71	24.86	24.62	24.61	26.76	26.67	26.68

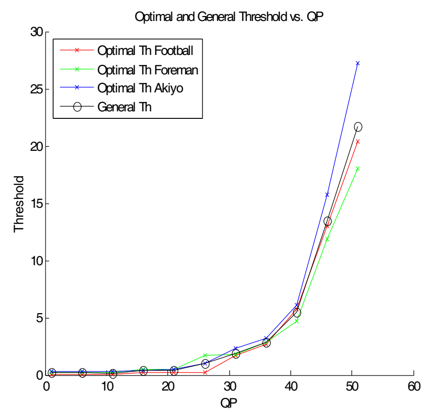

Figure 1. DCT variance threshold for frame 30 vs. QP

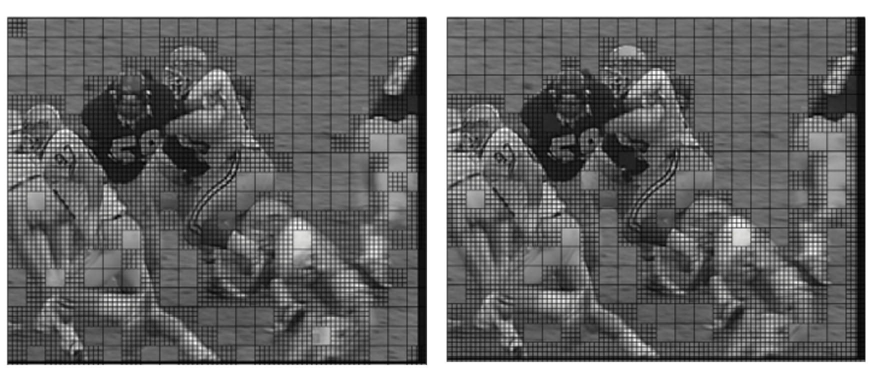

Figure 2. MB Mode estimation using the (a) reference transcoder and (b) the proposed transcoder