Hindawi Publishing Corporation
Advances in Multimedia
Volume 2007, Article ID 97535, 8 pages
doi:10.1155/2007/97535

Research Article

Interactive Multiview Video Delivery Based on IP Multicast

Jian-Guang Lou, Hua Cai, and Jiang Li

Media Communication Group, Microsoft Research Asia, Beijing 100080, China

Received 31 October 2006; Accepted 21 December 2006

Recommended by Jianfei Cai

As a recently emerging service, multiview video provides a new viewing experience with a high degree of freedom. However, due to the huge data amounts transferred, multiview video's delivery remains a daunting challenge. In this paper, we propose a multiview video-streaming system based on IP multicast. It can support a large number of users while still maintaining a high degree of interactivity and low bandwidth consumption. Based on a careful user study, we have developed two schemes: one is for automatic delivery and the other for on-demand delivery. In automatic delivery, a server periodically multicasts special effect snapshots at a certain time interval. In on-demand delivery, the server delivers the snapshots based on distribution of user requests. We conducted extensive experiments and user-experience studies to evaluate the proposed system's performance, and found that it provides satisfying multiview video service for users on a large scale.

1. INTRODUCTION

With the rapid development of electronic and computing technology, multiview video has recently attracted extensive interest due to greatly enhanced viewing experiences. For example, a system called EyeVision [1] was employed to shoot Superbowl 2001. Other systems, such as Digital Air's Movia [2] and Wurmlin's 3D Video Recorder [3], were also proposed to capture multiview video. Later, we proposed an interactive multiview video system (IMV System) for serving real-time interactive multiview video service [4]. Unlike conventional single-view video systems, a multiview video system allows the audience to change view direction and to enjoy some special visual effects such as view switch and frozen-moment. It greatly enhances user experience in interactive and entertainment orientated applications.

As a recently emerging service, multiview video provides a new viewing experience with a high degree of freedom. However, it also brings challenges to data delivery due to the huge data amounts transmitted. Hence, an interactive unicast solution was adopted by the previous IMV system to support a high degree of interactivity. However, unicast cannot meet the requirements of an increasing number of users due to restricted network bandwidth and limited server-processing capability. Different from the conventional unicast streaming, IP multicast is a promising technology that can handle users on a large scale. Many researchers have been investigating this area in the last decade. Among efforts is that work in VoD systems [5]. Cooperating with some delivery policies, such as command batching [6, 7] and video patching [8, 9], VoD multicast systems can provide users near VoD service and keep relatively low bandwidth costs.

When IP multicast technology is used for implementing a multiview video delivery system, interactivity has become an important issue since multiview video has unique features. In this paper, we proposed a multiview video multicast system to support a large number of users and a high degree of interactivity. Based on a detailed user study, we developed two schemes, one for automatic delivery and the other for on-demand delivery. In the automatic delivery, the server periodically multicasts special effect snapshots at a certain time interval. And, in the on-demand delivery, the server delivers the snapshots based on the distribution of user requests. The proposed system was also evaluated by extensive experiments and user-experience studies.

The rest of the paper is organized as follows. In Section 2, we outline the overall structure of multicast IMV system, and we present the video delivery schemes of our conventional and special effect videos in Section 2.2. Some experimental results are presented and discussed in Section 3. In Section 5, we conclude our work.

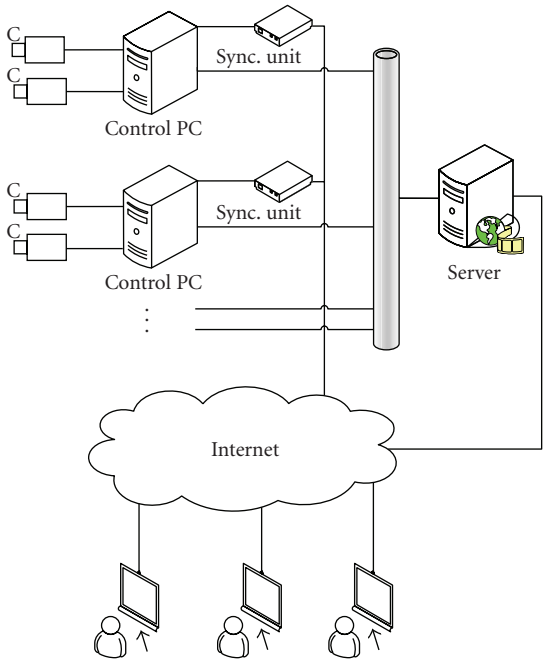


Figure 1: System architecture.

