FAST TRANSCODING OF INTRA FRAMES BETWEEN H.263 AND H.264

Jens Bialkowski, André Kaup

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

Klaus Illgner

Siemens AG,
CT IC 2,
Otto-Hahn-Ring 6, 81730 Munich,
Germany
Klaus.Illgner@mchp.siemens.de

ABSTRACT

In this paper we present algorithms developed for transcoding predicted intra macroblocks between H.263 and H.264. It can be used in video servers for fast adaption of video content in networks supplying different coding standards, e.g. internet and UMTS. It is shown that due to basic pattern similarities between frequency domain prediction of H.263 and spatial prediction of H.263 it is possible to reuse predictor side information. This side information is used to simplify the mode and direction decision for intra prediction. It results in a high reduction of complexity for mode and direction estimation at comparable PSNR and a slightly increased rate.

1. INTRODUCTION

Video transcoding is a technique to convert one bitstream into another and there exist two different scenarios for this: homogeneous and inhomogeneous transcoding. Homogeneous transcoding reencodes bitstreams within the same standard but with different parameters, e.g. interlaced to progressive conversion of MPEG-2. In contrast, inhomogeneous transcoding extends this scenario to a conversion between different standards. This type of conversion is useful for adapting bitstreams to a networks infrastructure, where a great variety of devices require video data in different formats. One example may be an internet video server, that stores its data using H.264 but has to supply UMTS mobiles, where H.263 is compulsory. In this scenario, one device - server, router or gateway - between internet and UTMS has to adapt the bitstream from H.264 to H.263 and vice versa.

Usually inhomogeneous transcoding comes along with a variation of the parameter sets, such as filtering or deviating motion vector range. In the video standards MPEG-2 or MPEG-4, the main differences are found for temporally predicted macroblocks. Many approaches (e.g. [1][2][3]) exist to improve transcoding of inter macroblocks. Usually they reuse the motion vectors and apply vector refining and rate controlled requantization on those macroblocks in order to reduce complexity. Intra macroblocks are not considered, since they are usually encoded without prediction information. In contrast, when transcoding between H.263 and H.264 there exists intra prediction for both standards in most profiles. The main problem for transcoding here is, that H.264 predicts from neighbouring pixel values while H.263 defines the prediction process within the frequency domain, i.e. on DCT coefficients. In this contribution we propose algorithms that enable reusing side information for this problem in order to reduce complexity of the estimation processes for both standards.

This paper is organized as follows. Section 2 gives an insight of the intra prediction modes and directions defined in H.263 [4] and H.264[5]. The general description of the cascaded pixel-domain transcoder is given in section 3. Algorithm details and simulation results for transcoding from H.264 to H.263 are explained in section 4. Equivalently transcoding intra macroblocks from H.263 to H.264 is explained in section 5. Section 6 concludes the research.

2. DESCRIPTION OF INTRA PREDICTION TYPES

This section describes the defined prediction modes and directions according to [4] and [5] respectively. Then, the encoder strategy with the best RD-performance but highest complexity is given. This strategy is used as performance comparison. For the new algorithm, it is simplified by reusing data from the input stream.

2.1. Intra Prediction in H.263

One single intra prediction mode is defined within frequency domain for H.263 in Annex I of [4] that is supported by all profiles but profile 0. One of three directions - vertical, horizontal or DC - is encoded for each macroblock. This direction is used for all 4 luminance and both chrominance blocks equivalently.

An optimal encoder has to transform each macroblock using the 8×8 DCT and calculate the SAD for each direction, vertical, horizontal or DC. The best direction is selected and after entropy coding transmitted in addition to the residual coefficients. Each actually coded 8×8 block is directly stored as reference for the following blocks.

