IMPLEMENTATION OF H.323 MULTIPOINT VIDEO CONFERENCE SYSTEMS WITH PERSONAL PRESENCE CONTROL

Chia-Wen Lin, Wen-Hao Wang
Computer and Communication Research Labs
Industrial Technology Research Institute
Hsinchu, Taiwan 310, ROC
{ljw,wwh}@n100.ccl.itri.org.tw

Ming-Ting Sun, and Jeng-Neng Hwang
Department of Electrical Engineering
University of Washington
Seattle, WA 300, USC
{sun,hwang}@ee.washington.edu

ABSTRACT

This paper presents an H.323 standard compliant video conference system architecture. Both the server side and client side implementations are addressed. Video processing for advanced personal presence features including dynamic rate allocation and video object manipulation are also presented.

1. INTRODUCTION

Video conferencing is a natural extension to voice-only communications. Due to the rapid progress in digital video processing, low-cost video codecs, networking technologies, and international standards, video conferencing is becoming more and more widely used in business and home applications. ITU-T H.323 [1] is the most widely adopted standard for multimedia communications over packet-switched networks worldwide. H.323 video conference systems are thus very attractive in current commercial market, for example, the Microsoft Netmeeting is the most popular PC-based point-to-point H.323 video conference system available now. We have developed several advanced techniques in enhancing the video quality in video telephony applications, especially multipoint video conferencing [3-5]. In this paper, we will show our implementation of H.323 standard compliant video conference systems incorporating the advance video transcoding techniques proposed in [3-5] which is featured with user friendly personal presence control features such as dynamic bandwidth allocation, object resizing, re-location, and manipulation, etc.

2. THE PROPOSED SYSTEM ARCHITECTURE

The proposed PPMCU is a PC-based prototype for multipoint video bridge and gateway which performs the multimedia communications and conversion and protocol translation among multiple LAN and WAN terminals. Here the Ethernet and ISDN are selected as the target networks for LAN and WAN users respectively. Fig. 1 depicts the architecture of the proposed H.323 Personal Presence Multipoint Control Unit (PPMCU). As shown in this figure, the proposed PPMCU cannot only serve as a Multipoint Control Unit but also play the role of a gateway to interwork between the H.323 (for LAN) and H.324 (for WAN) equipments. The video conference equipments are divided into two sides: the server side and the client side. The client side is basically an H.323/H.324/I video terminal, which generates an H.323/H.324/I-compliant bit-stream with integrated audio, video, and data content. The PPMCU server receives and terminates the bit-streams from the H.323 and H.324/I client terminals as well as performs the Multipoint Control (MC) and Multipoint Processor (MP) functions. In fact, the PPMCU server also includes the complete functions of the H.323/H.324/I client side terminals. The advance video transcoding functions are also performed in the PPMCU server.

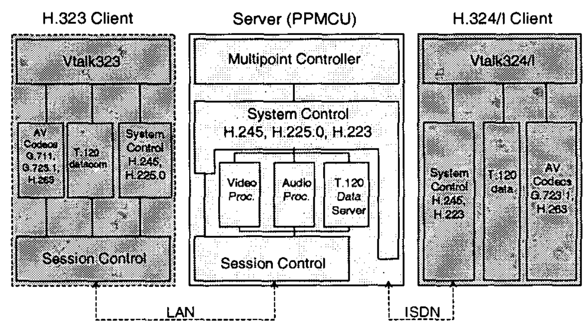


Fig. 1 The proposed architecture of H.323 PPMCU

Fig. 2 shows the baseline mode user interface of the video function of the PPMCU. It supports the split-screen display format so that four sub-windows can be seen simultaneously in a continuous presence fashion. A novel dynamic bit allocation scheme proposed in [3] is used to allocate bit-rates to the sub-windows in transcoding by taking into account the joint spatial-temporal complexity of each sub-window as follows:

$$B_m^* = S_m^{ST}\left[B_{\text{target}} - AN\sum_{m=1}^{M} C_m\right] + ANC_m, \quad m = 1, \ldots, M \tag{1}$$

where

$$S_m^{ST} = \alpha_\sigma \frac{S_m^\sigma}{\sum_{m=1}^{M_i} S_m^\sigma} + \alpha_{\text{MV}} \frac{S_m^{\text{MV}}}{\sum_{m=1}^{M_i} S_m^{\text{MV}}} + \alpha_{\text{MB}} \frac{S_m^{\text{MB}}}{\sum_{m=1}^{M_i} S_m^{\text{MB}}} \tag{2}$$

$$S_m^\sigma = \sum_{n=1}^{N} \alpha_{m,n}\sigma_{m,n} \tag{3}$$

$$S_m^{MV} = \sum_{n=1}^{N} \left|MV_{m,n}^x\right| + \left|MV_{m,n}^y\right| \tag{4}$$

