DYNAMIC FRAME SKIPPING FOR HIGH-PERFORMANCE TRANSCODING

Kai-Tat Fung, Yui-Lam Chan and Wan-Chi Siu
Centre for Multimedia Signal Processing
Department of Electronic and Information Engineering
The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

ABSTRACT

Transcoding is a process of converting a previously compressed video bitstream into a lower bit-rate bitstream. When some incoming frames are dropped for the frame-rate conversion in transcoding, the newly quantized DCT coefficients of prediction error need to be re-computed, which can create an undesirable complexity as well as introduce re-encoding error. In this paper, we propose new architecture of frame-skipping transcoder to improve picture quality and to reduce complexity. It is observed that re-encoding error is reduced significantly when the strategy of direct summation of DCT coefficients and the error compensation feedback loop are employed. Furthermore, we propose a frame-rate control scheme which can dynamically adjust the number of skipped frames according to the incoming motion vectors and the re-encoding error due to transcoding such that the decoded sequence can have smooth motion as well as better transcoded pictures. Experimental results show that, as compared to the conventional transcoder, the new frame-skipping transcoder is more robust, produces smaller requantization errors, and has simple computational complexity.

1. INTRODUCTION

To transport video over low bandwidth channels, a high transcoding ratio is required. However, the high transcoding ratio may result in an unacceptable picture quality when the incoming bitstream is transcoded with the full frame-rate[1-2]. Frame skipping[1-3] is often used as an efficient scheme to allocate more bits to the remaining frame, so that an acceptable quality for each frame can be maintained.

One straightforward approach for implementing a transcoding is to cascade a decoder and an encoder, commonly known as pixel-domain transcoding. The incoming bitstream is decoded in the pixel domain, and the decoded video frame is re-encoded at the desired frame rate. This involves high processing complexity, memory, and delay. As a consequence, some information reusing approaches[1-2] have been proposed, in which motion vectors extracted from the incoming bitstream after the decoding can be used to significantly reduce the complexity of the transcoding. Besides, the video quality of the pixel-domain transcoding approach suffers from its intrinsic double-encoding process, which introduces additional degradation. The effect of the re-encoding error is depicted in Figure 1 where the "Salesman" sequence was encoded. In the figure, the peak signal-to-noise ratio (PSNR) of the frame-skipping pictures is plotted to compare with that of the same pictures directly using a decoder without a transcoder. This figure shows that the re-encoding error leads to a drop in picture quality of about 3.5dB on average, which is a significant degradation.
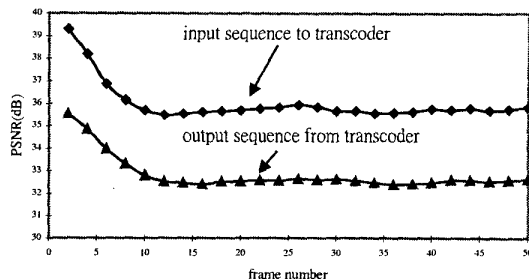

Figure 1. Quality degradation of conventional frame-skipping transcoder for "Salesman" sequence.

In recent years, the DCT-domain transcoding was introduced[4-5], under which the incoming bitstream is partially decoded to form the DCT coefficients and downscaled by the requantization of the DCT coefficients. Since the DCT-domain transcoding is carried out in the coded domain where complete decoding and re-encoding are not required, the processing complexity is significantly reduced. However, the frame-rate conversion has not been fully considered in the literature. When the frame rate changes, the incoming quantized DCT coefficients of the residual signal are no longer valid because they refer to the frames which have been dropped. Thus, it is difficult to perform frame-skipping in the DCT-domain since the prediction error of each frame is computed from its immediate past frames. In this paper, we provide a computationally efficient solution to perform frame skipping in a transcoder, mainly in the DCT-domain, to avoid the complexity and the quality degradation arising from pixel-domain transcoding. In addition, a frame-skipping control scheme with dynamic behaviour is proposed, which can provide a smoother and better transcoded sequence.

