

AN EFFICIENT ALGORITHM FOR VC-1 TO H.264 VIDEO TRANSCODING IN PROGRESSIVE COMPRESSION

Jae-Beom Lee[†] and Hari Kalva[‡]

[†]Sarnoff Corporation, Princeton, NJ 08540

[‡]Dept. of Computer Science and Engineering, Florida Atlantic University, Boca Raton, FL 33431

ABSTRACT

The high definition video adoption has been growing rapidly for the last two years. The two high definition DVD formats HD-DVD and Blu-ray have mandated MPEG-2, H.264 and VC-1 as video compression formats. The coexistence of these different video coding standards creates a need for transcoding. In this paper, an efficient transcoding algorithm from VC-1 video to H.264 video is discussed. While there has been recent work on MPEG-2 to H.264 transcoding, the published work on VC-1 to H.264 transcoding is non-existent. There is very limited amount of published work on VC-1. This paper gives a brief overview of VC-1 and discusses the opportunities for low-complexity tools for VC-1 to H.264 transcoding. The paper considers I and P frame transcoding from VC-1 to H.264 and proposes two approaches 1) Low Cost Design and 2) High Cost Design.

1. INTRODUCTION

The high definition video adoption has been growing rapidly for the last two years. The two high definition DVD formats HD-DVD and Blu-ray have mandated MPEG-2, H.264 and VC-1 as video compression formats. The VC-1 standard is derived from Microsoft's proprietary WMV-9 which is widely used on the Internet. VC-1 is pure video compression technology developed by Microsoft, and is expected to be deployed as a key engine in satellite TV, IP set-tops and high-definition DVD recorders. H.264 is an emerging standard that is expected to replace MPEG-2 for digital video applications. The coexistence of these different video coding standards creates a need for transcoding. In this paper, an efficient transcoding algorithm from VC-1 video to H.264 video is discussed. For H.264, mode decision for Intra MB and Inter MB is computationally intensive since each of these modes has to be checked to select the best coding mode. The key idea of the paper is to propose early termination algorithm of forth-coming H.264 encoding based on incoming VC-1 encoded bitstreams. If early termination algorithm of mode checks takes advantage of already encoded VC-1 bitstreams, this will save a lot of computation in such MBs that would, otherwise, undergo all mode checks in the H.264 encoding stage.

The emergence of H.264 has resulted in increasing research in the area of transcoding to H.264 format. Transcoding tools and algorithms have been proposed to transcode video from H.263 [1], MPEG-4 [2], and MPEG-2 [3] to the H.264 format. The techniques proposed in the literature attempt to reuse the information gathered during the decoding stage to improve the transcoder performance. The transform domain approaches convert MPEG-2 DCT coefficients to H.264 transform coefficients [4,5]. These techniques, however, have limited applicability due to the fact that H.264 uses directional predictors in intra coding and determining the optimal predictor in the transform domain is computationally expensive. The use of variable block size motion estimation and deblocking filter also make transform domain transcoding impractical for full transcoding applications. Another approach is pixel domain transcoding: the MPEG-2 video is fully decoded followed by an accelerated H.264 encoding stage that uses information gathered during the decoding stage. This approach has shown promising results and is reported in several papers [6-8].

While there has been recent work on MPEG-2 to H.264 transcoding, the published work on VC-1 to H.264 transcoding is non-existent. There is very limited amount of published work on VC-1 [9] and no work has been published on VC-1 to H.264 transcoding. This paper gives a brief overview of VC-1 and discusses the opportunities for low-complexity tools for VC-1 to H.264 transcoding. The rest of the paper is organized as follows: section 2 gives an overview of the VC-1 coding tools for progressive video.

2. COMPRESSION TOOLS FOR VC-1 PROGRESSIVE VIDEO

This section provides a brief overview of VC-1 with emphasis on the features that impact transcoding. Like all MPEG standards, VC-1 is based on motion compensated transform coding. There is no fixed GOP structure in VC-1. And, I, P, B, BI and Skipped P are defined as pictures/frames. Unlike MPEG standards, I (Intra) frame does not have to occur periodically. Therefore, if there is no big scene change for a lengthy period of time, there could be only P frames in the sequence after the first I frame.

Unlike H.264, B frames cannot be used as reference frames. The BI frames are almost identical to I frames. A BI frame is Intra-coded frame and is used instead of a B frame when the B frame coding is inefficient. If there is big scene change continuously, some B frames cannot capture any similarity from two reference frames. In such cases, BI frame compression is a good choice. Since BI doesn't have the overhead for motion compensation, the syntax is optimized for such scenarios. A BI frame is not used as a reference frame. Skipped P frame is signaled when the frame is exactly the same as the previous reference.

