A COMPRESSED-DOMAIN HETEROGENEOUS VIDEO TRANSCODER

Wan-Chi Siu, Kai-Tat Fung and Yui-Lam Chan

Centre for Multimedia Signal Processing
Department of Electronic and Information Engineering
The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

ABSTRACT

The forthcoming multimedia telecommunication services are expected to use pre-encoded videos for storage and transmission. The heterogeneities of the present communication networks and the clients' devices demand matching the encoding format of the video source to the constraints of the networks and clients' devices. In order to provide quality support services to heterogeneous clients or transmission channels, the video server should have the capability of performing heterogeneous transcoding, which is regarded as a process of converting a previously compressed video bitstream into another bitstream with a different format. However, much investigation has been conducted to focus on homogeneous transcoding. Only a limited number of methods have been proposed to realize the heterogeneous transcoding from MPEG-2 to H.263. The major difficulties for transcoding B-picture to P-picture is that the incoming DCT coefficients are predicted from the forward and backward prediction. In other words, the DCT coefficients predicted from the previous frame is not available. Therefore, the required new prediction errors need to be re-estimated in the pixel domain. This process involves high computational complexity as well as introduces re-encoding errors. Motivated by this, we propose a new transcoder architecture to convert a B-picture into a P-picture by making use of the techniques of motion compensation in the DCT domain and indirect addition of DCT coefficients. In this paper, we derive a set of equations and formulate the problem of how to obtain the DCT coefficients. Experimental results show that the proposed video transcoder achieve a better performance as compared to the conventional video transcoder in terms of both quality and complexity. Furthermore, we propose a fast algorithm to speed up the transcoding process making use of the correlation of motion activities between pictures.

I. INTRODUCTION

Homogeneous transcoding techniques of MPEG-2 to MPEG-2, H.261 to H.261, H.263 to H.263 have been investigated[1-4]. However, there are requirements for heterogeneous transcoding, such that decoders of one form (e.g., H.263) have to receive videos which were previously coded by another form (e.g., MPEG-2). This is particularly important for transmitting videos over low bandwidth channels or hostile environments such as mobile networks and internet. The emergence of the forthcoming Universal Mobile Telecommunication System (UMTS) carrying video, voice, and data is a good example. In this case, in contrast to homogeneous transcoding, picture type, directionality of motion vectors, and picture rate might all change. To resolve this problem, one straightforward approach for implementing heterogeneous transcoding is to cascade a decoder and an encoder, commonly known as pixel-domain transcoding as shown in Figure 1. The incoming MPEG-2 video bitstream is decoded in the pixel domain, and the decoded video frame is re-encoded by an H.263 encoder at the desired output bitrate according to the capability of the clients' devices and the available bandwidth of the network. This involves high processing complexity, large memory, and long delay. Recently, some fast algorithms have been proposed by making use of the information from the incoming bitstream [1-7] to reduce computational complexity. For example, motion vectors extracted from the incoming bitstream after decoding can be used to reduce significantly the complexity of the transcoding, since motion re-estimation can be avoided. In addition, the video quality of pixel-domain transcoding approach suffers from its intrinsic double-encoding process, which introduces additional degradation.

Motivated by this, the main objective of this paper is to investigate these bottlenecks and propose possible solutions to alleviate the problems. As a consequence, compressed-domain heterogenerous transcoding will be derived, under which the transcoding process is carried out in the DCT domain where complete decoding and re-encoding are not required; hence the processing complexity is significantly reduced. In addition, a fast DCT-domain video transcoding is proposed by assuming uniform motion activities between pictures.

The organization of this paper is as follows. Section II presents the details of the proposed DCT-based heterogeneous video transcoder. Experimental results are then given in Section III. Finally, some concluding remarks are provided in Section IV.

0-7803-8554-3/04/$20.00 ©2004 IEEE.

