DIVERSITY AND IMPORTANCE MEASURES FOR VIDEO DOWNSCALING

Kai-Tat Fung 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

In video downscaling, simply reusing the motion vectors extracted from an incoming video bitstream may not result in good quality pictures. Several refinement schemes have been proposed recently to correct their recomposed new motion vectors in order to optimize the coding efficiency during the transcoding process. However, the major concern is that an optimal resampled motion vector may not exist during video downscaling process. In other words, it is difficult to use one motion vector to represent four motion vectors when the diversity of the incoming motion vectors is high. Besides, redundant computation for refinement has also been carried out even if the resampled motion vector is already optimal. Motivated by this, we propose an adaptive motion vector re-composition algorithm using two new measures: the diversity and importance measures of motion vectors. Using the importance measure, our proposed scheme manages to differentiate the most representative motion vector as a consideration to recompose a new motion vector. In addition, the diversity measure provides information for the video transcoder controlling the size of the refinement window to achieve a significant reduction of computational complexity. Experimental results show that our proposed adaptive motion vector re-composition scheme provides a high coding efficiency in terms of both quality and complexity.

I. INTRODUCTION

With the rapid growth of DVD popularity and the availability of broadband access networks, there is a strong need for users to access videos originally captured in a high resolution through the mobile multimedia capable devices. Motivated by this, different kinds of video transcoder have been proposed recently[1-7]. A major difficulty of designing a video downscaling transocder is to avoid computational complexity. More importantly, re-encoding errors are introduced during the re-encoding process. The most conventional approach for implementing transcoding is to cascade a decoder and an encoder [1], commonly known as pixel-domain transcoding. To downscale an encoded video produced by one of the current video compression standards such as MPEG, H.261 or H.263 which employ motion compensated prediction to exploit the temporal redundancy to achieve low bitrates, the conventional approach needs to decompress the video and then perform downscaling of the video in pixel domain[3]. Then new motion vectors and DCT coefficients for this downscaled video need to be recomputed inside a transcoder. This involves high computational complexity, large memory, and long delay on a video server to generate the downscaled video. As a consequence, some information reusing approaches jointly consider the DCT coefficients and motion vectors such as adaptive motion vector resampling for downscaled videos were suggested to provide a computationally efficient solution to recompose the new motion vector. However, the recomposed new motion vector needs to be refined in order to optimize the coding efficiency during the transcoding process. Due to different motion activities of video frames, the control of the size of the refinement window becomes a critical step. In other words, it is beneficial to minimize the computational cost when the motion vector is near optimal and build a refinement window more dynamically such that redundant operations can be avoided. More importantly, the major concern is that an optimal resampled motion vector may not exist during video downscaling process. In other words, it is difficult to obtain one motion vector to represent four motion vectors when the diversity of the incoming motion vectors is high. In [3], the resampling scheme suggested to align the weighting toward the worst prediction to recompose an outgoing motion vector from the incoming motion vectors of the incoming frame which has a higher resolution. A hybrid AMVR system was proposed [3] to downscale the video such that the transcoded sequence can avoid full motion re-estimation. These techniques are useful for video downscaling transcoders in the pixel-domain. However, the motion vector obtained is not optimal especially when there is a high diversity for its related motion vectors. Although a motion re-estimation process can be performed to resolve this problem, this can lead to high computational complexity as well as introduce re-encoding errors.

Figure 1 shows the optimal case of the motion vectors during video downscaling. In this case, four motion vectors have the same direction and magnitude. Hence the new motion vector can be obtained by using align-to-average weighting (AAW), align-to-best weighting (ABW), align-to-worst weighting (AWW), or adaptive motion vector resampling (AMVR) to provide an optimal motion

vector[3]. However, when not all the motion vectors are well aligned as shown in Figure 2 and Figure 3, a good motion vector resampling or motion vector refinement[1] is required in order to reduce the re-encoding error.

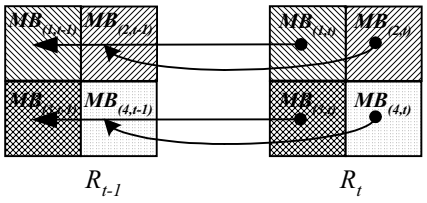

Figure 1. Diagram showing that all motion vectors have the same direction and magnitude.