In I frames of VC-1, no intra-prediction is used. For Intra-coded MBs (such as in I frames), only 8x8 transform size is used. For Inter-coded MBs (such as in P/ B frames), 4 transform sizes – 8x8, 4x8, 8x4, 4x4 – are potentially used on the residual data. Transform block size can change adaptively in P/ B frames with 4 different size options, while size of motion compensation is either 16x16 or 8x8 in VC-1. Note that this is quite the opposite to that of H.264. H.264 normally uses fixed size 4x4 transform with variable block size prediction for motion compensation. The transforms are 16 bit transforms where both the sums and the products of two 16 bit values produces results within 16 bits – the inverse transform can be implemented in 16 bit fixed point arithmetic. Note that the transform approximates a DCT, and norms of basis function between transforms are identical to enable the same quantization scheme through various transform types [10].

VC-1 supports a few options for motion compensation: 1) Half-pel or quarter-pel resolution motion compensation can be used. 2) Bi-cubic or bi-linear filter can be used for the interpolation. 3) 16x16 or 8x8 block size can be used. Only some combination of such options are defined to signal at the Frame level. Quantization is generally defined with two parameters in video standards – Qp and Dead-zone. The Qp varies from 1 to 31 as in MPEG, while there are two choices for Dead-zone in VC-1 – 3Qp and 5Qp. In I frames, PQAUNT is applied to all the MBs. However, DQUANT is used to adaptively describe Qp in each MB in P/B frames. Another Qp usage option is to use only two Qps for an entire frame depending on the MB positions – either boundary MB or non-boundary MB. There two techniques used in VC-1 to reduce blocky effects around transform boundary – Overlapped Transform (OLT) smoothing and In Loop deblocking Filtering (ILF). OLT and ILF are performed on reference frames I and P. Thus, the result of filtering affects only the quality of next pictures that use OLTed and/or ILFed frames as references.

3. VC-1 TO H.264 TRANSCODING

The transcoding algorithms discussed in this paper assume full VC-1 decoding down to the pixel level, followed by a reduced complexity H.264 encoding. The data gathered during the VC-1 decoding stage is used to accelerate the

H.264 encoding stage. We assume that the VC-1 encoded bitstreams were generated with an R-D optimized encoder. Table 1 shows a comparison of the VC and H.264 features from a transcoding point of view. The picture coding types used are similar. The transform size and type is different and makes transform domain transcoding prohibitively complex. The semantics of intra MBs are similar except for the intra prediction allowed in H.264 and the mixed inter MBs in VC-1. The inter prediction has significant differences including the block size of MC, block size of transform, and reference frames used. These similarities between the codecs can be exploited in reducing the transcoding complexity. We have developed two approaches 1) low cost design and 2) high cost design. These approaches are marked by the resources required for transcoding and represent quality vs. complexity tradeoffs.

TABLE 1: COMPARISON OF VC-1 AND H.264 FEATURES TRANSCODING

Feature	VC-1	H.264
Picture type	I, P, B, BI, Skip	I, P, B, SI, SP
Transform size	Adaptive	Fixed
Transform	Integer, similar to DCT—4 different transform sizes	Integer, similar to DCT – 4x4 and 8x8 transforms
Intra prediction	None	Directional predictors
Motion comp.	16x16, 8x8	7 variable block sizes
Reference frames	Max 2	Max 16

3.1 Low Cost Design

The low cost design performs direct MB mode mapping with the same reference frames used in VC-1. Picture type is also mapped to the same type picture. BI pictures can be mapped to B frames since the position of the frames are in B frames. In this paper, we focus on I and P frames' transcoding with the goal of transcoding VC-1 to H.264 at baseline profile.

3.1.1 Intra MB Mode Mapping

An intra MB in the incoming VC-1 bitstream is coded as H.264 intra MB. A VC-1 intra MB uses only 8x8 size transform. Based on uniformity of the incoming texture, either Intra 16x16 or Intra 4x4 modes can be mapped in output. The variance of the DC coefficients can be used to map to the H.264 mode as described in [3].

