Exploiting the Directional Features in MPEG-2 for H.264 Intra Transcoding

Hari Kalva, *Senior Member*, IEEE and Branko Petljanski, *Student Member*, IEEE

Abstract — *This paper presents an MPEG-2 to H.264 intra frame video transcoder. The DCT coefficients gathered from the MPEG-2 process are used to estimate the directional features in a picture. The estimated directional features are then used to compute the intra prediction modes in the H.264 encoding stage. The INTRA16x16 vs. INTRA4x4 MB coding mode decision is made based on the variance of the DC coefficients of the 8x8 DCT blocks in the input MPEG-2 video. The proposed approach completely avoids the MB mode and prediction mode computations in the H.264 encoding stage, there by reducing the transcoder complexity substantially. The results show that our approach results in over 30% drop in the intra frame encoding time with a negligible drop in PSNR. A comparison with a cascaded reference transcoder is presented.*

Index Terms — H.264 transcoding, MPEG-2 transcoding, DCT domain, directional features, edge angle.

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 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]. Reduced complexity transcoding is necessary when resources are limited or when real-time/high-density transcoding is desired. Complexity reduction can be usually achieved by reusing the information gathered in the decoding stage of the transcoder.

There is 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 is the transform domain [2]. A matrix multiplication based approach is developed to convert the MPEG-2 DCT coefficients to the H.264 transform coefficients. The paper does not discuss the issues of intra MB mode and the MB prediction mode selection.

H. Kalva is with the Department of Computer Science and Engineering, Florida Atlantic University, Boca Raton, FL 33431, USA (e-mail: hari@cse.fau.edu).

B. Petljanski is with the Imaging Technology Center, Florida Atlantic University, Boca Raton, FL 33431, USA (e-mail: pbranko@gmail.com).

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 selectively evaluating the prediction modes. INTRA16x16 mode is evaluated fully, and based on the cost of the INTRA16x16 block, the INTRA4x4 mode complexity is reduced by reducing the number of modes evaluated.

Su et. al. reported a transform domain transcoder for MPEG-2 to H.264 transcoding [4]. Motion vector reuse for fast inter mode prediction was presented in [5] and [6].

The transcoding work that has been reported focuses on MB mode mapping to reduce the complexity. Estimating the intra prediction directions has not addressed. Since obtaining the right directional predictor is essential for improving the prediction efficiency, selective mode evaluation based approaches can lead to sub-optimal performance. Our approach is not to reduce the number of prediction modes evaluated but to find a solution that can determine the right prediction mode without performing any mode evaluations. Such an approach will lead to low complexity intra transcoder.

In this paper we present an innovative approach to intra MB mode and intra prediction mode computation using the DCT coefficients gathered from the MPEG-2 process. Our approach relies on estimating the directional features in the input MPEG-2 video using the DCT coefficients and the properties of the DCT transforms to estimate MB mode and MB prediction mode. Unlike the other transcoding approaches that have appeared in the literature, our approach completely avoids the MB mode and prediction mode computations there by reducing the complexity substantially. The proposed approach also estimates the optimal prediction mode, again, without evaluating any modes in the H.264 encoder. Our earlier work presented the mode estimation and DCT transformation tools needed for the direct mode mapping [7].

This paper is organized as follows. Section II presents the complexity reduction tools developed for MPEG-2 to H.264 transcoding. Section III discusses threshold design for MB mode estimation, the results and discussion is presented in Section IV. Conclusions are presented in Section V.

II. COMPLEXITY REDUCTION TOOLS

The proposed approach to complexity reduction uses the MPEG-2 DCT coefficients to estimate the MB coding modes and the best MB prediction modes. Figure 1 shows the block diagram of the proposed reduced complexity intra frame transcoder. The transcoder is comprised of an MPEG-2 decoding stage followed by the H.264 encoding stage. The subsystem shown in the shaded rectangle corresponds to a

Manuscript received April 14, 2006

reference transcoder that performs MPEG-2 decoding followed by a full H.264 encoding stage. Complexity reduction is achieved by adding MB mode and MB prediction mode estimation modules. The DCT coefficients gathered in the MPEG-2 decoding stage are used to estimate the MB modes and MB prediction modes. The complexity reduction tools used to estimate the modes are described briefly in this section.

