TRANSCODER ARCHITECTURES FOR VIDEO CODING

Niklas Björk and Charilaos Christopoulos

Ericsson Telecom AB.
HF/ETX/PN/XML, CLAB
126 25 Stockholm, Sweden
Email: {niklas, ch.christopoulos}@clab.ericsson.se

ABSTRACT

Two different models for transcoding of H.263-based video streams are examined: rate reduction and resolution reduction. Results show that the computational complexity of the basic transcoding model can be reduced for each model by an average of 39% and 23% with less than 1 dB loss in quality for sequences with high motion. Comparisons with scalable video coding model are also presented.

1. INTRODUCTION

When video signals are encoded, the channel characteristics through which the resulting bitstream will be transmitted might be unknown. This is particular true when the same bit stream is expected to be distributed to several decoders using channels with different characteristics. Since the encoding needs to know the channel characteristics, these have to be assumed as known parameters to the encoder when it is running. This is a common problem in a multipoint video conference where the bandwidth of each participants is different. Normally, such a video conference will be accomplished through a multipoint control unit (MCU) and appropriate gateways. A typical example is multipoint communication where POTS, ISDN and ADSL lines are used. In such a situation, the common bitrate to be used would be the one used from the POTS subscriber, i.e. around 28.8 Kb/s in most environments. To alleviate this problem, we investigate various video transcoder architectures which will be located in an MCU/gateway and scale the incoming bitstream according to the requirements of the receivers. Video is supposed to be encoded using the H.263 standard [4,5].

Basically, the transcoder itself consists of a cascaded decoder and encoder [1] (figure 1). The incoming bit stream at bitrate R_1 is decoded by a variable length decoder (VLD), which gives the value of the quantized coefficients and the motion vectors. Next these coefficients are de-quantized (Q_1^{-1}) and transformed by an inverse Discrete Cosine Transform (IDCT). These operations yield the signal Δx_n of the picture X_n. The picture X_n is then reconstructed by adding the motion-compensated previously decoded picture, X_{n-1}.

After decoding the incoming signal, the encoder follows. As seen in figure 1, the first step is to subtract the prediction P_n from the picture X_n, which gives the residual signal ΔY_n. This signal is transformed using a DCT and further quantized (Q_2). The quantized DCT coefficients are put through a variable length coder and the bitstream at bitrate R_2 is transmitted at the encoder. In figure 1 E_n denotes the quantization error introduced at the encoder.

2. INVESTIGATED MODES OF TRANSCODING

Two different modes of transcoding are investigated. These two modes are rate reduction and resolution reduction.

2.1 Rate reduction

Rate reduction implies that R_2 is less than R_1. A simple way to achieve it is to increase the quantization step Q_2, at the encoder present in the transcoder. To reduce the computational complexity, motion vector information and macroblock coding type information can be passed directly to the encoder (of the transcoder) and need not to be evaluated. This is shown in figure 2 (please ignore the downsampling filter for the moment). However, when passing motion vectors (MV's) and macroblock (MB's) information from the decoder to the encoder some problems arise:

a. The passed MV's, coming from the decoder (figure 2), which were computed at the transmitter for quantization factor $Q_{transmitter}$, may be not suitable due to different quantization ($Q_{transmitter} \neq Q_2$).

b. MB's might be coded in the wrong mode. For instance, a MB that should be coded in a SKIPPED mode at the encoder of the transcoder, due to larger quantization making all coefficients zero, could be coded as an INTER MB since it was coded as an INTER MB at the transmitter.

The solutions to these problems could be:

- Since the passed MV's will almost be the same as the recalculated ones, refine them. The refinement can be done in a small search window around the motion vector passed. Results show that it is enough to do this refinement on nearby pixels, and at most in a search window of size [-3,+3] (in most cases [-1,+1] gives satisfying results). Fast search methods can be applied to further reduce the computational complexity of this operation [7]. The search method that has been used in the our simulations is the Full Search with the Sum of Absolute Difference (SAD) matching criterion [2].
- For the MB type: if it was coded as INTRA (at the transmitter) again code it in INTRA; if it was coded as SKIPPED again code it as SKIPPED; if it was coded in INTER, check to see if all coefficients are zero and if they are, code it as SKIPPED; else check again whether the MB has to be coded in INTRA or INTER mode. Alternatively, the encoding type of the MB can be re-evaluated.