TABLE 2: VC-1 AND H.264 INTRA MB MODE MAPPING

VC-1 Intra Transform	H.264 Intra Mode
All 8x8 – uniform	Intra 16x16
All 8x8 – non-uniform	Intra 4x4

The inter MBs in the P pictures in VC-1 can have up to three 8x8 sub-blocks coded as intra. This is a mixed mode intra case and is mapped to intra MB in H.264 since the

H.264 intra mode has a reasonably good performance. Based on these observations, intra mapping is done as follows:

TABLE 3: VC-1 AND H.264 MIXED MB MODE MAPPING

VC-1 Inter MB (at least 1 8x8 Intra)	H.264 Intra Mode
3 8x8 Intra blocks	If (8x8 size transform used in Inter block): Intra 16x16 Else: Intra 4x4 + propagation
2 8x8 Intra blocks	If (8x8 size transform used in 2 Inter blocks): Intra 16x16 Else: Intra 4x4+ propagation
1 8x8 Intra block	If (8x8 size transform used in 3 Inter blocks): Intra 16x16 Else: Intra 4x4+ propagation

3.1.2 Prediction Mode Computation in Intra MB

Prediction mode estimation is based on estimating the directional features of the image blocks in the compressed domain. Since the VC-1 transform is an approximation of the DCT the directional features can be estimated in VC-1 using an approach we have developed for MPEG-2 to H.264 transcoding [2]. Figure 1 shows the edge model used for computing the directional feature. 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 [11]. The edge angle can be obtained in the DCT domain using equations (1). Since VC-1 transform is similar to DCT, a similar approach can be used to estimate the edge angle.

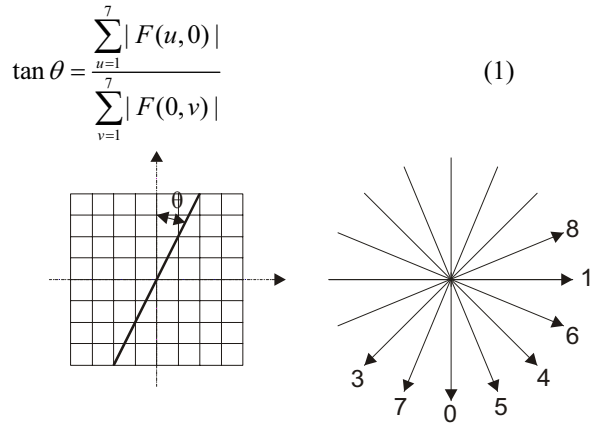


Figure 1. Edge orientation in an 8x8 block and intra prediction modes for 4x4 block

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. 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.

If all the 8x8 sub-blocks of a VC-1 MB use the 8x8 transform (when at least one of the 8x8 blocks is Intra-coded), the H.264 mode is selected as intra 16x16 and one prediction mode is determined for the MB. If one of the sub-blocks uses a smaller size transform in Inter-mode blocks, intra 4x4 is selected in H.264. However, for simplicity of the algorithm, the directional mode is propagated from adjacent Intra-mode direction. Figure 2 shows the prediction mode propagation that further reduces the complexity. Since Intra-coded 8x8 block in VC-1 bitstream corresponds to 4 4x4 blocks in H.264, a represented prediction mode is selected to propagate all other Inter coded area. If there are more than 2 intra 8x8 blocks and their directional modes are different, such a number of different directions is considered/ tested for the best coding mode.

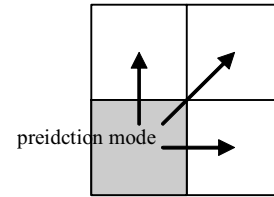


Figure 2. Coding mode decision in I/P with spatial neighborhood consideration for Inter blocks

3.1.3 Inter MB Mode Mapping

An inter coded MB in the incoming VC-1 bitstream is coded as inter MB in H.264. The inter MBs in VC-1 have 2 motion compensation modes – 1 MV mode and 4 MV mode. The 1MV mode is usually selected in VC-1 for areas that are relatively uniform. So, it is natural to choose 1MV mode in H.264 encoding – hence, the same MV of VC-1 is used for H.264 encoding. The 4MV mode is usually selected in VC-1 for areas that are non-uniform. The 16x16, 16x8, and 8x16 modes are eliminated for such non-uniform MBs. The MB is then mapped to one of the other variable block sizes allowed for H.264 motion compensation.

TABLE 4: VC-1 AND H.264 INTER MB MODE MAPPING

VC-1 Transform	H.264 Inter Mode
8x8	8x8
4x8	4x8
8x4	8x4
4x4	4x4

The variable block size for 8x8 sub-blocks of an MB is selected based on the transform size used in VC-1. Table 4 shows the proposed H.264 sub-partition modes based on VC-1 transform size. The shape of the transform used to code the MC residual in VC-1 indicates continuous regions and performing motion compensation with that block size is

likely to find a better match and improve the prediction. Once the MB coding mode is mapped, the next step is to determine the motion vectors for the MB. For inter 16x16 and the VC-1 4MV mode mapped to 4 8x8 blocks in H.264, the motion vectors are used without any refinement. For other block sizes, the VC-1 motion vector is used as a seed and the motion vectors are refined with a 5x5 refinement window. The reference frames used in VC-1 are also selected as references in H.264.

