A Hybrid Method on Video Mixing for Multimedia Videoconference

Xin-Gang Liu[1] and Kook-Yeol Yoo[2]
Department of Information and Communication Engineering,
University of Yeungnam, South KOREA
[1]liuxingang78@yahoo.com.cn, [2]kyoo@yu.ac.kr

Abstract

In this paper, we propose a fast video mixing method for reducing the computational complexity in the MCU (Multipoint Control Unit) used in the video conferencing. The conventional mixing method is based on the pixel-domain transcoder, of which computational complexity is linearly increased as the number of participants is increased. Basically the method requires N decoders and one huge encoder to mix the bitstreams from the N participants. To reduce the computational complexity, we propose a hybrid mixing method based on the bitstream-domain and pixel-domain transcoding methods. The proposed method reduces the computational complexity about 45% at the improved quality, compared with the conventional mixing method based on the pixel-domain transcoders.

1. Introduction

With the broad deployment of the Internet, videoconference will be more and more popular. For the better management of the multiple bitstreams from the participants in the videoconference, the MCU (Multipoint Control Unit) is often used. One of the important functions in the MCU is the mixing of the video bitstreams to distribute the mixed video bitstream to the participants. Without mixing function of MCU, each participant needs (N-1) decoders to display the other participants. For this reason, the mixing function of MCU is inevitable thing in the multiparty videoconference. By providing the mixed bitstream to each participant by the MCU, the participant needs only one decoder. The computational complexity of the video mixing, however, can be drastically increased as the number of participants is increased. For instance, if there are N participants in the videoconference, the MCU requires the N video decoders and one video encoder to produce the mixed video bitstream. It means that the computational complexity for the video mixing is still high for the practical application. To use the MCU for the practical videoconference system, the reduction of computational complexity in the video mixing is, therefore, critically required. To reduce the computational complexity for the change of the compressed video signals, the transcoding algorithm has been discussed [1, 2]. Basic concept of the transcoding is the cascade decoder encoder structure. Simply the transcoder performs the full decoding of the compressed video to produce the reconstructed video signals and then the video signals mixed at the pixel-domain by using the reconstructed signal are encoded to produce the single mixed video signals. For the reduction of encoding process power, the motion vectors and coding modes are often reused in the encoder in the transcoders. Since this kind of transcoding requires the pixel-domain manipulation for the motion compensation and DCT (Discrete Cosine Tra-nsform), there is called pixel-domain transcoder. This means that the great amount of video processing power is necessary.

To reduce the extensive processing power in the pixel-domain, the bitstream-domain transcoding algorithms are proposed in [3, 4]. However, these algorithms are only limited to the still image bitstream coding based on the DCT. From the analysis in our previous work [5], we show that the compressed motion video signals can be mixed without the full decoding of video signals. In [5], we show that the macroblock in the video bitstream can be independently handled if the syntax elements coded with the lossless spatial DPCM method are properly re-encoded. However, the popular video codec, ITU-T Recommendation H.263 used in the videoconference has the constraint on the dynamic range of such syntax elements [6]. To solve the problem in [5], we introduced the additional header description for the macroblock located at the boundary of the mixed video signals, namely H.263 Annex K header description.

However, the most video decoders in the market adopt the H.263 baseline algorithm for the patents issues and computational complexity. In this paper, we propose a hybrid video mixing method based on the bitstream-domain and pixel-domain transcoding methods to remove the limitation of special header description in the bitstream-domain mixing method in [5]. For the left-sided sub-video signals, we used the bitstream-domain mixing method in [5]. The right-sided sub-video signal, which possibly contains the syntax violation in spatial DPCM, is transcoded with the conventional transcoding algorithm. The simulation results show that the proposed method reduces the computational complexity about 45% at the improved quality, compared with the conventional mixing method based on the pixel-domain transcoders.

This paper is organized as follows: in Section 2, several conventional video mixing techniques are discussed. A proposed hybrid method applied in video mixing will be presented in Section 3. And in Section 4 the experimental results of the proposed method compared with the conventional methods are shown. The conclusions are drawn in Section 5.