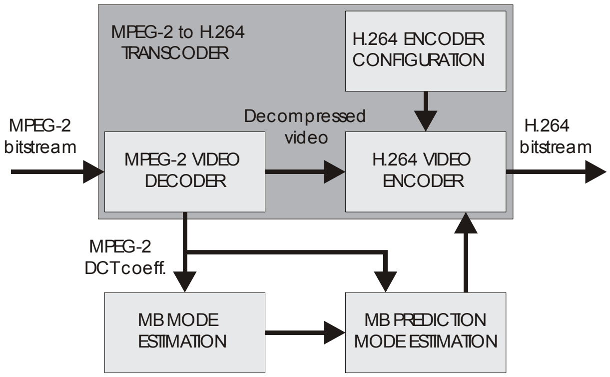

Fig. 1. Reduced Complexity Transcoder

A. MB Mode Estimation

The sum of absolute differences (SAD) of 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 FRExt extensions to the H.264 standard). There is a correlation among the activity within a MB, required quantization, and mode estimation and such a correlation can be exploited to determine the MB coding mode. These interdependencies are discussed Section III. Intensity variance in the spatial domain has been widely used as a metric for measuring activity within a region of an image. Variance of the pixels in the spatial and frequency domain is directly related and variance is good measure of the activity with in a block. Using the variance of DCT coefficients as a metric for a video segmentation between more and less active regions of an image has been reported in [8].

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, variance of four adjacent blocks can be used to determine the MB coding mode. A variance threshold is determined as a function of the selected H.264 quantization parameter. If the estimated variance of the 16x16 block is less than a threshold, then the MB is coded in INTRA16x16 mode; otherwise the MB is coded in INTRA4x4 mode. The process of determining the variance thresholds is discussed in Section III.

B. Prediction Mode Estimation

Prediction mode estimation is based on estimating the directional features of the image blocks in the DCT domain. The DCT coefficients of the 8x8 blocks gathered during the MPEG-2 decoding stage are used. The DCT coefficients have been used for analyzing image and videos in the compressed domain [9], including edge detection [10] and scene change detection [10].

Figure 2 shows the edge model used for computing the directional feature and the prediction directions in H.264. It has been shown that the ratio of vertical energy to the horizontal energy gives the tangent of the ideal edge passing through the center of the block [10][12]. Let $F(u,v)$ represent the DCT coefficients of an 8x8 block obtained during the MPEG-2 decoding stage. The edge angle, mean, and variance can be obtained in the DCT domain using equations (1)–(3).

$$\tan\theta = \frac{\sum_{u=1}^{7} |F(u,0)|}{\sum_{v=1}^{7} |F(0,v)|} \tag{1}$$

$$\mu = \frac{1}{64}\sum_{0}^{7}\sum_{0}^{7} f(x,y) = \frac{F(0,0)}{8} + 128 \tag{2}$$

$$\sigma^2 = \frac{1}{64}\sum_{u=0}^{7}\sum_{u=0}^{7} F(u,v)^2 \qquad (u,v) \neq (0,0) \tag{3}$$

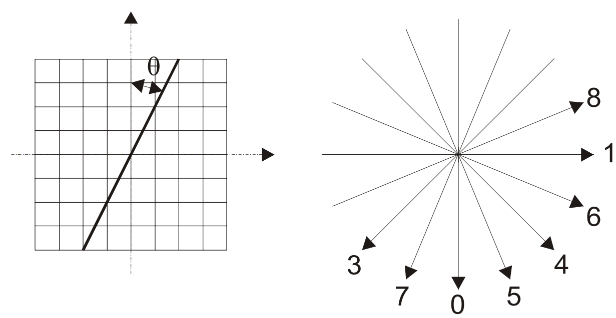

Fig. 2. Edge orientation in an 8x8 block (left) and H.264 intra prediction modes (right)

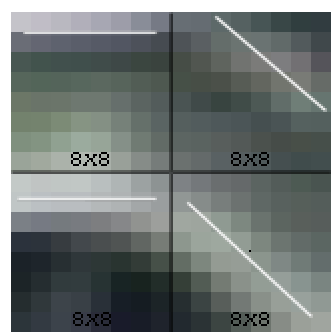

Fig. 3. Edge orientation in the 8x8 sub-blocks of a macro block and the estimated edge angle

The prediction mode estimation is based on the hypothesis that the edge orientation gives the direction of minimum energy variance within a block and hence can be mapped to an intra prediction mode. Figure 3 shows the prediction directions in 8x8 blocks overlaid on an actual image. The figure clearly shows that the prediction direction estimates closely correspond to the directional features in an image and the estimated directional features can be used to determine the intra prediction modes. Once the MB mode decision is made based on the block variance and the quantization parameter, the edge angle computed is used to estimate the prediction modes. The angle with a 10° window is used to map a direction to the prediction mode; i.e, orientation with 5° on either side of a H.264 prediction direction is mapped to that prediction mode.

