OVERVIEW OF LOW-COMPLEXITY VIDEO TRANSCODING FROM H.263 TO H.264

Jens Bialkowski, Marcus Barkowsky, and André Kaup

Chair of Multimedia Communications and Signal Processing,
University of Erlangen-Nuremberg, Cauerstr. 7, 91058 Erlangen, Germany
Email: {lastname}@LNT.de

ABSTRACT

With the standardization of H.264/AVC by ITU-T and ISO/IEC and the adaptatation into new hardware, the necessity of transcoding between existing standards and H.264 will arise to achieve interoperability between hardware devices. Because of the many new prediction parameters as well as the pixel-based deblocking filter and the new transform of H.264 this is a difficult task to perform.

In our work we propose a fast cascaded pixel-domain transcoder from H.263 to H.264 for both intra- and inter-frame coding. The rate-distortion (RD) performance of the encoded bitstreams is compared to an exhaustive full-search approach. Our approach leads to 9% higher data rate in average, but the computational complexity for the prediction can be reduced by 90% and more. It will be shown that the algorithms proposed for H.263 are applicable for transcoding MPEG-2 to H.264, too.

1. INTRODUCTION

Video transcoding is a technique to convert one bitstream into another. It is the key to transparent interoperability of end devices and supply networks. Homogeneous transcoding is used to change single parameters within the same standard, typically the data rate of the bitstream by requantization. In contrast, heterogeneous transcoding also includes the conversion from one compression standard to another compression standard for example from H.263 to H.264. A recent good overview for codecs based on the Discrete Cosine Transform (DCT) can be found in [1].

During the last decade the intensive research on block-based video compression led to the standardization of several video coding formats, for example the well known MPEG-2, MPEG-4 and H.263 standards. Although those standards differ in a small number of details they have many things in common, especially the process of motion compensation and the DCT. Due to similar motion compensation the motion vector (MV) can be reused very well [2]. Furthermore, the equivalent usage of the DCT of block size 8×8 makes a transcoder implementation within the DCT-domain possible [3].

With the standardization of H.264 the task of heterogeneous transcoding became much more difficult. This is an effect of the newly introduced integer-based transform (ICT) of size 4×4. Even if it is possible to approximate the coefficients from the DCT of size 8×8 [4, 5], a transcoder has further to cope with the non-linear deblocking filter defined in the pixel domain. The filtered image has to be passed back in order to avoid drift between transcoder and a successive decoder. This makes an all-frequency based (DCT-to-ICT) approach not desirable in terms of complexity. Another difficulty is the highly increased parameter range of H.264 compared to the DCT-based standards for intra- as well as for inter-predicted macroblocks (MB). For a good transcoding rate-distortion performance, reestimation of a number of coding parameters within a transcoder is required.

Table 1. Comparison of the coding standards H.263 and H.264

	H.263	**H.264**
	General	
Transformation	DCT, block size 8×8	Integer based transform; block size 4×4
Deblocking	Ann. J: inloop DCT-based; non-lin.	in-loop pixel based non-linear filtering across transform block edges
	Intra Macroblocks	
Prediction	Ann. I: 8×8 (4 Directions)	16×16 (Intra16; 4 Directions) 4×4 (Intra4; 9 Directions)
	DCT-domain	Pixel-Domain
	Inter Macroblocks	
MB Partitions	16×16	Free combination from 16×16 to 4×4 blocks including rectangular blocks
MV accuracy	half-pel, bilinearly filtered	half-pel using Wiener filter, quarter-pel bilinearly filtered
MV range	$[-16, 15.5]^2$	$[-1024, 1023.75] \times [-64, 63.75]$
References	1	up to 4 QCIF frames at lowest level

We presented our algorithms for intra transcoding between H.263 and H.264 in [6] and for inter-frame transcoding in [7]. This work gives an extended overview to both approaches. It extends these works by a more detailed insight into the choice of inter-frame coding decisions based on statistics evaluated using a full-search approach on pre-encoded video.

The work is organized as follows. In Sect. 2 we compare H.263 and H.264 in details for intra- and inter-coded frames with respect to the difference for transcoding. We draw consequences from this comparison and derive a proposal for a transcoder from H.263 to H.264, which is described in Sect. 3. The evaluations conducted will then be shown in Sect. 4. In Sect. 5 we explain, how our approach can be used for converting MPEG-2 to H.264, too. Finally, Sect. 6 will conclude the work.