3.2 High Cost Design

The Low Cost Design performs quick mode mapping to reduce computational complexity while sacrificing the quality. When computational resources are available, the High Cost Design approach can be used to improve the quality of the transcoded video.

3.2.1 Intra MB Mode Mapping

The intra MB mode and prediction mode in I frames are determined using the same set of rules specified for the Low Cost Design. For inter frames, the intra MB of VC-1 are coded as intra in H.264. The mixed mode MBs in VC-1 are evaluated further. A mixed mode MB in VC-1 indicates that coding the MB either as inter 16x16 or inter 8x8 requires more bits and up to three 8x8 sub-blocks are coded as intra. The inter coded sub-blocks are mapped to H.264 based on the VC-1 transform shape and the intra coded sub-blocks are evaluated using motion estimation (8x4, 4x8, and 4x4). Furthermore, if 3 sub-blocks are intra, the entire MB is coded as intra. If one or two sub-blocks are intra, the cost is evaluated for sub-partitions and the MB is coded as inter.

TABLE 5: HIGH COST INTER MB MODE MAPPING

VC-1 Transform	H.264 Inter Mode
1 MV (Inter 16x16)	Best of 16x16, 8x16, and 16x8
4 MV (Inter 8x8)	
All 8x8 transforms	Best of 8x16, 16x8, 8x8
8x4, 4x8 transform	Best of 8x8, 8x4, 4x8, 4x4
4x4 transform	Best of 8x8, 4x4

3.2.2 Inter MB Mode Mapping

An inter 16x16 MB in VC-1 is evaluated for all the 16x16 sub-partitions (16x16, 16x8, 8x16) with the VC-1 motion vector and a search window of 5x5. The 8x8 sub-partitions are not considered since an inter coded 16x16 MB indicates that the 8x8 mode evaluated during VC-1 encoding was found to require more bits. Similarly, an inter 8x8 MB in VC-1 (4 MV) is not evaluated for 16x16 partition. The sub-partitions evaluated are selected based on the transform shape. If 8x8 transform is used, inter 16x8 and inter 8x16 also evaluated before the final decision. If smaller transforms are used, only the 8x8 sub-partitions are evaluated for the best match. In all these cases, the reference frames are limited to the references used in VC-1.

4. CONCLUSION

This paper addresses an important problem of transcoding VC-1 to H.264 coding format. Both VC-1 and H.264 are hybrid video coding algorithms that exploit motion compensation and transform coding. The VC-1 coded video has enough similarities to H.264 to enable reduced complexity transcoding. The paper proposes reducing the MB coding mode complexity in H.264 by exploiting the variable transform size used in VC-1. The quality vs. complexity tradeoffs are addressed using a low cost design and a high cost design for the transcoder. If early termination algorithm of mode checks takes advantage of already encoded VC-1 bitstreams, this will save a lot of computation in such MBs that would, otherwise, undergo all mode checks in the H.264 encoding stage.

5. REFERENCES

- [1] J. Bialkowski, A. Kaup, K. Illgner, "Fast transcoding of intra frames between H.263 and H.264," International Conference on Image Processing, 2004. ICIP '04. vol.4, pp. 2785- 2788.
- [2] Y.-K. Lee, S.-S. Lee, and Y.-L. Lee, "MPEG-4 to H.264 Transcoding using Macroblock Statistics," Proceedings of the ICME 2006, Toronto, Canada, July 2006.
- [3] 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.
- [4] C. Chen, P.-H. Wu, and H. Chen, "MPEG-2 to H.264 transcoding," Picture Coding Symposium, 15-17 Dec, 2004.
- [5] 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.
- [6] 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.
- [7] J. Xin, A. Vetro, S. Sekiguchi, and K. Sugimoto, "MPEG-2 to H.264/AVC Transcoding for Efficient Storage of Broadcast Video Bitstreams," International Conference on Consumer Electronics, 2006. pp. 417- 418, Jan. 2006.
- [8] G. Fernandez-Escribano, H. Kalva, P. Cuenca, and L. Orozco-Barbosa, "Speeding-up the Macro block Partition Mode Decision for MPEG-2 to H.264 transcoding," Proceeding of ICIP 2006, Atlanta, September 2006.
- [9] S. Srinivassan, P. Hsu, T. Holcomb, K. Mukerjee, S. Regunathan, B. Lin, J. Liang, M.-C. Lee and J. Ribas-Corbera, "WMV-9: Overview and Applications," Signal Processing Image Communication, Oct. 2004, pp851-875.
- [10] SMPTE, "Standard for Television: VC-1 Compressed Video Bitstream Format and Decoding Process," SMPTE 421M-2006.
- [11] K.R. Rao and P. Yip, Discrete Cosine Transform: Algorithms, Advantages, Applications, Academic Press, 1990, ISBN0-12-580203-X.