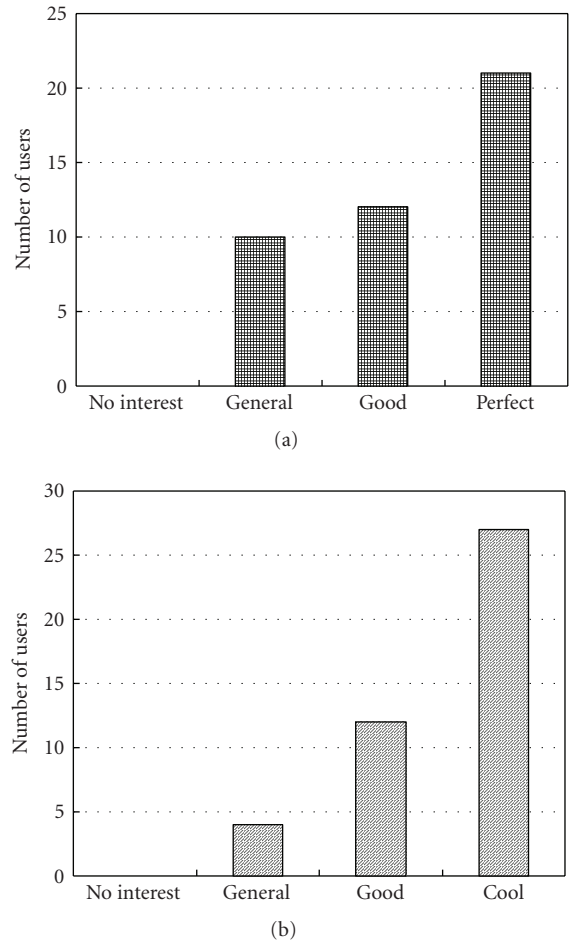


Figure 2: User study results on (a) view switching and (b) frozen moment.

2. VIDEO DELIVERY SCHEME

2.1. System overview

As described in Figure 1, our IMV system mainly consists of K video cameras, K pan-tilt units, a set of control PCs and synchronization units, a server, a network backbone, and many receivers (clients). These components can be classified into three parts [4]: capturing part, server part, and client part. Based on the synchronization signals generated by sync units, the capturing part acquires the same dynamic event simultaneously from multiple cameras with various view directions. The captured video signals are compressed in several control PCs and then sent to the server through a network backbone, for example, a gigabit Ethernet. The server part collects both the K compressed video streams from the control PCs and transcoded special effect snapshots from the transcoding servers. It then provides a multiview video service to end users.

Interactive special effects such as *frozen moment*, *view sweeping*, and *view switching* are three important features of our IMV system. In the frozen moment, time is frozen and the camera view direction rotates about a given point, while view sweeping involves sweeping through adjacent view directions while time is still moving (please refer to Figure 2 of [4] for a detailed description). View switching means that users are able to switch from one camera view direction to another as the video continues along time. Through a usability study, we found that users were highly interested in these new features. In the study, more than 40 people were invited as participants, including people with technical and nontechnical backgrounds. The results are summarized in Figure 2. Figure 2(a) indicates that about 75% of the participants consider view switching is a useful feature in an IMV system. Meanwhile, in Figure 2(b) about more than 90% of them consider that the frozen moment effect is an interesting feature.

Based on these observations, we mainly focused on how to provide multiview video features for users based on IP multicast techniques. The main challenge is to support large scale users with high interactivity using relatively low server bandwidth.