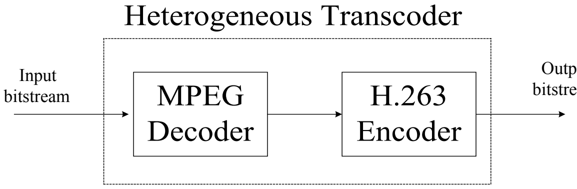

Figure 1. Conventional Heterogeneous transcoder

II. HETEROGENEOUS VIDEO TRANSCODER

Let us consider a heterogeneous video transcder, for example, which transcodes a video sequence from MPEG-2 to H.261 or H.263. Thus, we have to design efficient algorithms for transcoding among I-pictures, P-pictures and B-pictures. For standard codecs, a variety of transcoding format can exist. For the sake of discussion, let us confine our discussion within a reasonable size. In the following, we assume a group of pictures (GOP) in the incoming bitstream, which has a length of 8 pictures (*N=8*) and the distance between the anchor I/P pictures is set to 4 pictures (*M=4*) and the format of the output picture sequence is either in H.261 or H.263, with the sequence structures ($N=\infty$, $M=1$), as depicted in Figure 2. The superscript, *'* , is used to denote the picture after performing the transcoding. In Figure 2, pictures are presented in the display order, but are numbered in the encoding order. We consider these scenarios as two likely cases of the conversion, and the method can be generalized easily to other picture formats of the incoming and outgoing bitstreams.

For the first B-picture in the sub-group of an input bitstream, such as B_2, this picture is converted into an output P-picture, $P_1^{'}$, with a prediction from I_0, as shown in Figure 2. The new output motion vector, $mv^{out}_{0\to1}$, can be obtained directly from the forward motion vector of B_2, $mv^{in}_{0\to2}$,as depicted in Figure 3. However, the output prediction error in the DCT-domain, $DCT[e^{out}_{0\to1}]$, is not available from the incoming bitstream since the prediction error of incoming B_2 is derived from the previous I_0 and following P_1, such that:

$$e^{in}_{0\to2\leftarrow1}(k+j,l+i)=\frac{B_2(k+j,l+i)-I_0(k+j+mvy^{in}_{0\to2},l+i+mvx^{in}_{0\to2})}{2}+\frac{B_2(k+j,l+i)-P_1(k+j+mvy^{in}_{1\to2},l+i+mvx^{in}_{1\to2})}{2}\qquad(1)$$

where *(k,l)* represents the location of the upper left corner of a macroblock, *(i,j)* represents the spatial location within the macroblock, *mvx* represents the x-displacement of motion vector *mv* and *mvy* represents the y-displacement of motion vector *mv.*

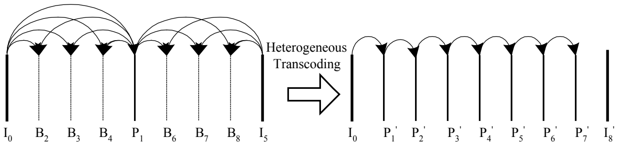

Figure 2: Typical heterogeneous transcoding.

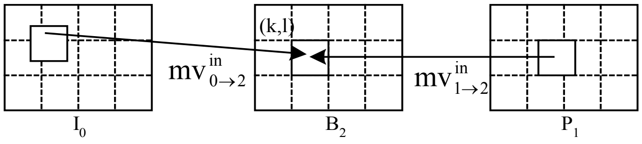

Figure 3: Forward and Backward motion vectors of B_2.

By applying the DCT to $e^{in}_{0\to2\leftarrow1}$ and taking into account of the linearity of the DCT, we obtain an expression for $e^{in}_{0\to2\leftarrow1}$ in the DCT-domain.

$$2DCT[e^{in}_{0\to2\leftarrow1}(k,l)]=DCT[B_2(k,l)-I_0(k+mvy^{in}_{0\to2},l+mvx^{in}_{0\to2})]+DCT[B_2(k,l)-P_1(k+mvy^{in}_{1\to2},l+mvx^{in}_{1\to2})]\qquad(2)$$