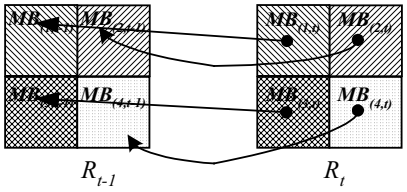

Figure 2. Different directions of the motion vectors with low diversity.

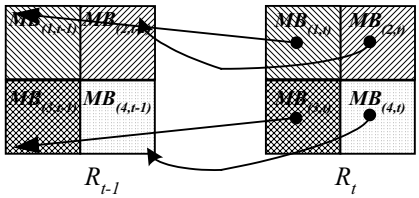

Figure 3. Different directions of the motion vectors with high diversity.

In [3], the spatial frames are reconstructed and downscaled in the pixel domain but the motion vectors are estimated directly from the existing motion vectors in the original sequence. Initially, a variable length decoding is performed and information of the motion vectors is extracted. Then inverse quantization and inverse DCT are performed for the incoming coefficients. After the motion compensation process, the pixel domain data are decoded and stored in the frame memory. This frame memory is used to reconstruct the next incoming frame. After the decoding process, a downscale process is applied to the decoded data in the pixel domain. In order to speed up the re-encoding process, motion re-estimation is not applied. In other words, the AMVR block is responsible for resampling adaptively the motion vectors, which can be described in the following equation:

$$\overline{mv}' = \frac{1}{2}\frac{\sum_{i=1}^{4} \overline{mv}_i A_i}{\sum_{i=1}^{4} A_i} \qquad (1)$$

where $\overline{mv}'$ denotes the new motion vector, $\overline{mv}_i$ denotes the motion vector of block *i* in the original *NxN* video and A_i denotes the activity measurement of the i^{th} residual block. The simplest way to calculate A_i is to count the number of non-zero AC coefficients. The effect of re-encoding errors is shown in Figure 4 where the "Table Tennis" sequence was transcoded at a quarter of the incoming frame size. This figure shows that re-encoding errors lead to a significant degradation in the picture quality.

Figure 4. Re-encoding error introduced by the video downscaling transcoder using AMVR system.

As shown in Figure 4, re-encoding errors are introduced, especially in the boundary regions or when the diversity of the motion vectors is high. For standards such as MPEG-4 Visual and H.263, there is support in the syntax for advanced prediction modes that allow one motion vector per 8x8 luminance block. This type of prediction modes allow us to have a more accurate representation of the motion activities. The major concern is that more bits are required since four motion vectors are coded instead of one(four-motion vector mode). Also, it is not necessary to use this mode for the well-aligned case.

In the next section, an adaptive motion vector re-composition for high performance spatial video transcoder using diversity and importance measures is proposed. Our architecture tries to differentiate the most representative motion vector as a consideration to recompose a new motion vector. The proposed architecture controls the size of refinement window dynamically to achieve a significant reduction of computational complexity and detects whether a four-motion vector mode or an intra-refresh technique is to be used to reduce the re-encoding error.

II. AN ADAPTIVE MOTION VECTOR RE-COMPOSITION FOR HIGH PERFORMANCE SPATIAL VIDEO TRANSCODER USING DIVERSITY AND IMPORTANT MEASURE (AMVR-DIM)

In this section, we present a new motion vector re-composition video downscaling transcoding architecture. The architecture of the transcoder is shown in Figure 5. The input bitstream is firstly 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. Switch *SW* is used to select appropriate tools to re-estimate the new motion vector. The selection depends on the motion vector and the amount of non-zero DCT coefficients. The switch positions for different coding modes are shown in Table 1. For non-MC macroblocks or the well-align case (e.g. all motion vectors having the same magnitudes and directions) as shown in Figure 1, the motion vector refinement process can be avoided. Hence, low computational complexity can be achieved. When the motion vectors are not well aligned as shown in Figure 2 and Figure 3, motion vector refinement process is necessary since the incoming prediction errors mismatch with the reconstructed new motion vector. However, high computational complexity is required for a large refinement window. Motivated by this, our proposed architecture estimates the new motion vector by considering its diversity and importance. The proposed diversity and importance measures are defined as follow:

The diversity measure of a resampled macroblock j is defined as

$$Diversity_j = Mv_{hi} - \hat{M}v_h + Mv_{vi} - \hat{M}v_v \qquad (2)$$

where Mv_{hi} and Mv_{vi} represent the horizontal and vertical components of the motion vector of the original macroblock i(i=0 to 3), respectively. $\hat{M}v_h$ and $\hat{M}v_v$ represent the average horizontal components and vertical components of all motion vectors, respectively.

The importance measure of a resampled macroblock j is defined as

$$Importance_j = \sum_{i=0}^{3} DCT_i \qquad (3)$$

where DCT_i represents the number of non-zero quantized DCT coefficients of the original macroblock i.

Then, high diversity of a resampled macroblock j can be defined as

$$\frac{Diversity_j}{\frac{1}{N}\sum_{j=0}^{N-1} Diversity_j} > 1 \qquad (4)$$

and high importance of resampled macroblock j can be defined as

$$\frac{Importance_j}{\frac{1}{N}\sum_{j=0}^{N-1} Importance_j} > 1 \qquad (5)$$

There are four types of conditions in this motion vector re-composition. The classification is made based upon their diversity and importance measures as shown in table 1.

Type 1 According to the diversity and importance measures, if both measures are low, the transcoder uses one of the motion vectors with the highest importance and performs motion vector refinement within ± 1 pixel.

Type 2 If the diversity measure is low but the importance measure is high, the transcoder uses the motion vector with the highest importance and performs motion vector refinement within ± 3 pixels.

Type 3 If the diversity measure is high but the importance measure is low, four-motion vector mode is employed in the transcoder.

Type 4 According to the diversity and importance measures, if both measures are high, Intra-refresh mode is employed in the transcoder.

If the diversity and importance measures are high, it is difficult to obtain a new motion vector to represent the original four motion vectors. Since the average energy is high in these macroblocks, intra-refresh is a good technique to apply in this case. On the other hand, if the diversity is high but the importance measure is low, it is beneficial to use four-motion vector mode since the incoming motion vectors have already minimized the prediction error successfully. It is advantageous to maintain these motion vectors instead of finding other solutions. If the diversity is low but the importance measure is high, a large refinement window is required since every macroblock contains a lot of prediction errors. A large refinement window can guarantee to find an optimal motion vector in this case. When the diversity and importance measures are low, only a small refinement window is necessary since the incoming motion vectors have already minimized the prediction error with good performance and it is easy to obtain a new motion vector when the diversity is low. By using this adaptive motion vector re-composition algorithm, the proposed transcoder can optimize the resultant motion vector in terms of good quality and low computational complexity. Besides, the transcoder can detect whether the new motion vector is good enough or has to choose an alternative solution to tackle the motion vector re-composition problem.

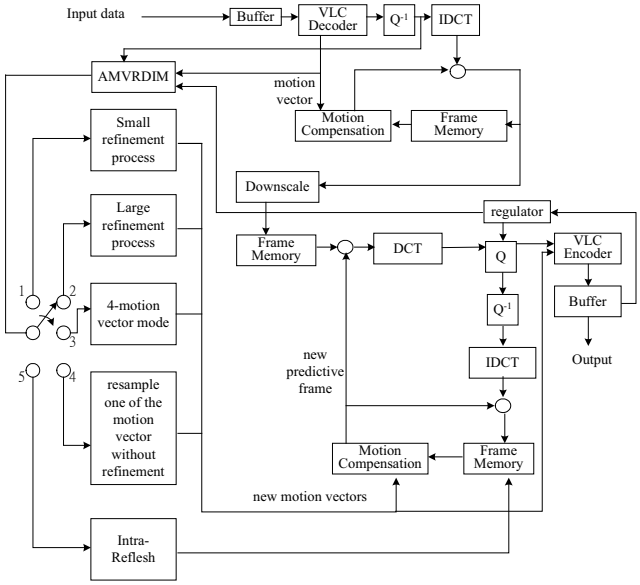

Figure 5. Architecture proposed for the video downscaling transcoder.