2. COMPARISON OF H.263 AND H.264

Heterogeneous transcoders always have to convert the entropy coding. This step is unique in each standard and so it has to be performed. A detailed comparison of signal dependent coding parameters and side information between H.263 [8] and H.264 [9] is given in Tab. 1. It is easy to see that for each parameter of H.263 there exists a greater or equal set in H.264. This is also valid if taking Profile 3 of H.263 into account which is often implemented. The single categories of Tab. 1 are now discussed more exactly with respect to transcoding to H.264.

General

H.263 as well as H.264 are block-based video compression stan-

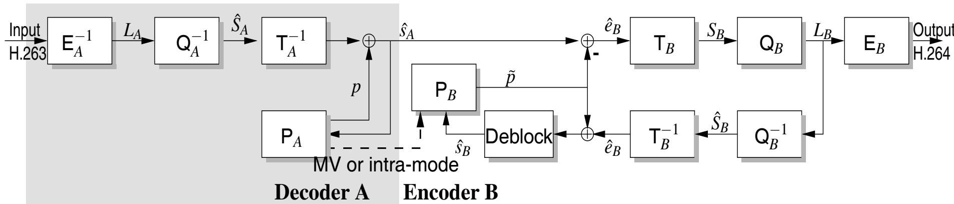


Fig. 1. Overview of the proposed cascaded pixel-domain transcoder

dards. This means that a video is divided into macroblocks and then each of them is encoded independently, possibly including prediction.

For each macroblock a transcoder has to convert coefficients from the DCT of block size 8 × 8 to the ICT of size 4 × 4 as defined in H.264. We have shown in [5] that requantization on coefficients of different transforms leads to quality losses of 3 dB PSNR at the same quantizer step size. This loss is decreasing if the output quantizer step size is larger than the input step size. An improvement for this problem has not been found so far, so any transcoded video will undergo the degradation.

Another transcoding problem is the inloop postprocessing in the pixel-domain by a non-linear deblocking filter defined in H.264. The performance gain of this filter is about 9% less rate at fixed quality [10] and so it is not recommended to switch it off. The signal difference of input and transcoded output due to the transform in combination with the deblocking filter has to be fed back for temporal prediction. Only then drift free transcoding is possible.

Intra-Coded Frames

Intra-prediction is defined only for H.264 and in Annex I of H.263, which is compulsory in most profiles, e.g. the Profile 3 (mobile). H.263 defines three directions for prediction within the DCT domain: DC, vertical or horizontal. Coefficients from the block to the left (horizontal), to the top (vertical) or both (DC) are used as predictor. This results in a horizontal, a vertical, or a flat prediction pattern. These basic patterns can be found in H.264, too. But the prediction 'Intra16' or 'Intra4' is performed using pixels from neighouring decoded blocks instead of coefficients. Furthermore, for Intra4 for each 4×4 block one out of 9 different directions can be signalled. The most probable patterns are those that are also defined in H.263. Therefore, the prediction directions of H.263 can be used as basis for H.264 at a transcoder.

Inter-Coded Frames

Inter-frames use blocks from a previously coded frame as predictor for the block to be coded. This principle is used both in H.263 and in H.264, but H.264 has a greater variety of coding options as shown in Tab.1. The most important differences are multiple references and flexible macroblock partitioning down to a size of 4×4 including rectangular partitions. Therefore, the motion vector fields of H.263 and H.264 are different. For our transcoding scenario this means that the coding decisions available from H.263 are not sufficient for H.264. For example, the approach to reuse motion vectors with successive half-pel refinement as known from the literature [1, 3] may not be sufficient. Instead, the performance of MV refinement has to be investigated again for transcoding from H.263 to H.264.

Another consequence of the extended parameter range of H.264 is the fact that often two different coding modes achieve a similar RD-result, e.g. multiple reference frames and quarter-pel vector accuracy. As a consequence, a transcoder does not have to check each single parameter available. Instead, the comparison of a selection of all possible parameters is sufficient. For example, we have found during our simulations that it is not necessary to use multiple references if using quarter-pel interpolation. But on the other hand in many cases testing the intra-prediction of H.264 against the inter-mode taken from H.263 results in a better RD-performance for the transcoded bitstream.

Also the direct mode of H.264 is very important to be used in a transcoder. For this mode the motion vector predictor is used for predicting a partition size of 16×16 and no coefficients are transmitted.

3. PROPOSED TRANSCODER FROM H.263 TO H.264

Cascaded Pixel-Domain Transcoder

From the observations made above we derive a transcoder model. Figure 1 shows a block diagram for the investigated cascaded pixel-domain transcoder. The input bitstream is fully decoded and side information of intra or inter coding parameters, such as intra directions or MVs, are transmitted to the prediction block P_B of the encoder B. An encoder B for the H.264 output bitstream is connected to the decoder A. The computational complexity of the prediction process P_B will be highly reduced by reusing the side information. Additionally selected advanced parameters of H.264 have to be compared to the reused side information in order to guarantee good RD-performance.