2.2. Multicast video delivery

The multiview video consists of not only conventional live videos from different views, but also the special visual effects mentioned in Section 2.1. In general, the server should broadcast videos of conventional views and special visual effects. Users should be able to enjoy special visual effects at any time, and the effects should be rendered immediately after users' actions. This application poses two requirements: QoS guaranteed transmission and low-delay interactivity. However, lower latency and more flexible action often result in higher bandwidth cost, especially when the number of views

becomes large. To handle this problem, we prepared two kinds of video streams at the server side. The first one is the conventional single-view video stream that is captured and compressed individually at the control PC. The other kind of streams is the frozen moment stream and the view-sweeping stream. These video contents are delivered through $M+N$ video multicast channels. Here, the M channels are assigned to multicast the videos from different views, while the N channels are used for delivering the special effect streams. The values of M and N depend on the available bandwidth of the server. As shown in Figure 3, each client simultaneously joins one conventional video channel and one or more special effects channels. The number of the special effects channels that a user joins depends on available downlink bandwidth. Therefore, users with higher available downlink bandwidth can join more special effects channels, and thus can enjoy the special effects with higher degree of interactivity.

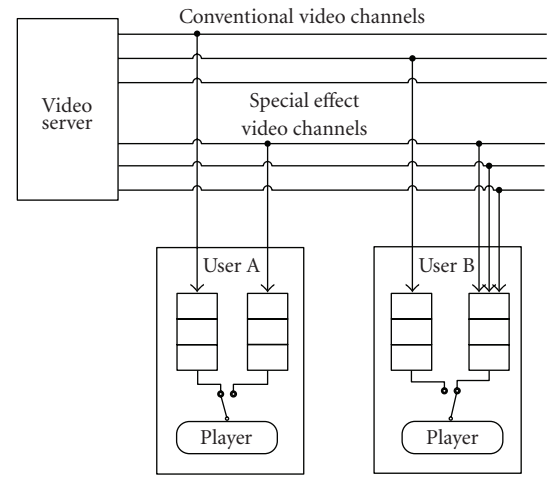


Figure 3: The overview of IP multicast for online user service.

2.2.1. Conventional video channels

One problem of designing the proposed delivery system is how to select views for the M conventional view channels. The simplest solution is that all K views are broadcast through $M = K$ channels to serve users' subscriptions. However, in real world scenarios, we found that the number of conventional view channels M is not necessarily the same as the number of views. Because of the small visual difference between two adjacent views, users are unlikely to do a switch operation between them. In our experiments, we found that $M = 6$ can usually meet user requirements for a total capture angle of 90°. It is only about 1/5 of the original view number (32 in our system). Such a sampling scheme can largely reduce the server bandwidth usage. Given the server available bandwidth, the saved bandwidth can be used to improve the interactivity of special effects.

In our system, view switching is realized by switching from the source to the destination conventional channel. The maximum latency of the view switching is $T_s + T_v$, where T_s is the time of network channel switching and T_v is the latency from the current frame to the next I frame. The value of T_v is determined by the group of picture (GOP) size when compressing the conventional view videos. Figure 4 shows two conventional video channels, view i and view j. Because each conventional video stream is compressed as I and P frames using an MPEG-liked encoder, to guarantee the smoothness of the visual experience, we can only switch one view to another at an I frame. For example, if a user sends a switching command from view i to j at time index $t_0 = t + 1$, it immediately joins channel j and receives video frames following $P_j(t+1)$ from channel j. At the same time, it still receives P frames following $P_i(t+1)$ from channel i until it leaves channel i when the next I frame $I_j(t+g)$ arrives. In our system, the GOP size is set to 30 and the video frame rate is 30 fps. Thus the maximum value of T_v is 1 second and the average value is 0.5 second. In [4], we have found that users can tolerate a relatively long, for example, 1 second, view switching latency, which is similar to the consumers attitude on the program

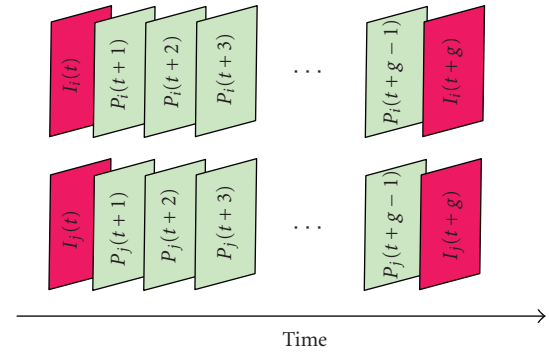


Figure 4: Conventional video channels. Here g is the value of GOP size.

switching latency in digital TV. Therefore, such a delay can meet most users' requirements.