2.2. Intra Prediction in H.264

In contrast to H.263, where the predicted signal consists of DCT coefficients, the prediction process in H.264 is defined within the pixel domain. It is compulsory for all H.264 compliant bitstreams according to [5]. For each macroblock, one prediction mode - Intra4 or Intra16 - which defines partitioning of the macroblock is transmitted. For each mode and each partition one of several prediction directions is transmitted as well. In particular, there exist 4 prediction directions for Intra16 and 9 directions for Intra4. While for Intra16 only one direction for Luma is encoded, it is necessary

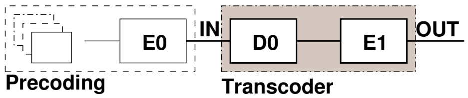

Fig. 1. Cascaded pixel domain transcoder

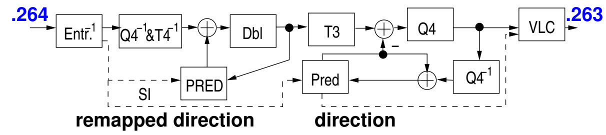

Fig. 2. Concept of pixel-domain transcoder from H.264 to H.263

to transmit 16 directions for Intra4. Chrominance is encoded in the same way as Intra16.

An encoder has to select the best combination of modes and directions using an RD-optimized algorithm for best performance. This involves, that for each existing direction of each mode the predictor within the pixel-domain is created from boundary pixels of the current partition and RD costs are calculated. The best combination of mode and directions is selected according to the minimum RD cost. The residual is encoded using a 4×4 integer based transform.

2.3. Discussion of Differences and Similarities

Although there are remarkable differences for the prediciton process, it will be shown, that reusing side information is possible.

The most significant difference between both standards is the prediction domain. In H.263 prediction is defined over coefficients of the DCT domain while H.264 uses pixel domain prediction. However, there is some similarity between the basic patterns of the directions which are defined for both standards. These patterns are vertical or horizontal stripes or a plane of fixed amplitude for DC prediction. Especially for the coded mode the similarity of the pattern is high. In consequence, the coded direction can be used as estimate for reencoding the bitstream.

The second difference is the capability for partitioning macroblocks in H.264 compared to the single mode of H.263. The Intra4 mode with 16 directions is often chosen when encoding with small quantisation parameter (QP) and brings highly improved RD performance for this case. In contrast, H.263 is not capable of a finer division and only one direction per macroblock is transmitted. For transcoding from H.263 to H.264 this means, that the capability for Intra4 must be provided, while for transcoding from H.264 to H.263 the 16 modes must be combined to one for the other direction.

3. CASCADED TRANSCODER

For the researches presented in this work, a cascaded pixel-domain transcoder is used. Two scenarios are considered: transcoding intra prediction from H.263 to H.264 and vice versa.

3.1. Cascaded Pixel-Domain Transcoder

A cascaded pixel-domain transcoder as shown on the right hand side of Fig. 1 is used here. For simulations, the video data has been coded once within stage 0 marked with 'E0'. This standard bitstream is sent to the transcoder which consists of a full decoder 'D0' and an encoder 'E1' for the new standard. The RD reference is an RD optimised full search which used only QP information of 'D0' and the decoded frame data as input, but no further side information. In contrast, for the complexity reduced algorithms side information used in 'D0' was additionally given to 'E1'.

Table 1. Mapping for prediction directions from H.264 Intra4 and Intra16 to H.263

H.264 Intra16	H.264 Intra4	H.263 output
DC	DC	DC
Plane	Horizontal-Down	
	Vertical-Right	
	Diagonal-Down-Right	
Vertical	Vertical	Vertical
	Vertical-Left	
	Diagonal-Down-Left	
Horizontal	Horizontal	Horizontal
	Horizontal-Up	

3.2. Necessity of Full Decoding

In the past, DCT domain transcoders have been presented by various authors ([1][2][3]). This is not applicable for drift-less transcoders in our scenario for several reasons. The main reason is the different prediciton domain which requires conversion from DCT to pixel-domain. Furthermore, H.264 defines a non-linear deblocking filter, which cannot be implemented efficiently in the transform domain, since the filter values are set on a pixel-by-pixel rule. At least, there is no algorithm known so far with small effort to convert between the 8×8 DCT domain of H.263 and the 4×4 integer based transform domain of H.264. As a result, for the developed drift-free intra prediction transcoder between H.263 and H.264, fully decoding is necessary and only the estimation process is accelerated.