This approach is chosen by the authors because of two reasons: The first reason is the improved flexibility for the reencoder in terms of coding modes - inter or intra - as well as MV refinement. Here, the pixel-domain approach assures a drift free bitstream. The second reason is the deblocking filter that has to be performed in the pixel-domain. Thus, a possible complexity advantage of a frequency-domain approach is reduced.

The *reference* for our simulations is an unsupervised exhaustive full-search transcoder, which leads to the best performance possible. The input bitstream is also fully decoded there, but no side information is transmitted and all possible coding parameters are evaluated. Then, the parameter set with the best RD performance is encoded. However, this approach has the highest computational complexity.

Parameter Extraction for Intra-Frames

If information of intra prediction is found at the H.263 input bitstream, i.e. for a profile including Annex I, this information is used as preference in H.264. We presented an algorithm in [6] that decreased the transcoding complexity for the intra prediction process P_B of the H.264 encoder B. The absolute sum of the residual energy $\hat{Z}$ of the predicted coeffcients is used as measurement for determination a suitable intra mode of H.264. The value of $\hat{Z}$ is compared to a threshold D which is empirically determined and which is dependent on the input QP. If $\hat{Z}$ is below D for all four blocks of size 8×8, then Intra16 is used with the direction of H.263 without any further estimation. If the threshold was exceeded for at least one block of size 8×8, then Intra4 is used and blocks with high threshold are fully reestimated. We found that at high input quality about

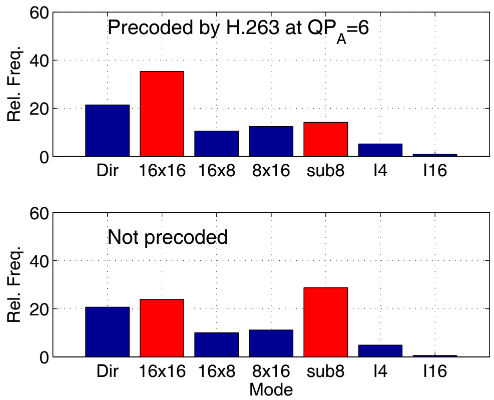


Fig. 2. Relative frequency of coding modes for reencoded data compared to encoding unprocessed data

Table 2. Average percentage of inter-modes for several QP_B chosen by the exhaustive search

	Direct	**16x16**	**16x8**	**8x16**	**sub 8x8**
$QP_A = 6$	9.4	43.0	11.1	13.1	15.2
$QP_A = 9$	14.5	47.1	9.8	12.1	8.7
$QP_A = 12$	17.9	51.4	7.9	9.9	5.3

30% of the blocks had to be reestimated and at low quality, i.e. using $QP_A > 10$, less than 15%. No Intra16 mode decisions have to be performed. The same algorithm was independently applied on the chrominance data.

Parameter Extraction for Inter-Frames

We evaluated the statistics for a number of sequences in order to find out, which inter coding parameters of H.264 are the most important ones. Fig. 2 shows the influence of the prequantization on the mode decision. For several sequences the average of the mode decisions when encoding unprocessed video (lower histogram) are compared to those of reencoded video (upper histogram). The input sequence of the reencoding was previously encoded by H.263 at high quality ($Q_A = 6$). But even for this small prequantization it can be seen that the mode decision for partition size 16×16 is highly increased while the mode 8×8 and its subpartions are used less. Our consequence is that these modes are of less importance especially for transcoding and thus we do not evaluate them. This approach is supported by the fact that partitions using one motion vector, i.e. 16×16 and Direct, are chosen more often when increasing the output QP. An example for this behaviour is given in Tab. 2. Direct mode and inter 16×16 mode use a 16×16 partition size and have altogether a proportion of over 50% for high quality input ($QP_A = 6$) and about 70% for medium quality input ($QP_A = 12$).