There are other methods of reducing the computational complexity of the transcoder and one is the simple re-quantization model [9]. However, although the computational complexity is decreased significantly with this method, it leads to drifting distortions because the decoded pictures are not those used as predictions at the encoder. In addition, we did not evaluate a method similar to the one proposed in [6], which is working in the DCT domain, since it is not suitable for resolution reduction transcoding. Furthermore, as it will be shown below, for good results during transcoding, the motion vectors have to be refined. This is not easily accomplished with the model of [6] which is working in the DCT domain and other suitable methods have to be investigated.

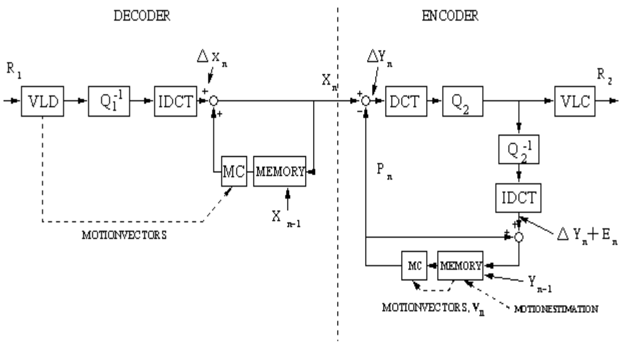

Figure 1 Outline of a basic video transcoder

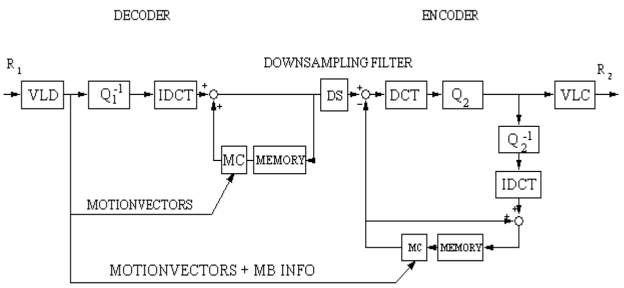

Figure 2 Outline of a transcoder where VM's are passed directly

2.2 Resolution reduction

By inserting a downsampling filter in the transcoder, as seen in figure 2 (denoted as DS), there is a possibility to perform resolution reduction in the incoming video. As in the case of rate reduction, problems arise when passing motion vectors and macroblock information directly from the decoder to the encoder. The problem is to evaluate the motion vector that will be used for a MB of the downsampled picture. This MB is produced from 4 MB of the incoming video (when resolution reduction by a factor of 2 in each dimension is performed). Therefore, the MV corresponding to this MB is produced from 4 MV. The motion vector has also to be scaled which will introduce additional errors. The same problem occurs for the macroblock coding type.

Possible solutions for the MV determination are:

- Calculate the average or median of four MV's, scale it (i.e. divide it by two) and pass to the encoder.
- Pick a single MV out of four, scale it (i.e. divide it by two) and pass to the encoder.

Since the solutions are sub optimal, a motion vector refinement might be needed at the encoder.

Concerning the MB type the following can be done:

- If there is at least one INTRA type among the 4 MB's then encode it as INTRA. Encode as INTER type if there is no INTRA MB and at least one INTER MB. Encode as SKIPPED if all MB's are of the SKIPPED type. Alternatively, the MB encoding types can be re-evaluated at the encoder.

3. EXPERIMENTAL RESULTS

The computational complexity of the transcoding operation (measured as the amount of processor time in a UNIX SUN SPARC platform), and the quality of the transcoded sequence (measured in PSNR), are used for the evaluation of the various transcoding schemes. The video sequences used for the simulations were coded according to the H.263 standard.

3.1 Rate reduction

The basic model where motion vectors and macroblock information are passed directly from the decoder to the encoder (figure 2) is compared with the direct coding method. When re-evaluation of the MV's is performed, the full search method using SAD as the matching criterion was used and the search area was [-7,+7].