2. Conventional Video Mixing Methods

There reported several mixing methods for multipoint video conference. One of the direct techniques is pixel-domain mixing, which uses a transcoder [1, 2]. For instance, let's consider four-party videoconference as shown in Fig 1. The pixel-domain mixing method is depicted in Fig 2. Each of the four participants generates a H.263 QCIF bitstream by encoding its own frame sequence. The transcoder consists of four cascaded decoders, one pixel-domain mixer and one encoder. In the transcoder, each bitstream is decoded and reconstructed. After this, all of the four reconstructed QCIF video signals are mixed into a CIF video signal in pixel-domain. Following, the encoder in the transcoder re-encodes the new mixing video sequence into a CIF bitstream. At the terminals, each participator decodes the bitstream and sees a CIF video including all of the four QCIF participants in real time.

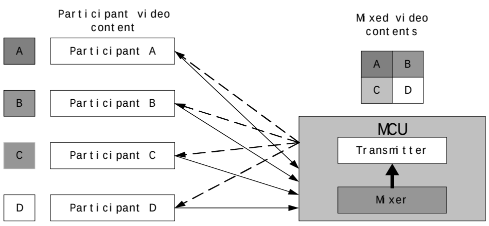


Fig. 1 Four-party videoconference

Pixel-domain transcoder mixing method is more flexible in terms of picture manipulation. However, this method has two major problems. The first is computational complexity. The transcoder system should independently decode the bitstreams and the raw video signals are mixed together. And the encoder re-encodes the mixed video signals. So, four decoders, one encoder and pixel-domain mixer are needed here. It requires large amount of expensive memory and decompression capacity. The other drawback in the transcoder is the quality degradation because the mixed video bitstream is a result of double encoding. Further, the re-encoding operation will cause more delay because of the steps of codec.

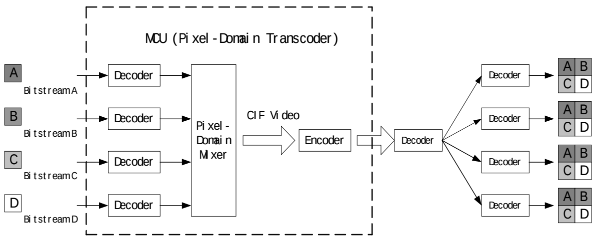


Fig. 2 Pixel-domain Transcoder mixing technique

3. Proposed Hybrid Mixing Method

To reduce the computational complexity, we proposed the bitstream-domain transcoder mixing method in [5]. Its basic structure is shown in Fig. 3. This method only needs to analyze the bitstream based on syntax of video coding standard and modify the control information without changing the video signal information in the video bitstream under certain constraints. Because this method is based on simple bitstream copy, it greatly reduces the computational complexity and saves process time. Further, it removes the re-encoder, so the quality of the output signals will be increased compared with the pixel-domain transcoder mixing method.

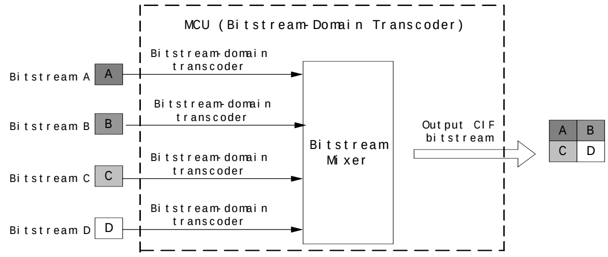


Figure 3 Structure of the Bitstream-domain video mixing method

In our previous work [5], we assume that the quantization step sizes of all the bitstreams are fixed since the difference of quantization step sizes between two successive macroblock in pane range is mixed by bitstream A and bitstream B shown in Fig. 4, both of which may have different quantizer. If the change of the quantizer exceeds the bound (-2 to 2) at the boundary of A and B, H.263 baseline codec cannot support the exact description of quantization step size and the incorrect dequantization causes quantization noise drift. To solve this problem, H.263 Annex K method is used for the description of the quantization step size [5]. In general, most video decoders in the market use the baseline codec of H.263 for the patents issues and computational complexity. For this reason, the constraints on the quantization step sizes should be solved in [5].