While mode decisions are needed for 4x4 and 16x16 blocks, only the DCT coefficients of 8x8 blocks are available from the MPEG-2 decoding stage. These 8x8 DCT blocks can be used directly to estimate INTRA8x8 modes. Estimating INTRA16x16 and INTRA4x4 modes requires an additional tool to transform 8x8 DCT coefficients into desired 16x16 and 4x4 DCT coefficients.

C. DCT Block Splitting and Recombination

While the 16x16 and 4x4 prediction modes can be estimated based on the 8x8 MPEG-2 DCT coefficients, the accuracy can be improved if 4x4 or 16x16 DCT coefficients are available. Computing the 16x16 and 4x4 DCT coefficients after applying IDCT to the 8x8 DCT blocks is computationally expensive. We developed an approach based on DCT matrix transformations for DCT block combination and splitting. Our approach is simple and is based on the fact that the DCT transform matrix is orthonormal and separable; i.e., the transform can be represented as $\widetilde{X} = CXC^T$, where X is an image of size $N_1 \times N_2$, $\tilde{X}$ is DCT transform of X, and C is cosine matrix. For MPEG-2 video, $N_1 = N_2 = 8$ and an 8x8 DCT is used.

Deriving four 4x4 DCT blocks from a given 8x8 DCT block is described below. Full details including combining four 8x8 blocks into a 16x16 DCT block is described in [7]. Let $\tilde{X}_{8x8}$ denote an 8 x 8 block of horizontal and vertical frequency coefficients, C_{8x8} is cosine transformation matrix for block of size 8 x 8 and C_{4x4} is cosine transformation matrix for blocks 4 x 4. The resulting DCT coefficients, $\widetilde{X}_{4x4}^{mn}$, are obtained using equation (4).

$$\begin{bmatrix} \tilde{X}_{4x4}^{00} & \tilde{X}_{4x4}^{01} \\ \tilde{X}_{4x4}^{10} & \tilde{X}_{4x4}^{11} \end{bmatrix}_{8x8} = \begin{bmatrix} C_{4x4} & 0 \\ 0 & C_{4x4} \end{bmatrix} C_{8x8}{}^T \tilde{X}_{8x8} C_{8x8} \begin{bmatrix} C_{4x4} & 0 \\ 0 & C_{4x4} \end{bmatrix}^T \quad (4)$$

$$T_{8x8} = \begin{bmatrix} C_{4x4} & 0 \\ 0 & C_{4x4} \end{bmatrix} C_{8x8}{}^T \quad (5)$$

$$\begin{bmatrix} \tilde{X}_{4x4}^{00} & \tilde{X}_{4x4}^{01} \\ \tilde{X}_{4x4}^{10} & \tilde{X}_{4x4}^{11} \end{bmatrix} = T_{8x8} \tilde{X}_{8x8} T_{8x8}^T \quad (6)$$

Since C_{8x8} and C_{4x4} are constant cosine matrices, they can be replaced with a new constant matrix T_{8x8} shown in equation (5). As a result, transformation in equation (4) can be rewritten as shown in equation (6).

The complexity of using block splitting for estimating the directions in 4x4 blocks is rather low. The transformation matrix T_{8x8} has five non-zero elements per row. Prediction mode estimation described in section II.B requires only the first row and first column of 4x4 DCT blocks. In addition to this, taking advantage of the symmetry of the transform matrix reduces the computation to 288 additions and 100 multiplications for each 8x8 block. The complexity of intra mode estimation for the four 4x4 blocks of an 8x8 block requires 2096 additions, 596 shift operations, and 12 multiplications. Assuming the cost of addition and shift operations is 1 cycle and the multiplication is 10 cycles, the exhaustive mode estimation takes 2704 cycles. The block splitting approach requires 1288 cycles for block splitting and another 100 cycles of computation for mode estimation which represents about 50% savings compared to the exhaustive intra mode prediction. The computational savings are even higher when INTRA8x8 is used since we do not need DCT block transformations to determine prediction directions for INTRA8x8 MBs.