M is the number of sub-windows (e.g., four in our case). S_m^{MV} is the sum of the magnitudes of the motion vectors in the mth sub-

0-7803-6301-9/00 $10.00

window and s_m^{MB} is the number of the coded macroblocks (i.e. the COD bit is "0") in the *m*th sub-window of the incoming video. The weighting factors $(\alpha_\sigma, \alpha_{MV}, \alpha_{MB}) \in [0,1]$ and

$$\alpha_\sigma + \alpha_{MV} + \alpha_{MB} = 1 \qquad (5)$$

Because most of the time, only one or two conferees are active at one time, the proposed dynamic bit allocation method can achieve significant improvement on the visual quality as shown in Fig. 3.

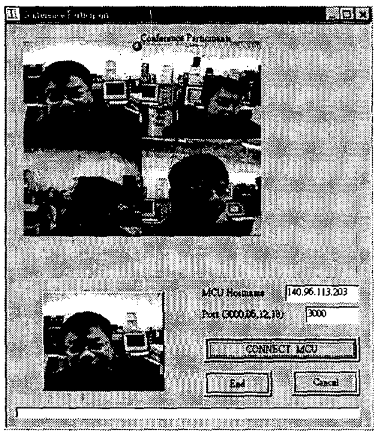
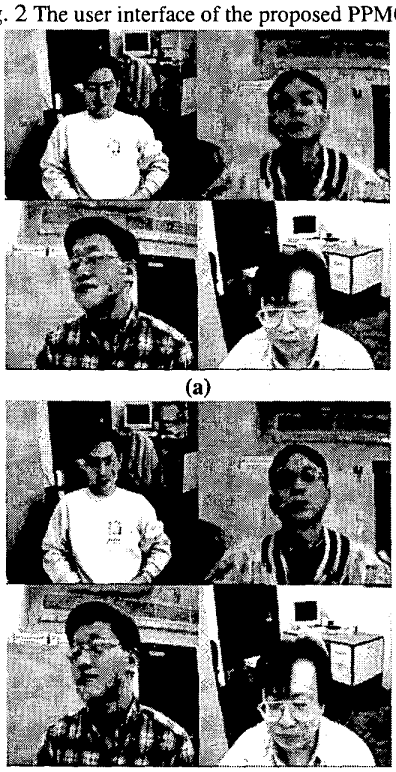


Fig. 2 The user interface of the proposed PPMCU



(a)

(b)

Fig. 3 Subjective performance comparison (a) Frame #136 without intelligent dynamic bit allocation; (b) Frame #136 with intelligent dynamic bit allocation. The active conferee on the upper right sub-window is drastically improved.

3. VIDEO TRANSCODING WITH ADVANCED PERSONAL PRESENCE CONTROL

The proposed PPMCU also supports several advanced personal presence control features. As shown in Fig. 4, we have implemented a chroma-key based approach to allow object extraction in the PPMCU server so that the video objects can be manipulated according to the users' need. After extracting the video objects, the server transcodes the video objects using the intelligent bit allocation method described above and inserts a chroma-key background again. The resultant bit-stream still conforms to the H.263/H.263+ standard [2]. The client side subsequently extracts the video objects from the received bit-stream and manipulates the 2-D video objects against a 3-D virtual background according to the 3-D loçation information of each conferee received from the server. The object manipulation functions are implemented using the OpenGL™ technologies which are supported by many commercial 3-D display card at the client side. Thus the computation load is shared, thus real-time processing is achievable. Fig. 4 shows an example of 2-D video objects compositing in a 3-D virtual background.

Fig. 4 Personal presence control features: virtual meeting with natural 2-D objects in synthetic 3-D background

4. REFERENCES

[1] ITU-T Recommendation H.323, "Visual telephone systems and terminal equipment for local area networks which provide a non-guaranteed quality of service". 1998.

[2] ITU-T Recommendation H.263, "Video codec for low bit-rate communication". 1996.

[3] T.-D. Wu, J.-N. Hwang, and C.-W. Lin, "Dynamic bit-rate conversion and bit reallocation in multipoint video transcoding," submitted to *IEEE Trans. Circuits Syst. Video Technol.*, May 1983.

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

[5] C.-W. Lin, "A study on video transcoding for multipoint video conferencing," *Ph.D. Dissertation*, National Tsing Hua Univ., Taiwan, R.O.C. Jan. 2000