Table 1 shows the results for the basic rate reduction model. In this table R_1 is the bitrate of the incoming bitstream, R_2 is the desired bit rate of the outgoing bit stream, B' is the size of the transcoded bit stream when passing MV and MB information, B is the size of the bit stream when the motion vectors and the macroblock type are being re-evaluated at the encoder of the transcoder. R_2 is the bitrate which is achieved when a certain quantization Q_2 and motion vector re-evaluation is performed with a full search method in a search window [-7,+7]. The value of Q_2 is then used in the transcoding modes that are investigated (i.e. when MV and MB information is passed to the encoder). R_2 denotes therefore the desired bitrate which the simplified models should approach. T_1 denotes the processor time that is needed when MB's types and MV's are re-evaluated and T_2 is the processor time that is needed for transcoding when passing MV and MB information. The values within the parentheses are valid for motion vector refinement.

Refinement was done by a full search method using the SAD matching criterion in a [-1,+1] search window for the 'Akiyo' and the 'Mother and daughter' (M&D) sequence and in a [-3,+3] search window for the 'Foreman' sequence. CIF resolution is used for 'Akiyo' and 'M&D' sequence and 30 frames/s, while QCIF is used for 'Foreman' and 15 fps. 150 frames were used for each sequence. The parameter *size* and *complexity* in Table 1 show the percentage of increase in bitrate and complexity reduction when the simplified model (passing MV's and MB types) is used compared to the re-evaluation of MV' and MB types (which give bitrate R_2).

As table 1 shows the computational complexity is improved, on average, 39 % by passing information from the decoder to the encoder and this with an average increase of 1.2 % bits in the outgoing bit stream. However, if refinement of the MV's is performed, the average increase in bitrate is 0.36% but the average gain in computational complexity is reduced by a factor of 2. This is due to the full search method used. This Full search method can be replaced by a fast algorithm which would significantly reduce the computational complexity without a noticeable increase in bit rate [2,7].

sequence	R_1 (kbits/s)	R_2 (kbits/s)	Size (%) (B'-B)* (100/B)	Complexi -ty (%) (T_1-T_2)* 100/T1
Akiyo	132	68	1.5(0.4)	36(18)
M &D	120	61	1.4(0.5)	37(18)
Foreman	136	68	0.7(0.2)	36(17)

Table 1 Rate reduction results, basic model

The quality of the transcoded video sequence can be seen in figure 3 for the 'Foreman' sequence (for the values shown in Table 1). The MV's have not been refined in figure 3. The loss of quality averages 0.5 dB (between 0.0-1.0 dB) over 150 transcoded frames. The result for the other test sequences are similar. Clearly, there is some loss in quality of the transcoded sequence which becomes insignificant when the MV's are refined in a small area. Notice that this loss is less evident in the other test sequences which do not have high motion.

3.2 Resolution reduction

The resolution reduction by a factor of 2 is studied. Results for transcoding from CIF to QCIF format are presented.

The refinement search windows were for the 'Akiyo' sequence [-1,+1], for the 'M&D' [-3,+3] and for the 'Foreman' sequence' [-5,+5]. Full search method using SAD was used as before. The frame rates and resolutions are the same as those used in the rate reduction mode. The macroblock type was re-evaluated at the encoder. As can be seen in table 2, the complexity is reduced when MV's are passed and refined, without a significant increase in the output bitrate. The median method seems to be the best method among the three tested ones.

The performance of the method is shown in figure 4 for the 'Foreman' sequence. Similar results were observed for the other test sequences.

4. COMPARISONS WITH SCALABLE VIDEO CODING

Scalable video coding is included in version 2 of H.263 (H.263+) [5]. Both SNR and spatial scalability are considered. Since this is a different way to achieve multipoint video conference, we compare the scalability modes with the transcoding modes. The results of the comparisons in quality can be seen in figures 5 and 6, which show a comparison between rate transcoding and SNR scalable coding and between resolution transcoding and spatial scalable coding for the 'Foreman' test sequence.

It can be observed that the base layer of the scalable bitstream provides better quality compared to the transcoder bitstream. This is because the coding of the base layer is the same as in non-scalable coding. However, the quality of the enhancement layer is not as good as the direct coded one. With the use of a transcoder the participants that can receive video encoded at the highest bitrate will not be affected.