4. TRANSCODING INTRA FROM H.264 TO H.263

Transcoding intra macroblocks from H.264 to H.263 within pixel-domain can be described by cascading a H.264 decoder followed by a H.263 encoder. A detailed system insight is shown in Fig. 2. For an incoming bitstream to the H.264 decoder entropy coding is reverted. Using the prediction mode and direction, the predictor is constructed using previously decoded neighbouring blocks. The reconstructed error residual is added in pixel domain and filtered. The resulting signal is stored as reference for following blocks in the pixel-domain. For transcoding, the direction is remapped to one H.263 direction depending on the mode and direction(s) used in H.264:

- Given Intra16 mode, the encoding direction for H.263 is directly reused after a remapping according to Tab. 1.
- Given Intra4 mode, each of the nine H.264 directions are mapped according to Tab. 1.
  A histogram for the remapped directions is calculated after remapping, using each of the 16 directions of Intra4. Then, the most often counted direction is used for encoding in H.263. If two or more directions have the same number of occurrences, the uppermost mode of the right column of Tab. 1 is chosen.

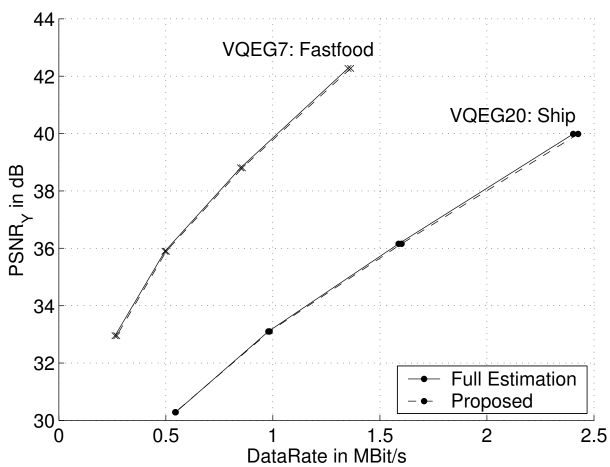

Fig. 3. RD results for transcoding intra macroblocks from H.264 to H.263

Then, the image signal and the remapped direction is sent to the successive H.263 encoder. For encoding no reestimation is processed in case of fast transcoding. Instead the remapped direction taken from H.264 is used directly and the predictor is constructed accordingly using DCT coefficients. The residual is calculated, quantized and entropy coded and written into the H.263 bitstream.

4.1. Complexity Reduction

Since transcoding Intra16 is effortless, the lower limit on complexity reduction is given when transcoding Intra4. Building the direction histogram needs 16 additions, while remapping is performed directly at generation of the predictor and needs no computation. The decision of the best representative out of three (horizontal, vertical, DC) modes requires 2 additional comparisons.

The full direction search has to compute the prediction signal and the SAD of 15 coefficients for all 6 8×8 blocks per macroblock (Y/C$_b$/C$_r$) and select the best. This affords 354 additions, 24 shifts and 2 comparisons. Altogether, the upper bound for computational complexity is reduced to 4.23% of a full search with equivalent comparisons.

4.2. Rate Distortion Results

For evaluating the algorithms performance, several video sequences were precoded with the reference software of H.264 [6] and afterwards transcoded to H.263. We implemented the algorithms into [7]. Every 30th frame of each sequence has been encoded with fixed QP as intra and only for these frames the average PSNR$_Y$ (in dB) and data rate (in MBit/s) were measured. The QP of the H.263 encoder has been empirically adjusted such that input and output data rate lay in the same range. An overview over the simulation results at various bitrates is given in Fig. 3. Notice, that the data rate is only slightly increased for the fast transcoding scheme compared to the full search, while the PSNR$_Y$ is equivalent for all simulated QP. In spite of the suboptimal choice of intra prediction direction, the proposed algorithm increases the data rate only very little compared to a full search. As a conclusion, the high complexity reduction at very small rate losses makes the proposed algorithm an interesting alternative to the full search.