From eqn (2), as it might be expected that $DCT[e^{out}_{0\to1}]$ cannot be obtained directly from the incoming bitstream. Let us make an approximation. Let the quantized DCT coefficients of the output prediction errors $DCT[e^{out}_{0\to1}]$ be given by the following equation,

$$DCT[e^{out}_{0\to1}(k,l)]=DCT[B_2(k,l)-I_0(k+mvy^{out}_{0\to1},l+mvx^{out}_{0\to1})]\qquad(3)$$

Since $mv^{out}_{0\to1}=mv^{in}_{0\to2}$, from eqn (2) and eqn (3), we obtain

$$\begin{aligned}2DCT[e^{out}_{0\to1}(k,l)]&=2DCT[e^{in}_{0\to2\leftarrow1}(k,l)]\\&+DCT[P_1(k+mvy^{in}_{1\to2},l+mvx^{in}_{1\to2})]\\&-DCT[I_0(k+mvy^{in}_{0\to2},l+mvx^{in}_{0\to2})]\qquad(4)\end{aligned}$$

By observing the prediction error of P_1 and making further manipulation of these equations, we can obtain an expression for the output prediction errors in the DCT-domain as shown below,

$$\begin{aligned}DCT[e^{out}_{0\to1}(k,l)]&=DCT[e^{in}_{0\to2\leftarrow1}(k,l)]\\&+\tfrac{1}{2}DCT[e^{in}_{0\to1}(k+mvy^{in}_{1\to2},l+mvx^{in}_{1\to2})]\\&+\tfrac{1}{2}DCT[I_0(k+mvy^{in}_{0\to1}+mvy^{in}_{1\to2},l+mvx^{in}_{0\to1}+mvx^{in}_{1\to2})]\\&-\tfrac{1}{2}DCT[I_0(k+mvy^{in}_{0\to2},l+mvx^{in}_{0\to2})]\qquad(5)\end{aligned}$$

From eqn(5), we can see that the first term can be obtained directly from the incoming bitstream. The second term is the motion-compensated prediction error of the incoming bitstream and it can be computed mainly on the DCT-domain. By using the DCT-domain inverse motion-compensation proposed in [5], we can also obtain the third and fourth terms in the DCT-domain.

Besides, we can approximate $DCT[e^{out}_{0\to1}]$, by assuming that motion activities between pictures are uniform(UMA), such that,

$$mv_{0\to 2}^{in} = mv_{0\to 1}^{in} + mv_{1\to 2}^{in} \qquad (6)$$

Hence eqn.(5) can be further simplified as

$$\begin{aligned}\mathrm{DCT}[e_{0\to 1}^{out}(k,l)] &= \mathrm{DCT}[e_{0\to 2\leftarrow 1}^{in}(k,l)] \\ &+ \tfrac{1}{2}\mathrm{DCT}[e_{0\to 1}^{in}(k + mvy_{1\to 2}^{in}, l + mvx_{1\to 2}^{in})] \qquad (7)\end{aligned}$$

In order to achieve high video quality for transcoding and low computational complexity, this assumption has to be verified. A process, designated as uniform motion checking(UMC), is applied to check whether the eqn (6) is satisfied or not. If the motion between the pictures cannot fulfil this requirement, SW_2 will be closed to position A_2 to estimate the new DCT coefficients using eqn (5) instead of eqn (7).

In the next section, the performance of using uniform motion assumption(UMA) and uniform motion checking(UMC) will be compared. Note that both approaches can obtain the targeted DCT coefficients in the DCT domain by reusing the incoming DCT coefficient in the above formulation. Using this formulation, the architecture of the proposed transcoder is as shown in Figure 4.