III. Designing Thresholds for MB Mode Estimation

The variance of pixel intensities in a 16x16 block gives a measure of the energy in the block. A block with large amount of detail has a larger variance and is more suitable for Intra 4x4 coding. However, with larger quantization, the details are lost during quantization and inverse quantization. The quantizer also has an impact on the bitrate. For larger quantizers, the differences between the neighboring pixels are lost in the de-quantized block because of the large quantization step size; the block thus becomes uniform after de-quantization. Larger quantizers can still maintain the differences among the neighboring pixels only when the pixel values differ substantially, i.e., variance is high. Smaller quantizers, however, maintain the differences among the neighboring pixels after de-quantization. This results in 4x4 blocks performing better with lower quantizers and 16x16 blocks performing better with higher quantizers. The key is to determine variance thresholds that can identify an intra MB mode. The threshold should also change depending on the quantizer used; i.e., a variable threshold that is a function of the quantizer has to be derived.

The thresholds were experimentally determined using Football, Foreman, and Flower sequences. These three sequences represent the content with high, medium, and low activity. This set of sequences were encoder at all allowed quantization parameters (QP) in H.264 (0-51) and for each QP a threshold is selected such that the threshold based MB mode decision would minimize the mode mismatches when compared with the modes determined by the encoder. Figure 4 shows an example of threshold determination. The figure shows a plot of the variance of the MPEG-2 DC coefficient of 8x8 blocks vs. the MB mode. For this particular example, majority of the blocks are codes as INTRA4x4. The range of variance of the MBs for INTRA4x4 and INTRA16x16 overlap as shown in the figure; i.e., there are certain number of MBs in a frame with variances that can fall in to the 16x16 or 4x4 class. A threshold is selected such that the mismatches in each MB coding mode are minimized; i.e., the selected threshold results in fewest number of incorrectly classified MBs.

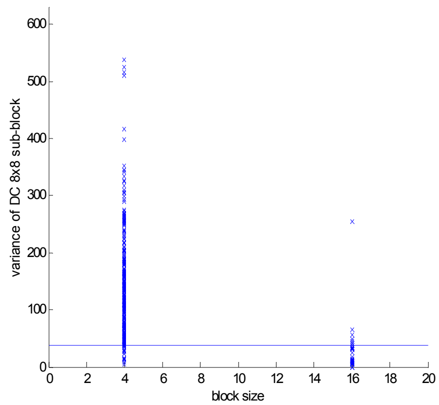

Fig. 4. Variance threshold to classify an Intra MB

A. Selecting a Threshold Metric

The level of activity in a MB that influences the MB coding mode the most can be estimated using a number of different metrics. A metric has to be selected such that the number of correctly classified MBs is maximized. We have experimented with a several metrics to find a suitable metric for MB mode estimation. The metrics evaluated are:

Metric 1: variance of AC coefficients of 8x8 blocks.
Metric 2: spatial domain variance of 16x16 block.
Metric 3: variance of DC coefficients of 8x8 blocks.
Metric 7: variance of DC and AC coefficients of 8x8 blocks
Metric 9: maximum variance of 4 8x8 blocks.