From the point of complexity, transcoding will add complexity and delay in the network while scalable video coding adds complexity to the video encoder and decoder (scalable encoder/decoder is required). For example, the encoder requires undersampling and up-sampling operations (base layers are used for prediction in the enhancement layers), as well as two motion estimation/compensation loops for two layer spatial scalable coding. The scalable decoder will require an extra up-sampling operation and two motion compensation loops. In contrast, when a transcoder is used in the network, scalable encoders/decoders are not required, and therefore they can be simple encoders and decoders. However, the delay introduced due to the transcoding operation and the loss in quality of the base layers need to be considered also. A direct advantage of video transcoders in comparison to scalable coding is that there is no need to know in advance the capabilities of the participants in the video conference.

sequence	R_1 kbits/s	R_2 kbits/s	Size (%) (B'-B)* (100/B)	Complexi -ty(%) (T1-T2)* (100/T1)
Akiyo				
Average	132	55	26.3(0.0)	14.0(7.0)
Median	132	55	10.9(0.0)	14.0(7.0)
Single	132	55	20.1(0.4)	14.0(7.0)
M&D				
Average	120	36	52.1(1.2)	18.8(12.5)
Median	120	36	21.3(0.7)	18.8(10.4)
Single	120	36	35.0(1.5)	20.8(12.5)
Foreman				
Average	136	31	79.5(6.1)	34.2(17.1)
Median	136	31	22.2(1.4)	34.2(17.1)
Single	136	31	29.5(4.1)	34.2(20.0)

Table 2 Results for the resolution reduction model

5. CONCLUSIONS AND FURTHER RESEARCH

Transcoder architectures for H.263 video coding were investigated. Rate and resolution transcoding were described and methods to speed up the computation of the transcoding operation were proposed. The transcoder can reside in a gateway or multipoint control unit and guarantee that the participants of the video conference are not dragged down by the lowest common denominator**.** Comparisons with scalable video coding show that the transcoder makes possible to have simple, non-scalable encoders and decoders, while at the same time guarantees that the participants will get the best quality. It would be interesting to compare scalable coding with transcoding, when more than two users with different capabilities in terms of bandwidth are connected. This is because in general scalability increases the total bit rate required to encode the enhancement layers to a quality that is comparable to the one achieved by direct coding .

Further research would be in the direction of computational complexity reduction methods, fast undersampling methods in the DCT domain, and motion estimation/motion vector refinement in the DCT domain. Recently published papers [3,8] show that some of these operations can be done in the DCT domain, but no direct applications of these for H.263 video coding has been reported to the knowledge of the authors.

6. REFERENCES

[1] P. A. A. Assuncao and Mohammed Ghanbari, Postprocessing of MPEG2 Coded Video for Transmission at lower bit rates, Proc. IEEE ICASSP '96, vol. 4, pp. 1998-2001, May 1996.

[2] V. Bhaskaran and K. Konstantinides, Image and Video Compression Standards, Algorithms and architectures, second edition, Kluwer Academic Publishers, 1997.

[3] S.F. Chang and D. Messerschmitt, "Manipulation and composition of MC-DCT compressed video", IEEE JSAC Vol. 13, No. 1, January 1995, pp. 1-11.

[4] ITU-T Recommendation H.263, "Video coding for low bitrate communication", 11/95.

[5] ITU - T, Study Group 16, Draft Text of recommendation H.263 version 2 (H.263+), September 1997.

[6] G. Keesman, R. Hellinghuizen, Fokke Hoeksema and Geert Heideman, Transcoding of MPEG bitstreams, Signal Processing: Image Communication, vol. 8, pp. 481-500, September 1996.

[7] L.-K. Liu, and E. Feig, "A block-based gradient descent search algorithm for block motion estimation in video coding", IEEE Trans. CAS, Vol. 6, No. 4, pp. 419-422, August 1996.

[8] B. K. Natarajan and B. Vasudev, "A fast approximate algorithm for scaling down digital images in the DCT domain", Proc. ICIP 95, Vol. II, pp. 241-243, October 1995.

[9] H. Sun, W. Kwok and J. W. Zdepski, Architectures for MPEG compressed bitstream scaling, IEEE Trans. on Circuits and Systems for Video Technology, vol. 6, No. 2, pp. 191-199, April 1996.

