

# Reduced Resolution MPEG-2 to H.264 Transcoder

Hari Kalva<sup>1</sup>, Gerardo Fernández-Escribano<sup>2</sup>, and Kelly Kunzelmann<sup>1</sup>

<sup>1</sup>Dept. of Computer Science and Engineering, Florida Atlantic University, Boca Raton, FL, USA.

<sup>2</sup>Instituto de Investigación en Informática de Albacete, Universidad de Castilla-La Mancha, Campus Universitario, 02071 Albacete, SPAIN

## ABSTRACT

This paper describes complexity reduction in MPEG-2 to H.264 transcoding with resolution reduction. The methods developed are applicable to transcoding any DCT based video such as MPEG-2, MPEG-4, and H.263 to the recently standardized H.264 video at a reduced resolution. H.264 is being adopted by mobile device industry and devices such as iPod use H.264. The mobile devices, however, need the video at a reduced resolution. The proposed transcoder accelerates the H.264 encoding stage by performing motion estimation for only one block size as determined by the trained decision trees. Our solution allows conversion of MPEG-2 video to the H.264 video format at a reduced resolution with substantially less computing complexity. We use machine learning based approaches to significantly reduce the complexity of this transcoding. Experimental results show a reduction in transcoding time of about 67% (a 3x speedup) with a less than -0.5 dB change in PSNR.

**Keywords:** H.263, VP6, transcoding , complexity reduction

## 1. INTRODUCTION

The H.264 [1] standard is intended for use in a wide range of applications including high quality and high-bitrate digital video applications such as DVD and digital TV, based on MPEG-2 [2], and low bitrate applications such as video delivery to mobile devices. However, the computing and communication resources of the end user terminals make it impossible to use the same encoded video content for all applications. For example, the high bitrate video used for a digital TV broadcast cannot be used for streaming video to a mobile terminal. For delivery to mobile terminals, we need video content that is encoded at lower bitrate and lower resolution suitable for low-resource mobile terminals. Pre-encoding video at a few discrete bitrates leads to inefficiencies as the device capabilities vary and pre-encoding video bitstreams for all possible receiver capabilities is impossible. Furthermore, the receiver capabilities such as available CPU, available battery, and available bandwidth may vary during a session and a pre-encoded video stream cannot meet such dynamic needs. To make full use of the receiver capabilities and deliver video suitable for a receiver, video transcoding is necessary. A transcoder for such applications takes a high bitrate video as input and transcodes it to a lower bitrate and/or lower resolution video suitable for a mobile terminal.

The key contribution of the proposed approach is the use of machine learning based approach that enables reduction to multiple resolutions. The proposed approach exploits machine learning based transcoding techniques we have developed [3]. Using this approach, the MB mode of the output H.264 MB is determined using decision trees without evaluating any other modes. This approach can be applied multiple times with resolution reduces by 2 in each iteration. The rest of the paper is organized as follows. Section 2 describes the recent work in reduced resolution transcoding. Section 3 introduces the proposed reduced resolution algorithms. Results are presented in section 4 and conclusion in section 5.

## 2. RELATED RESEARCH

Several different approaches have been proposed in the literature. An MPEG-4 to H.264 transcoder was proposed in [4]. The authors present algorithms that exploit the motion vectors of MPEG-4 (Part-2) video to reduce the search range in H.264. This method also reduces the number of partitions evaluated by classifying the input MBs as low, medium, and high complexity. Predicted motion vectors (PMV) in H.264 are very good estimates for MVs and a small search range is usually sufficient. Furthermore, MPEG-4 video uses 8x8 blocks for motion estimation and includes richer information

---

This work was partly supported by the Ministry of Science and Technology of Spain under CONSOLIDER Project CSD2006-46, CICYT Project TIN2006-15516-C04-02, the Council of Science and Technology of Castilla-La Mancha under Project PAI06-0106 and FEDER.

compared to MPEG-2 data. Transform domain transcoding similar to the MPEG-2 to MPEG-4 reduced resolution transcoding proposed in [5] are not practical when transcoding MPEG-2 to H.264 as the two standards are completely different. We use pixel domain transcoding where input video is fully decoded and H.264 encoding stage is accelerated.