2. PROPOSED FRAME-SKIPPING TRANSCODER

The architecture of the proposed transcoder is shown in Figure 2. 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. The two switches $S1$ and $S2$ are employed to update the DCT-domain buffer for the transformed and quantized residual signal depending on the coding mode originally used at the front encoder for the current macroblock being processed. When the macroblock is not motion compensated, the previous residual signal in the DCT-domain is directly fed back from the DCT-domain buffer to the summer, and the sum of the input residual signal and the previous residual signal in the DCT-domain is updated in the buffer. Note that all operations are performed in

425

the DCT-domain, thus the complexity of the frame-skipping transcoder is reduced. Also, the quality degradation of the transcoder is avoided. When the motion compensation is used, motion compensation, DCT, inverse DCT, quantization and inverse quantization modules are activated to update the DCT-domain buffer. The problem with those motion compensated macroblocks is that re-encoding error will accumulate, and it is observed that human eyes are sensitive to this type of abrupt quality degradation of the transcoded sequence. Thus, the feedback loop in Figure 2 tries to correct the re-encoding error introduced in the previous frames. Note that switch S3 is used to control the frame rate and refresh the frame buffer for non-skipped frames. Another advantage of the proposed architecture is that when multiple frames are dropped, it can be processed in the forward order, thus eliminating the multiple DCT-domain buffers that are needed to store the incoming quantized DCT coefficients of all dropped frames. Thus, only one DCT-domain buffer is needed for all the dropped frames.
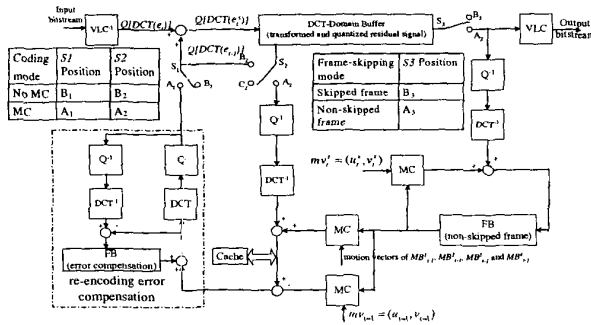

Figure 2. The proposed frame-skipping transcoder.

2.1 Direct summation of DCT coefficients for macroblock without motion compensation

In Figure 3, a situation in which one frame is dropped is illustrated. We assume that MB_t represents the current macroblock and MB_{t-1} represents the best matching macroblock to MB_t. Since MB_t is coded without motion compensation, the spatial position of MB_{t-1} is the same as that of MB_t, and MB_{t-2} represents the best matching macroblock to MB_{t-1}. Since the previous reconstructed frame, R_{t-1}, is dropped, for MB_t, we need to compute a motion vector, (u_t^s, v_t^s), and the prediction error in the quantized DCT-domain, $Q[DCT(e_t^s)]$, by using R_{t-2} as a reference. Since the motion vector in MB_t is zero, then $(u_t^s, v_t^s) = (u_{t-1}, v_{t-1})$. Since re-encoding can lead to an additional error, it could be avoided if $Q[DCT(e_t^s)]$ can be computed in the DCT-domain. In Figure 3, the pixels of MB_t can be reconstructed by performing inverse quantization and inverse DCT of the incoming quantized and transformed prediction error in frame t, $Q[DCT(e_t)]$, and summing this residual signal to pixels in MB_{t-1} which can be similarly reconstructed by performing inverse quantization and inverse DCT of the incoming quantized and transformed prediction error in frame $t-1$, $Q[DCT(e_{t-1})]$ and summing this residual signal to pixels in the corresponding MB_{t-2}. However, by considering the linearity of inverse DCT and inverse quantization, we obtain,

$$Q[DCT(e_t^s)] = Q[DCT(e_t)] + Q[DCT(e_{t-1})] \qquad (1)$$

Equation (1) implies that the newly quantized DCT coefficient $Q[DCT(e_t^s)]$ can be computed in the DCT-domain by summing directly the quantized DCT coefficients between the data in the DCT-domain buffer and the incoming quantized DCT coefficients, whilst the updated DCT coefficients are stored in the DCT-domain buffer, as depicted in Figure 2, when switches S_1 and S_2 are connected to B_1 and B_2 respectively. Since it is not necessary to perform motion compensation, DCT, quantization, inverse DCT and inverse quantization, the complexity is reduced. Furthermore, since requantization is not necessary for this type of macroblock, the quality degradation of the transcoder due to the re-encoding is also avoided. For a typical video sequence, the majority of the video signals are coded without motion compensation [6]. By using a direct summation of DCT coefficients for non-moving macroblocks, the computational complexity involved in processing these macroblocks can be reduced significantly and the additional re-encoding error can be avoided.
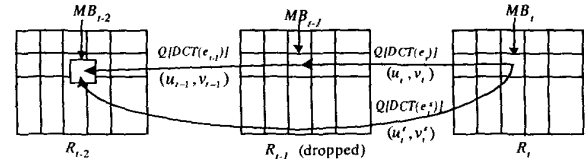

Figure 3. Residual signal re-computation of frame skipping for macroblocks without motion compensation.