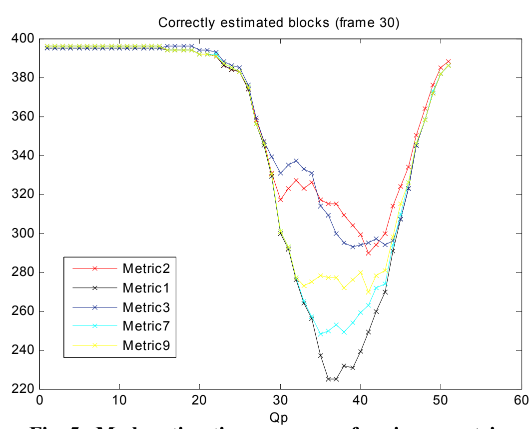

Fig. 5. Mode estimation accuracy of variance metrics

Metric 2, the variance of 16x16 pixel block computed in the spatial domain serves as a baseline. The goal is approximate this metric using DCT domain information in the incoming MPEG-2 compressed video. A threshold was selected for each of the metrics as described in the previous section and MB mode was determined based on the thresholds for each QP. Figure 5 shows the plot of correctly estimated MB modes across the QP range when using different metrics for the Football sequence at CIF resolution with 396 MBs per frame. As shown in Figure 5, Metric 3, the variance of the DC coefficients of 8x8 blocks performs the best and MB mode is estimated in the proposed transcoder based on this metric. Metric 3 also closely approximates the pixel domain variance which is a true indication of the activity in an MB. Using Metric 3 results in the classification of more than 75% of the MB modes correctly. Figure 6 shows the Metric 3 thresholds for the QP range. This QP-threshold function together with the variance of the DC coefficients of the 8x8 blocks in a MB are used to quickly determine the H.264 MB coding mode in the transcoder. The MB mode evaluation process is thus reduced to the computation of the variance of 4 integers followed by a threshold comparison.

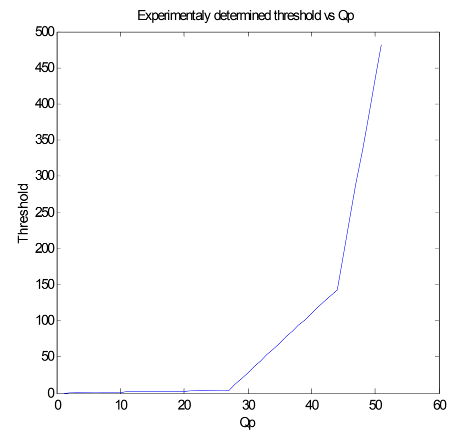

Fig. 6. QP-threshold function for Metric 3

The QP-threshold function shown in Figure 6 was used to compute the MB modes of the Flower, Foreman, and Football sequences at CIF resolution and encoded as MPEG-2 I frames only sequences using the default MPEG-2 intra quantizer. Figure 7 shows the number of correctly estimated MB modes across the QP range for frame 30 in all the sequences. For all cases, over 75% of the MBs are classified correctly. The accuracy of the mode estimation drops around QP 40 and this corresponds to the variance range overlap. The overall accuracy is always over 75%.

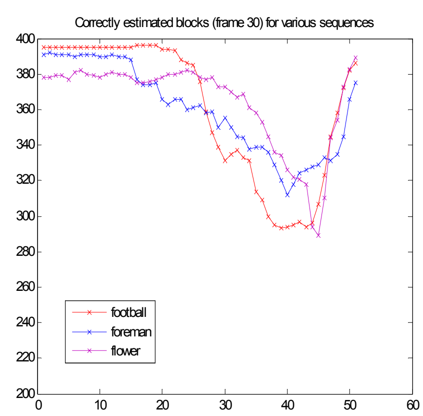

Fig. 7. Number of correctly estimated blocks in one frame using the threshold from Figure 6 for various sequences

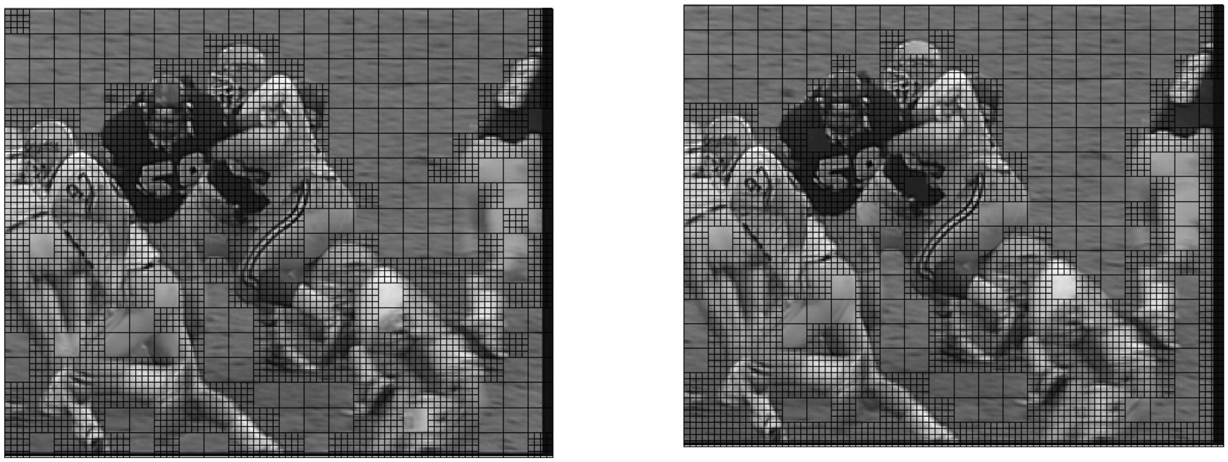

Fig. 8. MB Mode computation using the (a) reference transcoder and (b) the proposed transcoder

Figure 8 shows the MB mode decisions for the luma component of frame 30 of the Football sequence made using the reference transcoder and the proposed transcoder with a H.264 QP of 35. 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 costs, however, is substantial.