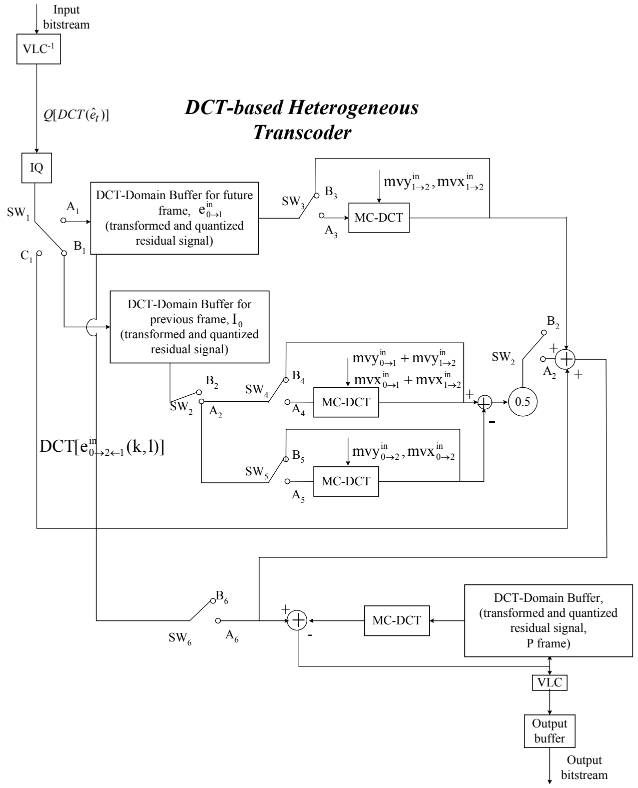

Figure 4. Architecture of the proposed heterogeneous video transcoder.

The input bitstream is first parsed with a variable-length decoder to extract the header information, coding mode, motion vectors and quantized DCT coefficients for each macroblock. Each macroblock is then manipulated independently. Switches SW_1, SW_2 and SW_6 are used to update the DCT-domain buffer for the transformed and quantized residual signal depending upon the coding mode originally used in the front encoder for the current macroblock being processed. SW_1 is used to update the DCT-domain buffer based on the frame type from the incoming bitstream. SW_2 is used to check the condition on uniform motion assumption(UMC) and SW_6 is used to convert the incoming P frame to a new reconstructed P frame since the reference frame is changed. Also, SW_3, SW_4 and SW_5 are used to control the MC-DCT modules. The function of switches are shown in Table 1 to Table 4. Note that all operations are performed in the DCT-domain, full re-encoding process are not required. Therefore, the complexity of the transcoder can be reduced significantly. Also, quality degradation during the transcoding process can be avoided.

Table 1. Switch positions for different frame types.

Frame Type	SW_1 Position
Previous frame	B_1
Future frame	A_1
B	C_1

Table 2. Switch positions for using uniform motion assumption(UMA).

Uniform motion assumption(UMA)	SW_2 Position
Yes	B_2
No	A_2

Table 3. Switch positions for converting different frame type.

Incoming B frame	SW_6 Position
Yes	B_6
No	A_6

Table 4. Switch positions for different coding modes.

Coding mode	SW_3 Position	SW_4 Position	SW_5 Position
Non MC	B_3	B_4	B_5
MC	A_3	A_4	A_5

III. EXPERIMENTAL RESULTS

To evaluate the overall efficiency of the proposed heterogeneous video transcoder, all test sequences of QCIF (*176 × 144*) were encoded at 128kb/s using a fixed quantization parameter with I-B-P structure(M=3, Intra period=15). All B frames have been converted to P frames inside the transcoder. Results of our experimental work show that our proposed video transcoders, namely 1)DCT-domain transcoder with uniform motion assumption (DA+MCDCT+UMA) and 2)DCT domain transcoder with motion checking(DA+MCDCT+UMC), outperform the conventional pixel domain transcoder(CPDT) in all cases as shown in Table 5 and Table 6. The results are more significant for sequences with high motion activities because our proposed architecture transcodes the DCT coefficients in the DCT domain to achieve low computational complexity as well as reduce re-encoding errors. Significant improvement can be achieved which is about 2.2-2.8dB as compared with the conventional video transcoder. Also, the speed up factor is about 2.1 to 7.3 times as compared with the conventional video transcoder[7].