2.2 DCT-domain buffer updating for motion-compensated macroblock with error compensation

For motion-compensated macroblocks, direct summation cannot be employed since MB_{t-1} is not on a macroblock boundary, as depicted in Figure 4. In other words, $Q[DCT(e_{t-1})]$ is not available from the incoming bitstream. It is possible to use the motion vectors and quantized DCT coefficients of the four neighboring macroblocks with MB_{t-1}, MB_{t-1}^1, MB_{t-1}^2, MB_{t-1}^3 and MB_{t-1}^4, to come up with $Q[DCT(e_{t-1})]$. First, inverse quantization and inverse DCT of the quantized DCT coefficients of MB_{t-1}^1, MB_{t-1}^2, MB_{t-1}^3 and MB_{t-1}^4 are performed to obtain their corresponding prediction errors in the pixel-domain. Figure 4 shows that the MB_{t-1} is composed of four segments. Thus, each segment of the reconstructed pixels in MB_{t-1} can be obtained by summing its prediction errors and its motion-compensated segment of the previous non-skipped frame stored in the frame buffer, as shown in the block diagram of Figure 2. After all pixels in MB_{t-1} have been reconstructed, we need to find the prediction error, e_{t-1}. Actually, e_{t-1} is equal to the reconstructed pixel in MB_{t-1} subtracted from the motion-compensated macroblock from the previous non-skipped frame stored in the frame buffer. In order to obtain the motion-compensated macroblock, we need to find a motion vector of MB_{t-1}. The dominant vector selection approach[1-2], which selects one dominant motion vector from four neighboring macroblocks, is employed. A dominant motion vector is defined as the motion vector carried by a dominant macroblock. The dominant

macroblock is the macroblock that has the largest overlapped segment with the MB_{t-1}. Hence, e_{t-1} is transformed and quantized to form $Q[DCT(e_{t-1})]$. Since re-quantization is performed in the formation of $Q[DCT(e_{t-1})]$, some quantization error, Δ, is introduced. The newly quantized DCT coefficient of a motion-compensated macroblock can be computed by

$$Q[DCT(e_t^s)] = Q[DCT(e_t)] + Q[DCT(e_{t-1})] + \Delta \quad (2)$$

The re-encoding process will introduce additional re-encoding errors. Since each non-skipped P-frame is used as a reference frame for the following non-skipped P-frame, quality degradation propagates to later frames in a cumulative manner. If the accumulated magnitude of re-encoding error is large, it means that the reconstructed quality of the transcoded sequence is degraded significantly. Human eyes are very sensitive to this type of accumulated re-encoding errors. Thus, the feedback loop in Figure 2 is used to compensate the re-encoding error introduced in the previous frames. The forward and inverse DCT and quantization pair in the feedback loop are mainly responsible for minimizing the re-encoding errors. For those motion compensated macroblocks, the quantized DCT coefficients are inversely quantized and an inverse DCT is then performed to form e_{t-1} with a re-encoding error, which is subtracted from e_{t-1} to generate the feedback re-encoding error signal. This feedback signal is stored in the frame buffer which is added to the prediction error of the following P-frame to compensate the re-encoding error. Use of the feedback loop for motion-compensated macroblocks can minimize significantly the accumulated error generated in processing these macroblocks.

In order to reduce the implementation complexity of the motion-compensated macroblock, a cache subsystem is added to our proposed transcoder, as depicted in Figure 2. Since motion compensation of multiple macroblocks may require the same pixel data, a cache subsystem is implemented to reduce redundant inverse quantization, inverse DCT and motion compensation computations.
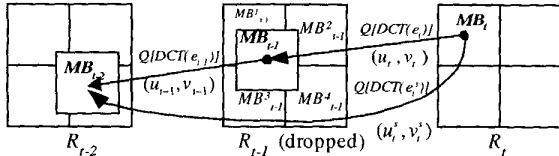

Figure 4. Residual signal re-computation of frame-skipping for motion-compensated macroblocks.