A motion re-estimation method for H.263 to H.264 transcoding with spatial down sampling was reported in [6]. The focus of this paper is motion vector reuse. While motion vector reuse is useful, most of the gains in transcoding come from MB mode selection. Another method for reduced resolution transcoding was proposed in [7]. In this approach, the authors scale MPEG-2 motion vectors to estimate 8x8 motion vectors. The distance between the motion vectors of these four 8x8 blocks is used to determine the partition size of the MB in H.264. This method becomes complex for reduction by more than a factor of 2.

### 3. SPATIAL RESOLUTION REDUCTION

The key problem in spatial resolution reduction is the H.264 MB mode determination. Instead of evaluating the cost of all the allowed modes and then selecting the best mode, the proposed algorithms directly determine the MB mode. The transcoding algorithms we have reported in our prior work transcode video at the same resolution [8]. While resolution reduction to any resolution is possible, reduction by multiples of 2 leads to optimal reuse of MB information from the decoding stage and gives the best performance. Resolution reduction by a factor of 2 in horizontal and vertical direction is discussed below.

Four MBs in the input video result in one MB in the output video. The coding mode in the reduced resolution can be determined using the MPEG-2 information from all the input MBs. An approach similar to our previous work on MPEG-2 to H.264 transcoding can be adapted here to determine the H.264 MB modes [8]. This approach, however, gives one H.264 mode for each MPEG-2 MB. For reduced resolution, we need one H.264 MB mode for four MPEG-2 MBs. Figure 1 shows an example of resolution reduction.

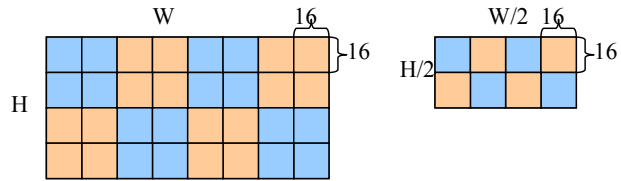


Figure 1. Resolution reduction by a factor of 2

Mode determination for the reduced resolution video can be performed in two ways: 1) use the information from four MPEG-2 MBs to determine single H.264 modes and 2) determine H.264 MB modes for each of the MPEG-2 MBs, and then determine one H.264 MB mode from four H.264 MB modes. The former approach is referred to Mode Reduction in the Input Domain (MRID) and the later approach is referred to as Mode Reduction in the Output Domain (MROD). Figure 2 shows the two approaches for resolution reduction in MPEG-2 to H.264 video transcoding.

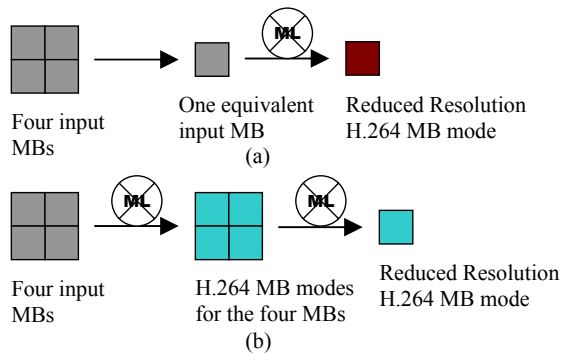


Figure 2. (a) MRID (b) MROD; ML is the machine learning process

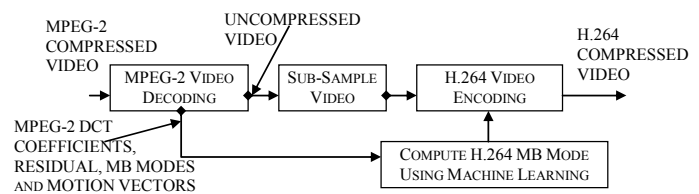


Figure 3. Reduced resolution transcoder architecture

#### 3.1. Mode Reduction in the Input Domain

Figure 3 shows the block diagram of the proposed pixel domain reduced resolution transcoder. The input video is decoded and MB information is collected for each MB. The decoded video is sub-sampled to the reduced resolution. The H.264 encoding stage is accelerated using the proposed MRID approach. The key idea of MRID is to reduce the MB information from the decoded MPEG-2 video (or other input video format) to the equivalent of one MB in the reduced

resolution and then determine the H.264 MB mode from the reduced input information. MB information from four input MBs is reduced to the equivalent of one input MB by scaling down the MPEG-2 residual data. A set of attributes is computed using the input MPEG-2 MB to characterize the H.264 MB mode. The attributes used are: mean of 4x4 sub-blocks and variance of the means of 4x4 sub-blocks that correspond to the various H.264 block sizes. These attributes are then used in a machine learning process to construct a decision tree for H.264 MB mode decisions. A similar approach to the use of machine learning in transcoding is reported in our prior work where transcoding was performed at the same resolution [3].