2.2.2. Special effects channels

Due to the limited downlink bandwidth, users cannot get all the special effects snapshots in real time. Fortunately, through the user study in [4], we found that different users often have a similar sense of exciting moments in a multiview video, and they will subscribe to special effects when there is an exciting moment. Furthermore, the visual experiences of neighboring snapshots in the time domain are close to one another. This means that not all snapshots need to be sent to end users. Then the problem is that, for a given available downlink bandwidth, how to select proper special effects snapshots for end users?

For an offline multiview video, suppose that the distribution of all user subscriptions $f(t)$ on special effects snapshots are known beforehand. Then, we can find the optimal

snapshots p_i $(i = 1, \ldots, n)$ by minimizing the total differences from the snapshot that a user wants:

$$\arg\min_{p_i} \left(\sum_{i=1}^{n} \int_{\phi_i}^{\phi_{i-1}} (t - p_i)^2 f(t) dt \right). \quad (1)$$

Note that (1) is very similar to the classic scalar quantization problem. The iterative method proposed by Lloyd [10] can be used to estimate the optimal values of p_i. However, in online multiview streaming, the distribution function $f(t)$ cannot be known in advance. In other words, we are not able to determine the proper snapshots beforehand.

A simple strategy, named as automatic delivery scheme, is that the special effects channels multicast the snapshots with a fixed time interval d_s ($d_s \geq T$, $T = b/B$ is the minimal time interval of sending a snapshot that is determined by the average snapshot size b and the available bandwidth B). In the automatic delivery, all of the sent snapshots are equally distributed in the special effects channels. Obviously, the disadvantage is that the sent snapshots may not be the ones that most users subscribe, because there is no interactivity between users and the server.

To overcome the problem in the automatic delivery, we also design an on-demand delivery scheme that takes user subscriptions into consideration. In the on-demand delivery, the server collects user requests and fetches an appropriate snapshot for most users. The snapshot will be sent when both of the following formulas are satisfied:

$$\begin{aligned} C &\geq \tau \times S, \\ T_\eta &\geq T, \end{aligned} \quad (2)$$

where T_η is the time interval between the sent time of the two snapshots, C is the sum of user requests in the period of T_η, τ is a threshold ($0 \sim 100\%$), S is the total number of logon users, and $T = b/B$.

To better illustrate the process of our on-demand delivery scheme, we give an example in Figure 3. The curve is the distribution of user requests. t_0 is the sent time of the last snapshot, t_s is the snapshot ordered by a user request x, t_1 is the sent time of the current snapshot, and t_c is a proper snapshot for most users. In the on-demand delivery, the special effects video service tries to meet the interaction requirements of most users. It can dynamically adjust the sending frequency based on the number of user requests and the threshold τ. Therefore, more bandwidth cost can be saved when there are fewer requests. However, the disadvantage is that it brings extra interaction latency for the server needs to collect user requests.

To demonstrate the performance and features of the system, we carried out streaming experiments and user-experience studies. The results can help us select a proper video streaming strategy.

3. EXPERIMENTS

In this section, we describe the experiments on the performances of the automatic delivery scheme and the on-demand delivery scheme under various network conditions.