2.3 Frame-skipping Transcoder with Dynamic Behaviour

Another objective of our proposed frame-skipping transcoder is to develop a strategy for determining the length of the skipped frame such that it can reduce the quality degradation as well as minimize the motion jerkiness perceived by human beings. Traditionally, a motion vector is used to serve as a good indicator for dynamic frame skipping[7]. When multiple frames are dropped in the frame-skipping transcoder, re-encoding errors in the motion-compensated macroblock cannot be avoided entirely even though a feedback loop is applied to compensate the accumulated errors, as mentioned in the previous section. It is observed that human eyes are sensitive to this type of quality

degradation. Thus, it is necessary to regulate the frame rate of the transcoder by taking into account the effect of re-encoding. The goal of the proposed dynamic frame-rate control scheme is to minimize the re-encoding error as well as preserve motion smoothness. To obtain a quantitative measure for frame-skipping, we define the frame-skipping criterion, FSC_t^s, which is the accumulated magnitude of the motion vectors and the re-encoding error due to transcoding for the macroblocks in the current frame. The criterion is given by,

$$FSC_t^s = \frac{\sum_{i=1}^{N}(MA_t^s)_i}{\sum_{i=1}^{N}(RE_{t-1}^s)_i} \quad (3)$$

where N is the total number of macroblocks in the current frame and $(RE_{t-1}^s)_i$ is the re-encoding error after error compensation in the feedback loop for the ith macroblock with the corresponding motion activity, $(MA_t^s)_i$ which is defined by,

$$(MA_t^s)_i = \left|(u_t^s)_i\right| + \left|(v_t^s)_i\right| \quad (4)$$

where $(u_t^s)_i$ and $(v_t^s)_i$ are the horizontal and vertical components of the motion vector of the ith macroblock which uses the previous non-skipped frame as reference. It can be observed from Figure 2 that the re-encoding error $(RE_{t-1}^s)_i$ is obtained by summing all requantization error of e_{t-1} coefficients for the ith macroblock.

$$(RE_{t-1}^s)_i = \sum \left| \left\lfloor \frac{DCT(e_{t-1})}{q} \right\rfloor \times q - DCT(e_{t-1}) \right| \quad (5)$$

where q is the quantization step-size and the floor function $\lfloor a \rfloor$ extracts the integer part of the given argument a.

If the value of FSC_t^s following a non-skipped frame exceeds the predefined threshold, T_{FSC}, this incoming frame should be kept. $(MA_t^s)_i$ in equation (4) is used to detect the activity level of the ith macroblock. If $\sum_{i=1}^{N}(MA_t^s)_i$ has a significant value, this implies that the incoming frame contains a certain amount of motions. It is reasonable that the frame-rate control scheme be used to keep this frame since the previous non-skipped frame is not sufficient to represent the current frame. As a consequence, it is much better that the incoming frame be refreshed. However, we cannot guarantee the quality of the reconstructed frame due to the re-encoding error in the transcoder in case where only the motion activity is used. Since the quality of the selected frame directly affects the quality of all of the following P frames, it is usually beneficial to maintain the selected frame at a good reconstruction quality. The conventional algorithm fails to fulfill this mission. Thus, $\sum_{i=1}^{N}(RE_t^s)_i$ in equation (5) is used to measure the re-encoding error of the incoming frame. Also, a larger value of $\sum_{i=1}^{N}(RE_t^s)_i$ implies a larger re-encoding error, and it will diminish the value of FSC_t^s such that the incoming frame is not kept even

427

though it contains a certain amount of motion activities. The simulation results presented in Section 3 show the benefit of choosing a high quality reference frame.

3. SIMULATION RESULTS

To evaluate the overall efficiency of the proposed frame-skipping transcoding approach, all test sequences of QCIF (*176 × 144*) were encoded at 128kb/s using a fixed quantization parameter. At the front encoder, the first frame was coded as intraframe (I-frame), and the remaining frames were encoded as interframes (P-frames). These picture-coding modes were preserved during the transcoding. The PSNR performance of the proposed frame-skipping transcoder without frame-rate control for the "Salesman" sequence is shown in Figure 5. At the front encoder, the original test sequence "Salesman" was encoded at 128kb/s with 30 frame/s and then transcoded into 32kb/s with 7.5 frame/s. As shown in Figure 5, the proposed transcoder outperforms the conventional pixel-domain transcoder. Also, Table 1 shows that it has a speed-up of about 10 times faster than that of the conventional transcoder for the "Salesman" sequence. The result is further enhanced by incorporating the proposed frame rate control scheme. This is because our proposed scheme can reduce the re-encoding error by preserving the high quality reference frame. These demonstrate the effectiveness of the proposed frame-skipping transcoder and the new frame rate control scheme. Details of our simulation results for more test sequences are summarized in Table 1.