Table 1. Different coding modes of switches *SW* of the proposed transcoder.

Position	Diversity	Importance	Motion vector re-composition
5	High	High	Intra-refresh
3	High	Low	Four-motion vector mode
2	Low	High	Large refinement window is used
1	Low	Low	Small refinement window is used
4	Zero(well-align)	High/Low	No refinement process

III. EXPERIMENTAL RESULTS

Extensive simulations and performance comparison have been done with reference to a conventional pixel-domain transcoder (CPDT) for employing align-to-average weighting (AAW), align-to-best weighting (ABW) or adaptive motion vector resampling (AMVR) which is to resample a downscaled motion vector from the incoming motion vectors of the four macroblocks. In the front encoder, the first frame was encoded as intraframe (I-frame), and the remaining frames were encoded as interframes (P-frames). Picture-coding modes were preserved during transcoding. The pre-defined threshold can be determined adaptively by considering jointly the size of the buffer and the bitrate requirement. Larger diversity threshold and larger importance threshold will be set for the low bitrate application. Under this condition, the transcoder tends not to use the 4-motion vector and Intra Refresh mode since more bits are required in these two modes. This architecture is able to make the video downscaling transcoder adjust the video quality more dynamically according to the bandwidth requirement. Results of our experimental work show that the proposed video transcoders outperform CPDT+AAW, CPDT+ABW and CPDT+AMVR in all cases as shown in Table 2. The results are more significant for sequences with high motion activities because the proposed transcoder provides an optimal solution to resample the motion vectors according to the diversity and importance measures. Significant improvement in motion compensated regions can be achieved, which is about 0.7-1.9dB as compared with the conventional video downscaling transcoder. As compared with the refinement scheme with +/-3, 40 to 60% of the computational time can be saved for the proposed transcoder with the similar video quality of the transocded video.

IV. CONCLUSION

In this paper, we have proposed a simple architecture to form a low-complexity and high quality video downscaling transcoder to resolve the problem of Motion vector resampling. Using the diversity and importance measures, the proposed video transcoder is able to detect whether the optimal motion vector can be obtained during the resampling process. By jointly considering the diversity and importance measures, the proposed video transcoder can control the refinement window dynamically to achieve low computational complexity. Besides, the four-motion vector and Intra-Refresh modes provide alternative solutions for the video transcoder to tackle the problem of motion vectors with high diversity. Results of our experimental work show that the proposed architecture produces pictures quality with better as compared with the conventional video downscaling transcoder at the same reduced bitrates.

Table 2. Average PSNR of the proposed transcoder, where the frame rate of the incoming bitstream was 30 frames/s. H.263 was used as the front encoder for encoding "Salesman", "Miss_America", "Hall", "Tennis", "Football" and "Flower".

Sequences	Input bitrate	Average PSNR difference as compared with CPDT+AAW for MC macroblock transcoding.		
		CPDT+ABW	AMVR[3]	AMVR-DIM
Salesman (352x288)	512k	0.06	0.41	0.76
	256k	0.05	0.39	0.70
Miss_America (352x288)	512k	0.09	0.39	0.74
	256k	0.07	0.36	0.71
Hall (352x288)	512k	0.11	0.42	1.24
	256k	0.08	0.38	1.16
Tennis (352x240)	3M	0.12	0.43	1.45
	1.5M	0.08	0.38	1.41
Flower (352x240)	3M	0.18	0.47	1.71
	1.5M	0.15	0.43	1.55
Football (352x240)	3M	0.21	0.50	1.91
	1.5M	0.27	0.56	1.87

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] Vetro, A.; Christopoulos, C.; Huifang Sun; "Video transcoding architectures and techniques: an overview," in IEEE Signal Processing Magazine, vol.20, pp.18-29, March 2003.

[5] YongQing Liang; Lap-Pui Chau; Yap-Peng Tan, "Arbitrary downsizing video transcoding using fast motion vector reestimation," in IEEE Signal Processing Letters, vol.9, pp.352-355, Nov. 2002.

[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.