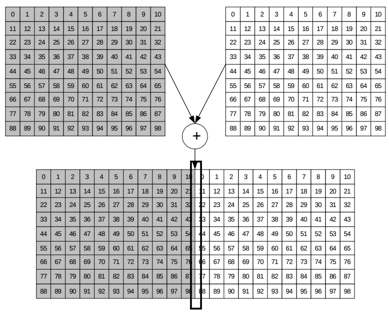


Fig. 4 Bitstream-domain mixing system of two bitstream (Horizontal mixing)

From the survey on the conventional methods, the pixel-domain transcoder mixing method reduces the quality of videos and also has high complexity of computation. Although the bitstream-domain mixing method requires low computation and also has good output quality, the demerit is Annex K of H.263 is necessary. It is not realistic for the mixing system because of the popularity of market and additional cost. To solve these problems, we propose a new hybrid video mixing method shown in Fig. 5. Through investigating, we know that bitstream-domain video mixing method is efficient in both complexity and visual quality. But for the bitstreams from B and D, we cannot apply the bitstream-domain video mixing method because of the quantization step sizes. Since the sub-pictures from A and C start from the left side of pictures, there is no constraint on the quantization step size. So we use the bitstream-domain transcoders for the bitstream A and C. And we use the pixel-domain transcoders for the bitstream B and D to solve the constraint on the quantization step size.

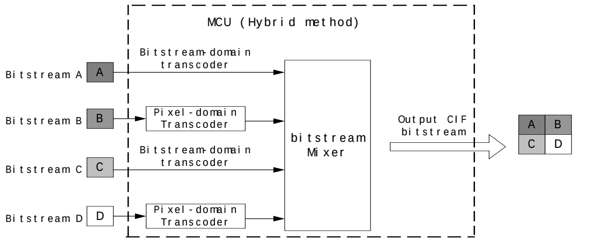


Figure 5 Structure of the proposed hybrid video mixing method

4. Simulation Results

For the evaluation of the proposed video mixing method, we perform some simulations, and all of simulations are done under the environments shown in table 1.

Table 1 Simulation Environments

Recommendation	Picture Name	Source Format	Encoding Bit Rate	Frame Number
H.263	A: Forman	QCIF, 176*144 4:2:0, 10kHz	64kbps	300
H.263	B: Akiyo	QCIF, 176*144 4:2:0, 10kHz	64kbps	300
H.263	C: News	QCIF, 176*144 4:2:0, 10kHz	64kbps	300
H.263	D: Silent	QCIF, 176*144 4:2:0, 10kHz	64kbps	300

4.1 Performance Comparison between Pixel-domain Transcoder and Bitstream-domain Transcoder

First we compare the performance between pixel-domain transcoder and bitstream-domain transcoder methods. QCIF bitstream "*Foreman*" listed in Table 1 is coded by these two methods, respectively. In Fig. 6, PSNR comparison shows that the video quality using pixel-domain Transcoder method is reduced due to the re-encoder. Further more, the computational complexity of pixel-domain transcoder is higher than bitstream-domain transcoder because of its double coding steps. The configurations of the simulation computer are: Pentium 4 CPU 2.89 GHz, 1.00GB memory. In Fig. 7, CPU time comparison proves the computation of pixel-domain Transcoder method is more complex. This is the reason that we select bitstream-domain mixing method as the main basic idea of the proposed hybrid mixing method.

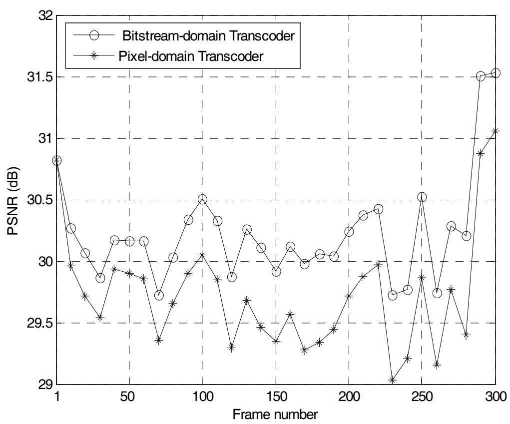


Figure 6 PSNR comparisons between Pixel-domain Transcoder and Bitstream-domain transcoder (QP = 15)