The main idea of this approach is to use Data Mining techniques for the Knowledge Discovery in Multimedia Databases. In the case that we are reporting, the databases are made with information gathered from the cascade transcoder, which performs the full process with training sequences (these training sequences are not used for testing the model). A large training set was created using training video sequences. Since mode decisions are made for each MB, each MB in a video represents an instance for training the model. The WEKA Data mining Tool [8] was used to train the model where each instance contains the attributes computed from the input MPEG-2 and a mode decision from the cascaded transcoder. The decision trees were created using C4.5 learning algorithm [9] and the resulting hierarchical mode decision tree built is shown in the figure 4. As shown in the tree, the H.264 MB mode is quickly computed using a set of if-else statements and the motion estimation is performed only for one block size there by reducing the complexity.

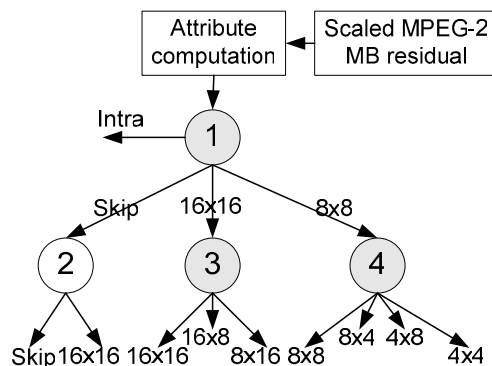


Figure 4. Decision tree used in transcoding

#### 4. RESULTS AND ANALYSIS

Simulations run for the following experiments use JM 10.2 H.264 [10] and MPEG-2 reference software from the MPEG Software Simulation Group. Each sequence evaluated was encoded with the following configuration: 6 Mbps input MPEG-2 for 4CIF (2 Mbps for CIF), 150 frames, GOP size of 12 with I and P frames only, and H.264 QPs of 24, 28, 32, 36, 40. Changes in time to transcode were measured in two parts; overall transcoding time, and time to complete the motion estimation process of the H.264 encoding stage.

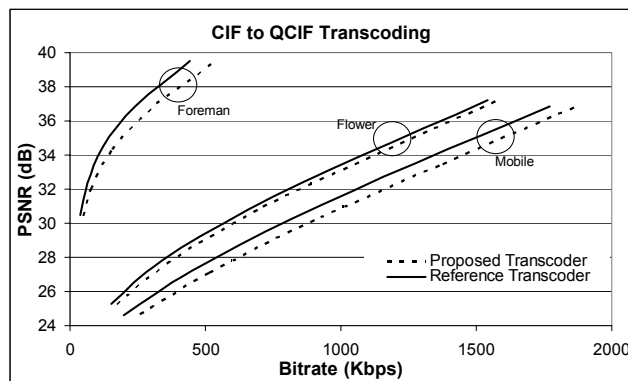


Figure 5. Performance of CIF to QCIF reduction

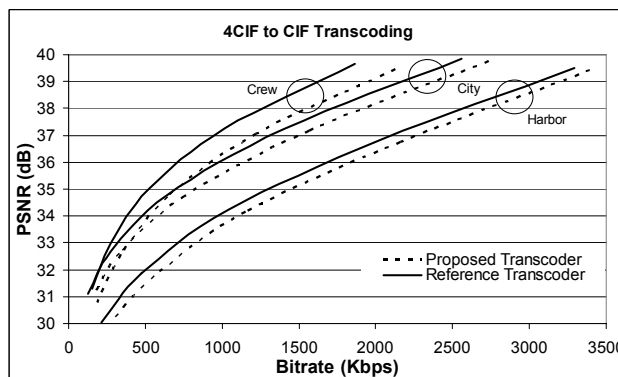


Figure 6. Performance of 4CIF to CIF reduction

##### 4.1. Video Quality and Bitrate

Changes to video quality for the proposed method were negligible and ranged in the -0.47 to -0.04 range. Selection of some sub-optimal macroblock modes through the machine learning process, however, did introduce an overall increase in bitrate. Figure 5 illustrates the rate distortion changes for the following common sequences: flower, foreman, and mobile. The RD performance of resolution reduction from 4CIF to CIF is shown in Figure 6. For practical bitrates, the RD curves show about .5 dB loss. PSNR drop is higher for Crew sequence possible because of poor mode estimation due to flashing in the sequences. at very low QPs (QP < 15). The RD performance is good across all resolutions and videos tested.