4. CONCLUSIONS

The paper proposes a low-complexity and high quality frame-skipping transcoder. Its low complexity is achieved by: 1) a direct summation of the DCT coefficients for macroblocks coded without motion compensation to deactivate most complex modules of the transcoder, and 2) a cache subsystem for motion-compensated macroblocks to reduce redundant IDCT and inverse quantization. We have also shown that a direct summation of the DCT coefficients on macroblocks without motion compensation and error compensation on motion-compensated macroblocks can reduce significantly the re-encoding error due to transcoding. Furthermore, we also propose a new criterion which employs incoming motion vectors and re-encoding error for dynamically adjust the frame rate. Overall, the proposed frame-skipping transcoder produces a better picture quality than the conventional frame-skipping transcoder at the same reduced bitrates.

5. ACKNOWLEDGMENTS

6. REFERENCES

[1] J. Youn, M.-T. Sun and C.-W. Lin, "Motion vector refinement for high-performance transcoding," IEEE Trans. on Multimedia, vol. 1, pp. 30-40, March 1999.

[2] J. Youn, M.-T. Sun and C.-W. Lin, "Motion estimation for high performance transcoding," IEEE Trans. on Consumer Electronics, vol. 44, pp. 649-658, August 1998.

[3] S.H.Kwok, W.C.Siu and A.G.Constantinides, "Adaptive temporal decimation algorithm with dynamic time window," IEEE Trans. on Circuits and Syst. for Video Tech., vol.8, pp. 104-111, Feb. 1998.

[4] H. Sun, W. Kwok and J.W. Zdepski, "Architectures for MPEG compressed bitstream scaling," IEEE Transactions on Circuits and System for Video Technology, vol. 6, pp. 191-199, April 1996.

[5] Y. Nakajima, H. Hori and T. Kanoh, "Rate conversion of MPEG coded video by re-quantization process," in IEEE International Conference on Image Processing, ICIP95, vol. 3, pp. 408-411, October 1995, Washington, DC.

[6] K.T. Fung, Y.L. Chan and W.C. Siu, "Low-Complexity and High Quality Frame-Skipping Transcoder," paper accepted, to be published on proceedings, ISCAS'2001.

[7] J.-N. Hwang, T.-D. Wu and C.-W. Lin, "Dynamic frame-skipping in video transcoding," 1998 IEEE Second Workshop on Multimedia Signal Processing, pp. 616 –621, 1998.
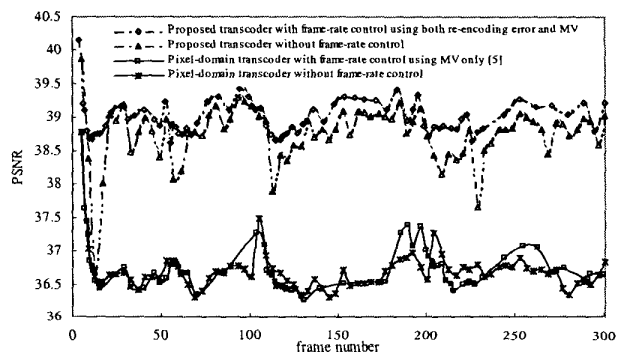

Figure 5. Performance using different dynamic frame-skipping transcoders of "Salesman" sequence.

Table 1. Performance of the proposed transcoder for various test sequences, with incoming bitstreams of 128kbps with 30fps.

Sequence	Method	Average PSNR	Speed-up ratio (as compared with method A)	Average encoded frames per sec
Salesman	A	36.69	-	7.5
	B	36.72	0.94	7.8
	C	38.74	10.15	7.5
	D	38.98	8.31	7.5
Carphone	A	34.87	-	7.5
	B	34.92	0.96	7.4
	C	36.21	7.70	7.5
	D	36.49	6.42	7.4
Foreman	A	34.60	-	7.5
	B	34.62	0.96	7.4
	C	36.18	6.65	7.5
	D	36.46	5.54	7.4

A: Pixel-domain transcoder without frame-rate control
B: Pixel-domain transcoder with frame-rate control using MV only [5]
C: Proposed transcoder without frame-rate control
D: Proposed transcoder with frame-rate control using both re-encoding error and MV