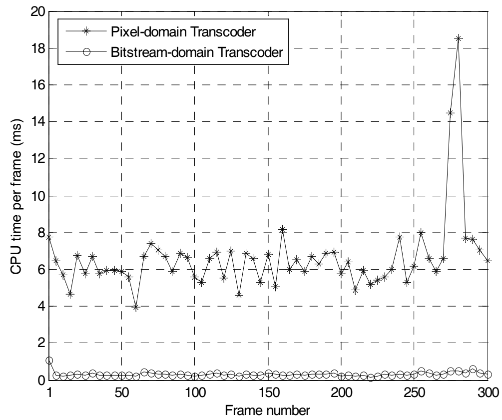
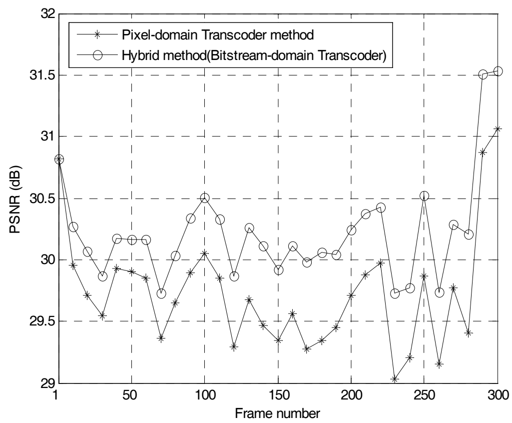


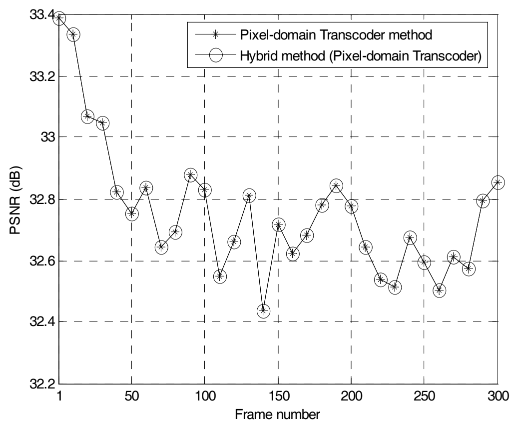
Figure 7 CPU time comparisons between Pixel-domain Transcoder and Bitstream-domain transcoder (QP = 15)

4.2 Performance Comparison of Various Mixing Methods

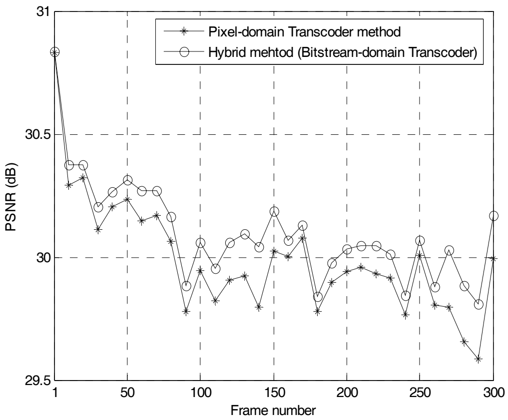
Then, we do the simulation with two mixing methods: pixel-domain transcoder mixing method and proposed hybrid mixing method. Figure 8 give the PSNR comparison of the two video mixing methods. In the proposed method, up-left and down-left parts use bitstream-domain transcoder mixing method and upright and downright parts use pixel-domain Transcoder mixing method.



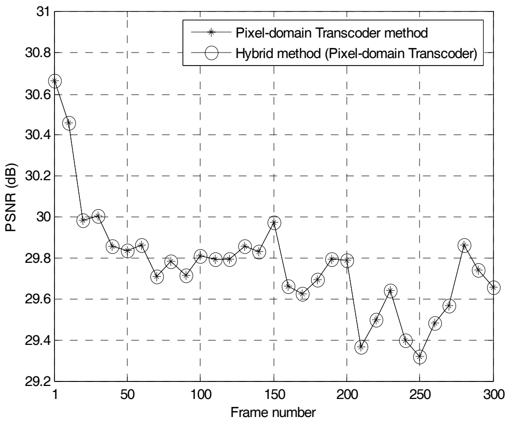
(a) Foreman (Bitstream-domain Transcoder method)



(b) Akiyo (Pixel-domain Transcoder method)



(c) News (Bitstream-domain Transcoder method)



(d) Silent (Pixel-domain Transcoder method)

Figure 8 PSNR Comparison between proposed and pixel-domain transcoder mixing method (QP = 15)