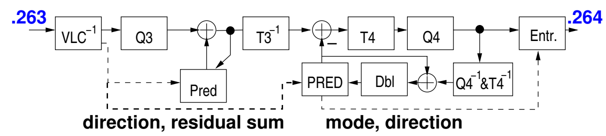

Fig. 4. Concept of pixel-domain transcoder from H.263 to H.264

5. TRANSCODING INTRA FROM H.263 TO H.264

Figure 4 shows the detailed overview of the transcoder algorithm. At the decoder stage, an incoming bitstream is parsed in order to reverse entropy coding. For each intra macroblock the extracted prediction direction is used to form the predictor within DCT domain and the requantized residual coefficients are added. This signal is used as reference for following blocks and is transformed to the pixel domain and transmitted to the H.264 intra macroblock encoder. For the proposed transcoder, the direction and the residual sum of the intra block is given to the prediction stage 'PRED' for threshold controlled simplification of the estimation process. More exactly, the direction is used as estimation for the H.264 direction, whereas the decision of the mode Intra4 or Intra16 depends on comparsion of a threshold T and the error residual sum. The exact algorithm works as follows:

1. For each 8×8 block i calculate the error residual sum R_i at the predicted coefficients c_j (depending on direction) as $R_i = \sum_{j \in \text{modei}} c_j$
2. For all 4 **luminance blocks** ($i \in \{1 \dots 4\}$): If all $R_i < T, i \in \{1 \dots 4\}$ use Intra16 mode, else use Intra4 mode:
   - For Intra16: Do not reestimate, use direction of H.263 for encoding directly instead
   - For Intra4: Reestimate directions of all 4 4×4 partitions of block i only if $P_i > T$, else use the direction of H.263 directly
3. For **chrominance** ($i \in \{5, 6\}$) blocks: If $R_5 > T$ or $R_6 > T$ reestimate chrominance prediction, else use direction of H.263
4. For chrominance and luminance: If the direction is not available in H.264 because of missing reference pixels, reestimation is used instead.

The idea behind this algorithm is to enable the enhanced partitioning as well as the greater variety of directions supplied by H.264 for those macroblocks, where the prediction of H.263 was unsatisfying according to T. The used measure is the residual sum. A small residual sum is equivalent to a high similarity of the predicted block and its predictor and is considered a good estimation for H.264 prediction. For blocks with high residual sum it is assumed, that there may be better directions for H.264 and therefore those blocks are reestimated. In this work, the reestimation has been done by a full search at the related 4×4 partitions of an 8×8 block, but it may be done by any suboptimal algorithm instead. Using this mode and direction setup, the error residual is calculated, transformed and - after entropy coding - written into the H.264 output bitstream.

Simulations have shown, that the main impact on performance loss is due to a wrong mode decision which strongly depends on the QP. To adapt this QP dependent selection we introduced an

Table 2. Thresholds used for simulations

QP	9	27	35
T_{AC}	91	114	170
T_{DC}	4	5	8

Table 3. Average number of reestimated blocks in % for different QP

	Full Est.	**New Algorithm**			
QP	all	9	15	27	35
	100	34	33	25	14

exponential mapping of T as function of QP according to the optimum Lagrange RD mode selection in [6]. For further improvements we introduced a weighting factor for the threshold of DC prediction, because for low QP values the DC prediction of the pixel-domain can differ strongly from the prediction within DCT-domain.