Considering these information we proposed an algorithm for transcoding inter coded MBs in [6]. We copy the motion vector of H.263 to the prediciton process P_B. This vector will then be refined by quarter-pel. As shown in [6], better results are possible only if the search range is increased over half-pel which leads to increased computational complexity. We also conducted a series of experiments on finding a suitable subpartition. This was done by performing a full-search on partitions smaller than 16×16. For the sequences investigated this resulted in a performance gain of less than 1% rate. A reason for this is the vector field used from H.263 which is not optimized for such a division and additionally the prequantization artifacts introduced by H.263. Furthermore, from Tab. 2 it can be seen that all modes apart from partitions with one motion vector (16×16, Direct) are used at very similar proportion of about 10%. This gives a hint that a selection of an appropriate partition is very difficult and thus results in high computational efforts. Instead, checking the best intra mode and the direct mode needs comparatively less computations but both modes increase the transcoding performance. The usage of the direct mode decreases the cost for inter-blocks, because no data has to be transmitted here. Using intra-prediction further improves the RD-performance much in cases where inter-prediction as reused from H.263 and then refined is not sufficient.

4. RD RESULTS AND COMPLEXITY

Our algorithms have been implemented into the reference software TM3.5 of H.263 and for H.264 into JM5.0g for intra-transcoding and JM7.3 for inter-transcoding. We conducted simulations on many sequences including the well known collection of the Video Quality Experts Group (VQEG), e.g. 'Mobile' (VQEG10) or 'Fastfood' (VQEG7). We encoded 220 frames of size Quarter Common Intermediate Format (QCIF) at 25 fps. A fixed quantization parameter (QP) is used for encoding the input and output bitstreams, which is independent on side effects of rate control selections and gives the highest reliablity regarding the proposed techniques. The relationship for input and output QP has been chosen such that the bitrate of the output bitstream was equal or slightly lower than the input bitrate. The output bitrate and the objective quality in dB PSNR are used as measurements here. Also, we compare the reduction of computational complexity in terms of SAD calculations rather than giving timing results here, because time measurements strongly depends on the implementation, the hardware used and the operating system as well and thus gives unreliable results.

Intra-Coded Frames

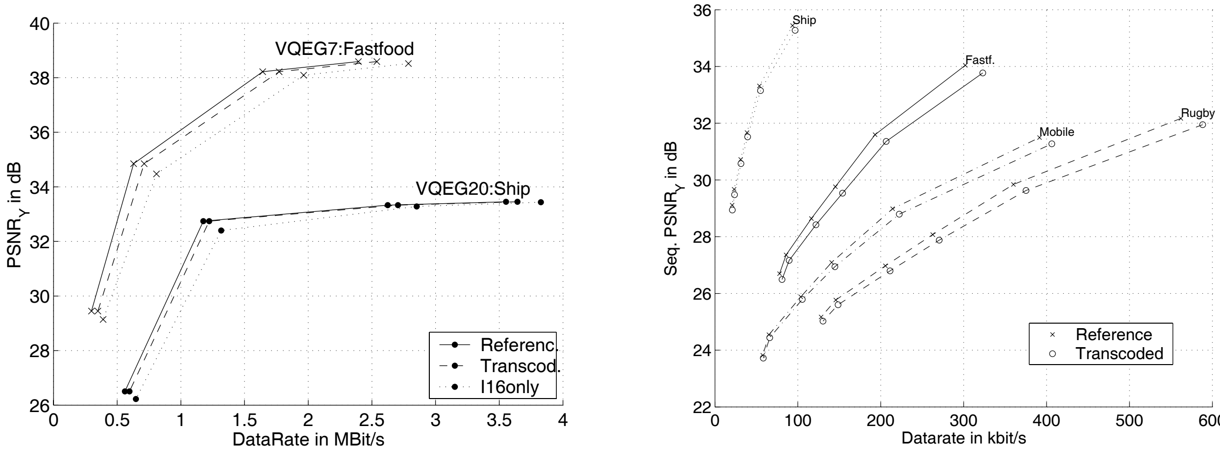
For intra coding the reduction of computational complexity for the mode estimation is between 75% at high input data rates and 92% for lower rates [6]. The performance of the proposed algorithm is shown in Fig. 3(a) for 'Fastfood' and 'Ship'. At equal objective quality the data rate of the proposed algorithm compared to the reference is increased by 7% or less for data rates over 500 kBit/s. Only at low data rates of less than 500 kBit/s the output rate of the proposal is about 10% higher than the reference. This can be explained by the fact that the DC prediction is transmitted relatively often in those cases. Furthermore, the DC residual used as cost is often low because of the high QP. This leads to the result that reestimation is performed only for a small number of MBs.

Inter-Coded Frames

For inter-coded frames the complexity reduction is a factor of approx. 200 [7]. It is the main contribution for complexity reduction in transcoders. This is because the reference uses an exhaustive full estimation over all search positions and modes possible resulting in the optimum RD-result at the highest complexity. The performance for several sequences is shown in Fig. 3(b). At input with high input data rates the output rate is increased by less than 10% in average compared to the full-search approach. At input data rates less than 300 kBit/s the output rate of the transcoder is less than 5% higher than the reference.