From figure 8 we can know that the quality of up-left and down left output sub-pictures utilizing proposed hybrid-mixing method (A and C) are better than that using pixel-domain transcoder method respectively. On the other hand, the quality of up right and downright output sub-pictures are very similar with that using pixel-domain transcoder mixing method respectively. With the same simulation computer, Table 2 shows the CPU time for coding out each CIF picture by the two methods, respectively. It means, that the proposed method is more efficient than the pixel-domain transcoder method, and it reduces the computational complexity about 45%. Further more, although the proposed method is less efficient than the method of bitstream-domain video mixing proposed in [5], it does not use H.263 Standard Annex K to deal with the problem of DQUANT. For this reason, the proposed method in this paper is more suitable to implement.

Table 2 Comparison of average computational complexity in CPU time

Method / Compare option	Pixel-domain Transcoder method	Proposed method
CPU time (Microsecond)	17.573	9.65

5. Conclusion

In this paper, we propose a hybrid method of video mixing for the videoconference system. From the study of video mixing techniques we know that the conventional video mixing method that based on pixel-domain transcoder requires the higher computational complexity: four decoders and on encoder with four times encoding speed and memories. There exist additive delay due to the steps of four-decoder→mixer →one -encoder. And this method also reduces quality of video because it is a result of double encoding. Although the bitstream-domain mixing method is smart, it requires H.263 Standard's Annex to prevent DQUANT error. To reduce the computational complexity, we propose a hybrid mixing method based on the bitstream-domain bitstream modification and pixel-domain transcoder. The proposed method reduces the computational complexity about 45% at the improved quality, compared with the conventional mixing method based on the pixel-domain transcoders. Actually, the method proposed in this paper is just a primary idea of the authors; we will research on this field deeply in the future. We believe that this video mixing system is valuable for implementation such as videoconference system, video chatting and so on.

References

[1] A. Vetro, C. Christopoulos, and H.Sun, "Video transcoding architecture and techniques: an overview", *IEEE Signal Processing Magazine*, vol. 20, no. 2 pp. 18-29, March 2003;
[2] M.Lukacs, "An Advanced digital network video bridge for multipoint with individual customer control," *Bell Commun, Res.*, NJ. May 1992;
[3] T.Shanableh and M.Ghanbari, "Transcoding architectures for DCT-Domain heterogeneous video transcoding" *Pro. Of IEEE Int'l Conf. of Image Processing*, vol. 1, pp. 433-436, Oct.2001
[4] S. -F. Chang and D.G. Messerschmitt, "Manipulation and compositing of MC-DCT compressed video," *IEEE Journal of Selected Areas in Communications (JSAC)*, vol. 13, no. 1, pp. 1 – 11, Jan. 1995;
[5] Kook-yeol, Yoo, "Syntax-based mixing method of H.263 coded video bitstreams," Proc. Of *IEEE Int'l Conf. of Consumer Electronics*, p. 9.18, Jan. 2005;
[6] ITU-T Recommendation H.263, "Video coding for low bit rate communication", Nov. 95;
[7] Niklas Bjork, Charilaos Christopoulos, "Transcoder architectures for video coding," *IEEE Transactions on Consumer Electronisc*, vol. 44, no. 1, pp. 88-98, Feb. 1998;
[8] K.R. Rao and J.J. Hwang, *Techniques & Standards for Image, Video, and Audio Coding*, Pretice Hall, 1995
[9] W. H. Chen. C. H. Smith, and S. Fralick, "A fast computational algorithm for the discrete cosine transform," IEEE Trans. Commun, vol. COM-25. pp. 1004-1009, Sept. 1977.
[9] Khanlid Sayood, Introduction to data compression, 2nd ed., Academic Press, 2000
[10] W.H.Chen.C.H.Smith, and S.Fralick, 'A fast computational algorithm for the discrete cosine transform," IEEE Trans. On Commun., vol.25.[[.1004-1009, Sept. 1977
[11] R. M. Haralick, "A storage efficient way to implement the discrete cosine transform."*IEEE Trans. Computer,* vol. C-25, pp. 764-765,July 1976.
[12] R. C. Reininger and J. D. Gibson. "Distributions of the two Dimensional DCT coefficients for images." *IEEE Trans. Commun,* vol. COM-31, pp. 835-839. June 1983.