Table 5. Performance comparison for all transcoded B-frames using the proposed DCT-based heterogeneous transcoder with uniform motion assumption (UMA) and conventional video transcoder

Sequences	Input bitrate	CPDT	DA+MCDCT+UMA	
		Average PSNR	Average PSNR	Speed-up ratio as compared with CPDT
Salesman (176x144)	64k	33.34	35.61	6.89
	128k	36.78	39	7.33
Foreman (176x144)	64k	30.53	32.69	4.65
	128k	34.28	36.56	4.85
Carphone (176x144)	64k	32.11	34.15	5.55
	128k	34.74	36.93	5.72
Table Tennis (352x240)	1.5M	32.41	34.54	5.06
	3M	35.04	37.33	5.30
Football (352x240)	1.5M	30.18	32.32	4.47
	3M	33.89	36.20	4.73

Table 6. Performance comparison for all transcoded B-frames using the proposed DCT-based heterogeneous transcoder with uniform motion assumption (UMA) and uniform motion checking(UMC)

Sequences	Input bitrate	DA+ MCDCT+ UMA	DA+MCDCT+UMC	
		Average PSNR	Average PSNR	Speed-up ratio as compared with CPDT
Salesman (176x144)	64k	35.61	35.62	5.30
	128k	39	39.01	6.33
Foreman (176x144)	64k	32.69	33.22	2.32
	128k	36.56	37.1	2.49
Carphone (176x144)	64k	34.15	34.34	3.22
	128k	36.93	37.08	3.43
Table Tennis (352x240)	1.5M	34.54	34.78	2.69
	3M	37.33	37.54	2.94
Football (352x240)	1.5M	32.32	32.91	2.16
	3M	36.20	36.80	2.38

IV. CONCLUSION

In this paper, we have proposed a new architecture for a heterogeneous video transcoder to convert a B-picture into a P-picture in the DCT-domain by using techniques relating to motion compensation in the DCT domain and an indirect addition of DCT coefficients. Also, we have derived a set of equations and formulated the problem of how to obtain the DCT coefficients. Experimental results show that the proposed video transcoder achieves better performance as compared to the conventional video transcoder in terms of both quality and complexity.

V. ACKNOWLEDGMENTS

This work is supported by the Centre for Multimedia Signal Processing, Department of Electronic and Information Engineering, Hong Kong Polytechnic University and the Research Grant Council of the Hong Kong SAR Government (PolyU 5234/03E). K.T. Fung acknowledges the research studentships provided by the same University.

VI. REFERENCES

[1] Jeongnam Youn, Ming-Ting Sun and Chia-Wen Lin, “Motion vector refinement for high-performance transcoding,” IEEE Trans. Multimedia, vol. 1, pp. 30-40, March 1999.

[2] K. T. Fung, Y. L. Chan and W. C. Siu, “New architecture for dynamic frame-skipping transcoder,” IEEE Trans. Image Processing, vol.11, pp. 886-900, August 2002.

[3] Bo Shen, Ishwar K. Sethi and Bhaskaran Vasudev, “Adaptive motion-vector resampling for compressed video downscaling,” IEEE Transactions on circuit and systems for video technology, vol.9, no.6, September 1999.

[4] Tap-Peng Tan; Haiwei Sun; YongQing Liang, “On the methods and applications of arbitrarily downsizing video transcoding,” in Proc. IEEE International Conference on Multimedia and Expo, vol.1 , pp.609-612, 2002

[5] J. Song and B.L. Yeo, “A Fast Algorithm for DCT-Domain Inverse Motion Compensation Based on Shared Information in a Macroblock,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 10, no. 5, pp. 767-775, August 2000.

[6] J. Chalidabhongse and C.-C. Jay Kuo, “Fast motion vector estimation using multiresolution-spatio-temporal correlations,” IEEE Trans. Circuits Syst. Video Technol., vol. 7, pp. 477-488, June 1997.

[7] Shanableh, T.; Ghanbari, M., “Heterogeneous video transcoding to lower spatio-temporal resolutions and different encoding formats,” IEEE Transactions on Multimedia, vol.2, pp.101-110, June 2000.