## 4.2. Transcoding Time

Overall gains in time to transcode averaged around 67% (i.e, the proposed method is about 3x faster than the reference implementation). As the proposed technique is an attempt to reduce the numbers of MB modes evaluated it follows then, that most of the reduction will be from reducing the cost of variable block motion estimation, a time consuming inter mode prediction process. This is shown experimentally by the fact that the motion estimation speed up is 9x on average. With RD optimization enabled, the proposed method is 5x faster with RD performance similar to RD\_Opt off case discussed above. Table 1 summarizes the performance results.

Table 1: Performance results for reduced resolution transcoding

CIF to QCIF Transcoding				4CIF to CIF Transcoding			
Sequence	$\Delta$ Y PSNR	$\Delta$ Time	$\Delta$ MEmTime	Sequence	$\Delta$ Y PSNR	$\Delta$ Time	$\Delta$ MEmTime
Mobile	-0.10 dB	-67%	-89%	Crew	-0.49 dB	-66%	-87%
Flower	-0.05 dB	-67%	-89%	City	-0.07 dB	-67%	-88%
Foreman	-0.20 dB	-65%	-89%	Harbor	-0.06 dB	-68%	-89%

The transcoder developed performs a resolution reduction by a factor of 2. For reducing the resolution by a factor of 4 (4CIF to QCIF), the same machine learning process can be employed. The decision tree would be applied once to reduce the resolution by 2 and then applied again to reduce the resolution by 2 again; resolution is reduced by 2 in each iteration. Mode reduction in the output domain that we are currently developing will work well for resolution reduction using this iterative approach.

## 5. CONCLUSIONS

It is intuitive that there is a stronger correlation between MPEG-2 and H.264 MB modes of similar resolution sequences than those of differing resolution. Machine learning is used to determine the MB modes in the reduced resolution space. Attributes of the input MPEG-2 video are used in the mode decision process. Results show that the proposed approach work well for reducing transcoding complexity. RD performance suffered slightly for the Crew sequence. This can be improved by improving the machine learning/training data. Preliminary exploration of the MRID method shows that this approach can improve the performance and reduce the complexity further as this method doesn't have to scale the MPEG-2 residual. The proposed methods reduce the complexity by over 67% (a 3x speedup) with negligible drop in PSNR.

## REFERENCES

1. ITU-T RECOMMENDATION H.264 "Advanced Video Coding for Generic Audiovisual Services". May 2003.
2. ISO/IEC 13818-2, "Generic Coding of Moving Picture and Associated Audio", MPEG-2 International Standard, 1994.
3. G. Fernandez-Escribano, H. Kalva, P. Cuenca, and L. Orozco-Barbosa, "Very Low Complexity MPEG-2 to H.264 Transcoding Using Machine Learning," Proceedings of the ACM Multimedia 2007, Santa Barbara, CA, October 2006, pp. 931-940.
4. T. D. Nguyen, G.-S. Lee, J.-Y. Chang, and H.-J. Cho, "Efficient MPEG-4 to H.264/AVC Transcoding with Spatial Downscaling," ETRI Journal, Volume 29, Number 6, December 2007, pp. 826-828.
5. A. Vetro, et. al, "Complexity-quality analysis of transcoding architectures for reduced spatial resolution," IEEE Transactions on Consumer Electronics, Vol. 48, Issue 3, AUG 2002, pp. 515- 521.
6. Y.-P. Tan and H. Sun, "Fast motion re-estimation for arbitrary downsizing video transcoding using H.264/AVC standard," IEEE Transactions on Consumer Electronics, Volume: 50, Issue: 3, Aug. 2004, pp. 887-889.
7. B. Hu, P. Zhang, Q. Huang, and W. Gao, "Reducing Spatial Resolution for MPEG-2 to H.264/AVC Transcoding. Advances in Multimedia Information Processing-PCM 2005: 6th Pacific-Rim Conference on Multimedia, Jeju Island, Korea, Nov. 2005, pp 830-840.
8. I. 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. 2004.