IV. Results and Discussion

The experiments are conducted with JM8.5 and the H.264 baseline profile. The times reported are for the transcoder running on a Pentium-4 PC at 2.6 GHz and 512 MB RAM. The executables are compiled with the Intel Compiler. 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 transcoder uses the default intra quantization matrix to quantize the DCT blocks. The RD optimization was disabled in the H.264 encoding stage. The H.264 encoder uses SAD to decide the optimal block size for intra MB coding. The reference transcoder determines the MB modes and MB prediction modes by performing an exhaustive evaluation of all the modes and prediction directions. The proposed transcoder first determines the MB coding mode based on the variance thresholds and then computes the prediction mode using the directional features of the MB.

The MB mode decisions are made based on a variance threshold that is a function of the QP used. The thresholds are based on the variance and are hence sequence independent. Figures 9-11 show the rate distortion curves for the reference and proposed transcoders with Football, Foreman, and Flower sequences. The proposed transcoder has an RD curve almost identical to the reference transcoder but with a substantial drop in computation. The average PSNR drop is less than 0.1 dB at the same bitrate.

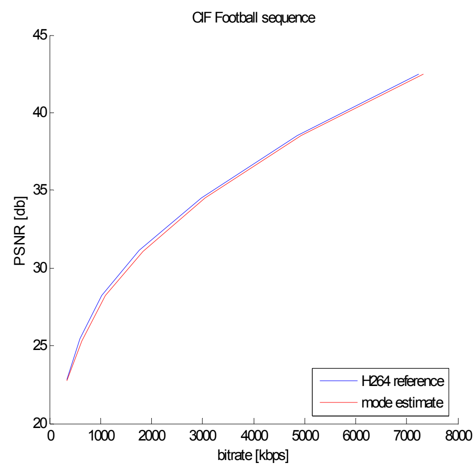

Fig. 9. RD performance of the reference and the proposed transcoder for the Football sequence

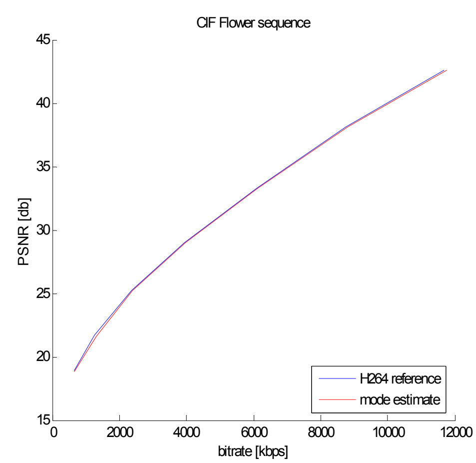

Fig. 10. RD performance of the reference and the proposed transcoder for the Flower sequence

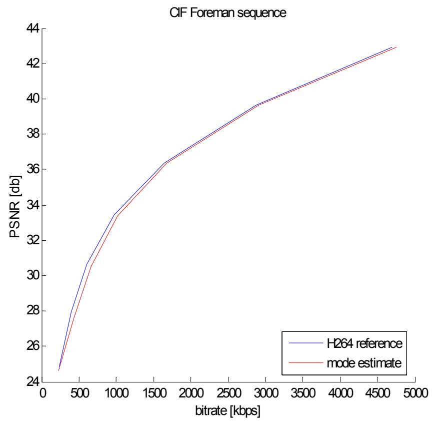

Fig. 11. RD performance of the reference and the proposed transcoder for the Foreman sequence

The MPEG-2 decoding stage is the same for the reference and proposed transcoder. For the H.264 transcoding stage, the proposed transcoder reduces the complexity by over 30% on average. This reduction is for the full H.264 encoding stage. The reduction for intra coding component is over 50%.