5. EXTENSION TO TRANSCODING FROM MPEG-2 TO H.264

We selected H.263 as a representative for DCT based coding standards. However, for progressive video input MPEG-2 [11] and H.263



(a) Intra-coded frames

(b) Inter-coded frames

Fig. 3. RD-performance of the proposed algorithms for transcoding (a) intra- and (b) inter-coded frames from H.263 to H.264

are different in only a few details. MPEG-2 has a larger motion vector range of $[-1024; 1023.5] \times [-64; 63.5]$ pixels which is nevertheless in the allowed range of H.264. No deblocking filter is used for MPEG-2. Because of this similarity of H.263 and MPEG-2, our proposed inter-transcoder can also be applied for transcoding from MPEG-2 to H.264. Additionally MPEG-2 does not support intra-prediction. So in the case of intra-frames or intra-macroblocks a transcoder to H.264 has to find an appropriate intra mode without preference.

6. CONCLUSIONS

Much research on transcoding between the DCT-based coding standards such as MPEG-2 or H.263 has been published so far. Transcoding within the DCT domain is state-of-the-art here. By giving a detailed comparison of these standards to H.264 we show why frequency domain transcoding from DCT-based standards to H.264 has no advantages. We selected H.263 as a representative for DCT-based standards. The algorithms published by the authors for transcoding intra- and inter-coded frames are discussed here. By exploiting side information of the input bitstream and evaluating selected modes of H.264 it is possible to decrease the computational complexity of the prediction process by 90% and more at the cost of 7% higher data rate compared to the full-search.

Our approaches are a first step for transcoding to H.264 and may be improved in the future by more detailed algorithms. Another interesting future research is the investigation of a QP based rate control at the transcoder. Because our algorithms are applicable over the full range of QP values it can be used very well for this kind of rate control methods.

Another comparison to be done is to compare our approach to a state-of-the-art low-complexity H.264 encoder which has a comparable complexity and also does not achieve the best RD-performance possible.

7. REFERENCES

[1] H. Sun A. Vetro, C. Christopoulos, "Video Transcoding Architectures and Techniques: An Overview," *IEEE Sig. Proc. Mag*, pp. 18–29, March 2003.

[2] N.Feamster and S.Wee, "An MPEG-2 to H.263 Transcoder," in *Proc. SPIE Int. Symp. on Voice, Video and Data Comm.*, September 1999.

[3] P. Assunção and M. Ghanbari, "A Frequency-Domain Video Transcoder for Dynamic Bit-Rate Reduction of MPEG-2 Bit Streams," *IEEE Trans. on Circuits and Systems Video Technol.*, vol. 8, pp. 953–967, 1998.

[4] Bo Shen, "From 8-tap to 4-tap Integer-Transform for MPEG to H.264/AVC Transcoding," in *Proc. Int. Conf. on Imag. Proc. (ICIP)*. IEEE, Oct 2004, pp. 115–118.

[5] J. Bialkowski, M.Barkowsky, and A. Kaup, "On Requantization in Intra-Frame Video Transcoding with Different Transform Block Sizes," in *Proc. Mult. Sig. Proc. Works.(MMSP'05)*, Nov. 2005.

[6] J. Bialkowski, A. Kaup, and K. Illgner, "Fast Transcoding of Intra Frames between H.263 and H.264," in *Proc. Int. Conf. on Imag. Proc.(ICIP)*, 2004.

[7] J. Bialkowski, M.Menden, M.Barkowsky, A. Kaup, and K. Illgner, "Fast Transcoding of Inter Macroblocks with Motion Vector Refinement from H.263 to H.264," in *Proc. Picture Coding Symposium.(PCS)*, Dec. 2004.

[8] G. Sullivan, Ed., *ITU-T Recommendation H.263: Transmission of Non-Telephone Signals; Video Coding for Low Bit Rate Communication*, ITU-T SG11, Jan. 1998.

[9] Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, *Study of Final Committee Draft of Joint Video Specification (ITU-T Rec. H.264; ISO/IEC 14496-10 AVC) (JVT F-100)*, ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, Dec. 2002.

[10] P. List et al., "Adaptive Deblocking Filter," *IEEE Trans. on Cir. and Sys. for Vid. Tech.*, vol. 13, no. 7, pp. 614–619, July 2003.

[11] *ISO/IEC 13818-2 (2nd Ed.): Information Technology - Generic Coding of Moving Pictures and Associated Audio- Video (MPEG-2)*, ISO/IEC JTC 1/SC 29/ WG11, Sydney, 2nd edition, 2000.