5.1. Rate-Distortion Results and Complexity Reduction

The proposed algorithm has been implemented into the reference software of H.264 [6]. Simulations with equivalent parameters as in section 4.2 have been made. The results for some sequences are shown in figure 5. For each sequence a full mode and direction search ('Full Est.') is compared to the proposed algorithm ('Prop.'). Additionally one simplified mode 'I16only' is shown. For this scheme, the mode of H.263 has been used directly without further reestimations. First the proposed scheme is compared to the full estimation scheme. Here, the PSNR is only slightly decreased but due to the suboptimal mode and direction choice, the data rate is slightly increased. Depending on the output bitrate, we measured an increment between 5% at high data rates and 16% at low data rates. Further, the simplified simulation 'I16only' increases the rate by 10 to 30%, which originates from the restricted mode selection. This is one evidence, that the wrong mode decision leads to a remarkable rate increment.

The average number of reestimated blocks of the sequences is given in Tab. 3. The left column is the reference RD encoder and requires 100% blocks to be estimated. The columns to the right show the number of reestimated blocks for the simulated QP. For high QP the threshold values are higher and thus more blocks are reestimated as for low QP. This value also depends on the image content.

The performance of the proposed algorithm is closer to the full estimation than to the simplified, which justifies the increased complexity for partial reestimation. This result was also valid for other simplified schemes such as Intra4 only.

6. CONCLUSIONS

Two algorithms for intra prediction transcoding have been proposed in this work. The first is designed for transcoding from H.264 to H.263. It is easy to use and reduces the computational complexity to 4.2% of a full direction search. It operates in a suboptimal way, but increases the data rate insignificantly at equivalent PSNR.

The second algorithm transcodes intra macroblocks from H.263 to H.264 and is dependent on image content. It is suited for fast

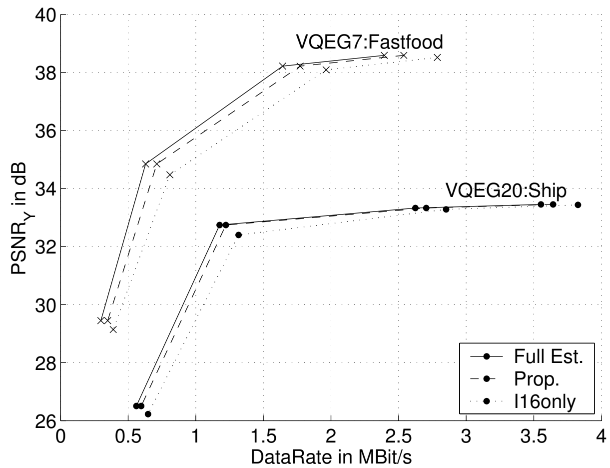

Fig. 5. RD-Results for transcoding intra macroblocks from H.264 to H.263

transcoder implementations because of its complexity reduction, even if the data rate is increased between 5 and 16% at equivalent PSNR.

The proposed algorithms can be used as basis for a full low complexity transcoder, since they are applicable in the full QP range. Furthermore, they are adaptable for any QP based rate control mechanism. The principle of reusing side information for fast transcoding can also be applied for inter frame coded macroblocks, which is currently under investigation by the authors.

7. REFERENCES

[1] K. Seo, S. Heo, and J. Kim, "Adaptive Rate Control Algorithm Based on Logarithmic R-Q Model for MPEG-1 to MPEG-4 Transcoding," *Sig. Proc.:Imag. Comm. 17*, pp. 857–875, 2002.

[2] T. Shanableh and M. Ghanbari, "Transcoding Architecture for DCT-Domain Heterogeneous Video Transcoding," in *Proc. Int. Conf. on Imag. Proc. (ICIP)*, 2001, pp. 433–436.

[3] S. Jang and N. Jayant, "An Adaptive Non-Linear Motion Vector Resampling Algorithm for Down-Scaling Video Transcoding," in *Int. Conf. on Mult. and Expo (ICME)*, 2003.

[4] 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.

[5] Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, *Study of Final Comitee 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.

[6] Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, *Reference Software to Commitee Draft JVT-F100 JM50g*, ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6, Feb. 2003.

[7] ITU-T SG11, *H.263+ Public Domain Codec (TMN 3.2)*, University of British Columbia, http://spmg.ubc.ca/h263plus, 3.2 edition, 1998.