V. Conclusion

This paper presented algorithms and tools for MB mode estimation in MPEG-2 to H.264 intra MB transcoding. The proposed method reduces complexity by over 30% with negligible loss in RD performance. The complexity reduction is achieved by reducing the MB coding mode evaluation to a classification problem and performing the classification based on the variance of the DC coefficients in 8x8 intra blocks in MPEG-2. The proposed mode estimation resulted in over 75% match in MB mode estimation. Once the mode has been determined, the prediction mode is determined using the directional features extracted directly from the DCT coefficients. Our results indicate that the proposed transcoder results in an RD performance very close to the reference transcoder with substantial drop in the computational complexity.

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] Y. Su, J. Xin, A. Vetro, and H. Sun, “Efficient MPEG-2 to H.264/AVC intra transcoding in transform-domain,” IEEE International Symposium on Circuits and Systems, 2005. ISCAS 2005. pp. 1234- 1237 Vol. 2, 23-26 May 2005.

[5] M. Kucukgoz and M.T. Sun, "Early-stop and motion vector re-using for MPEG-2 to MPEG-4 transcoding," Proceedings of the SPIE, Visual Communications and Image Processing, Volume 5308, pp. 932-936 (2004).

[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 International Conference on Multimedia, 2004, pp. 300-303..

[7] 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.

[8] S.Ji and H.W. Park, “Region-Based Video Segmentation Using DCT coefficients,” Proceedings of the International Conference on Image Processing, ICIP 99, Vol. 2, pp. 150-154

[9] H. Wang, A. Divakaran, A. Vetro, S.-F. Chang, and H. Sun, “Survey of compressed-domain features used in audio-visual indexing and analysis,” Journal of Visual Communication and Image Representation, 14(2):50–183, June 2003.

[10] B. Shen and I.K. Sethi, “Direct feature extraction from compressed images”, Proc. SPIE Storage & Retrieval for Image and Video Databases IV, Vol.2670, 1996.

[11] S.-W. Lee, Y.-M. Kim, and S.W. Choi, “Fast Scene Change Detection using Direct Feature Extraction from MPEG Compressed Videos,” IEEE Transactions on Multimedia, Vol 2, No. 4, December 2000, pp. 240-254.

[12] K.R. Rao and P. Yip, Discrete Cosine Transform: Algorithms, Advantages, Applications, Academic Press, 1990, ISBN0-12-580203-X.

[13] F. Pan et. al, “A directional field based fast INTRA mode decision algorithm for H.264 video coding,” Proceedings of the IEEE International Conference on Multimedia & Expo, June 2004, Vol. 2, pp.1147-1150.

Hari Kalva (M’92-SM’05) Dr. Kalva joined the Department of Computer Science and Engineering at Florida Atlantic University as an Assistant Professor in August 2003. Prior to that he was a consultant with Mitsubishi Electric Research Labs, Cambridge, MA. He was a co-founder and the Vice President of Engineering of Flavor Software, a New York company founded in 1999, that developed MPEG-4 based solutions for the media and entertainment industry. Dr. Kalva is an expert on digital audio-visual communications systems with over 12 years of experience in multimedia research, development, and standardization. He has made key contributions to the MPEG-4 Systems standard and also contributed to the DAVIC standards development. His research interests include pervasive media delivery, content adaptation, video compression, and communication. He has over 30 published papers and four patents (six pending) to his credit. He is the author of one book and co-author of five book-chapters.

Dr. Kalva received a Ph.D. and an M.Phil. in Electrical Engineering from Columbia University in 2000 and 1999 respectively. He received an M.S. in Computer Engineering from Florida Atlantic University in 1994, and a B. Tech. in Electronics and Communications Engineering from N.B.K.R. Institute of Science and Technology, S.V. University, Tirupati, India in 1991.

Branko Petljanski (M’04) Branko Petljanski joined Imaging Technology Laboratory at Florida Atlantic University as a Research Engineer in August 2001. He has been working on designing high resolution cameras, displays and recording video systems. His research interests are video compression, video and image processing, and high speed signal processing. Branko Petljanski received an M.Sc. in Electrical Engineering from Florida Atlantic University, Boca Raton, USA in 2001, and a B.Sc. in Electronic and Communication Engineering from the University of Novi Sad, Novi Sad, Serbia and Montenegro in 1996. He is currently a Ph.D. student in the Department of Computer Science at Florida Atlantic University..