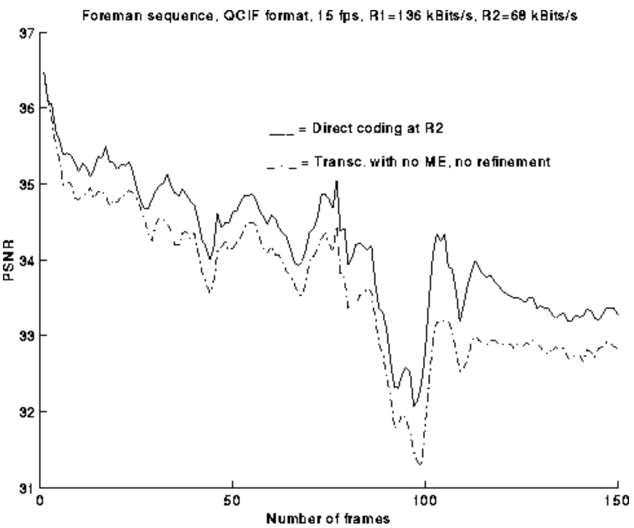

Figure 3 Quality comparison between transcoding from 136 to 68 Kb/s (MV's are passed directly without any refinement) and direct coding at 68 Kb/s

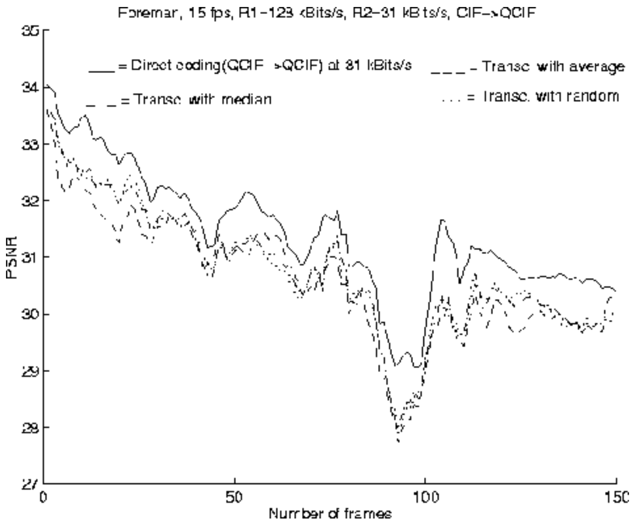

Figure 4 Comparison between direct coding at 31 Kb/s and transcoding from 136 to 31 Kb/s, CIF to QCIF

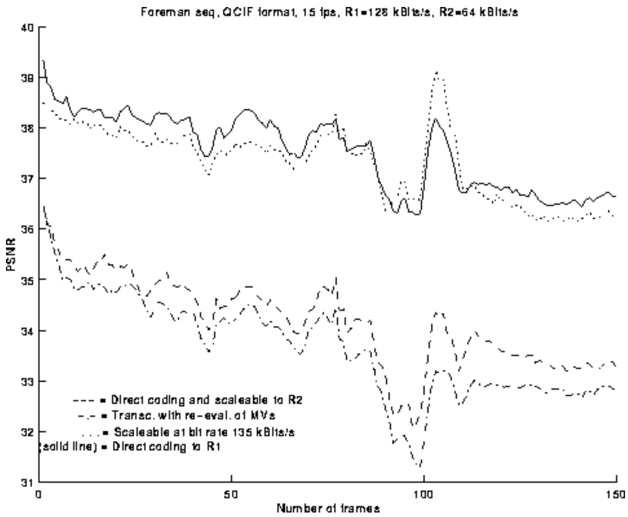

Figure 5 Comparison between SNR scalable coding and transcoding

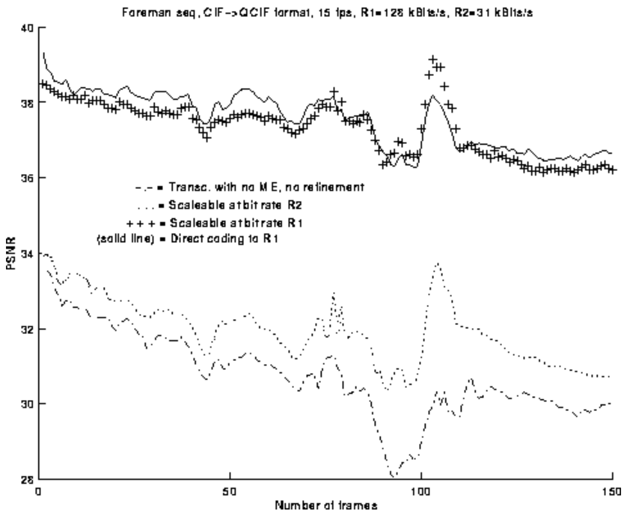

Figure 6 Comparison between spatial scalable coding and transcoding