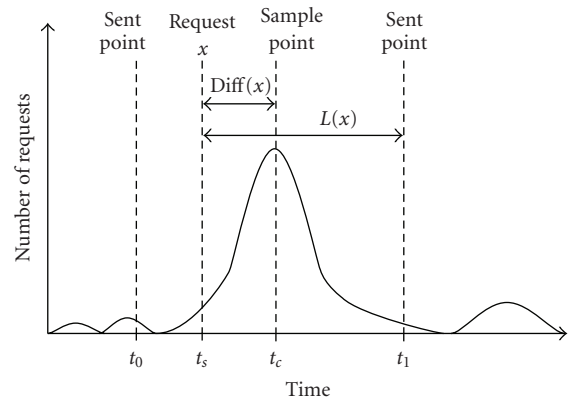


Figure 5: On-demand delivery scheme. Here, t_0 is the sent time of the last snapshot, t_s is the snapshot ordered by a user request x, t_c is the snapshot of most users, t_1 is the sent time of the current snapshot.

3.1. Performance metrics

Before the experiments, we first figured out two metrics that can be used to evaluate the system performance. Here are the definitions.

Special effects latency $D(x)$

Special effects video latency is the time interval from the moment that users send out requests to the moment that the special video starts to be played. In Figure 5, $L(x)$ is the time interval from the request time t_s to the sent time t_1. If t_n is the network RTT, the latency of the command x should be $D(x) = L(x) + t_n$, because a user sends command x at $t_s - t_n/2$, while receives the response at $t_1 + t_n/2$.

Special effects difference $\mathrm{Diff}(x)$

Special effects video difference is the time difference between the effects snapshot the server sends out and the one requested by a user. For example, in Figure 5, $\mathrm{Diff}(x)$ is the video difference of the command x from the time stamp t_s of the snapshot that a user requests to the time stamp t_c of the snapshot that the server sends out.

3.2. User-experience study

Even given the values of latency and difference, we still have no clear knowledge about whether they can meet most user requirements. Therefore, it is necessary to study user experiences on various values of latency and video difference. In this paper, we conducted a user experience study, which is formed from the feedback from 43 users after they observed the videos (including Chinese martial arts and gymnastics) with different latency and difference values. The result is shown in Figure 6, where the height of a bar represents the number of users who consider the interactivity with corresponding latency and difference as acceptable.

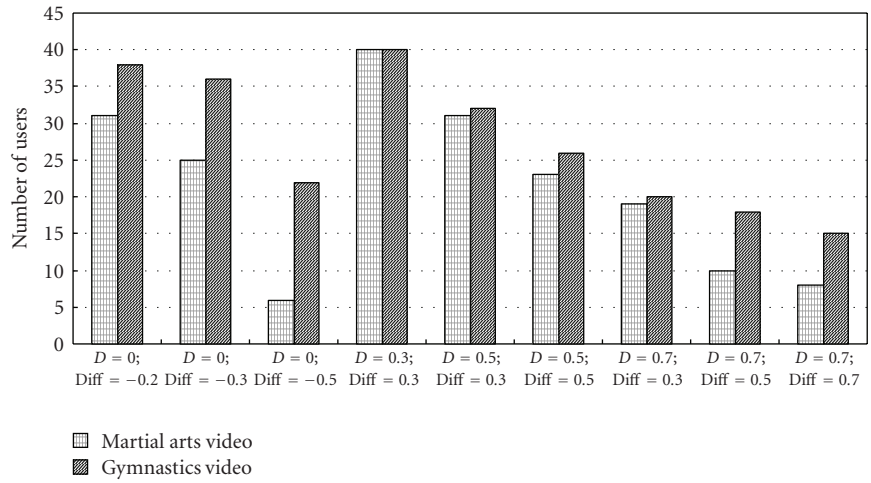


Figure 6: User study on latency and difference.

From Figure 6, we find out that more than 90% of users consider the performance of special effects video as very good when the difference and latency are set to 0.3 seconds. Less than 15% of participants can tolerate the 0.7 seconds latency and video differences. And the configuration of 0.5 seconds latency and 0.3 seconds difference is also acceptable. Most users felt that the latencies and differences in Chinese Martial Art videos are not as comfortable as they are in the gymnastics videos. This means that user responses to different video contents are slightly different. Based on the results, we find out that for a practical system, the latency and difference should be less than 0.5 and 0.3 seconds.

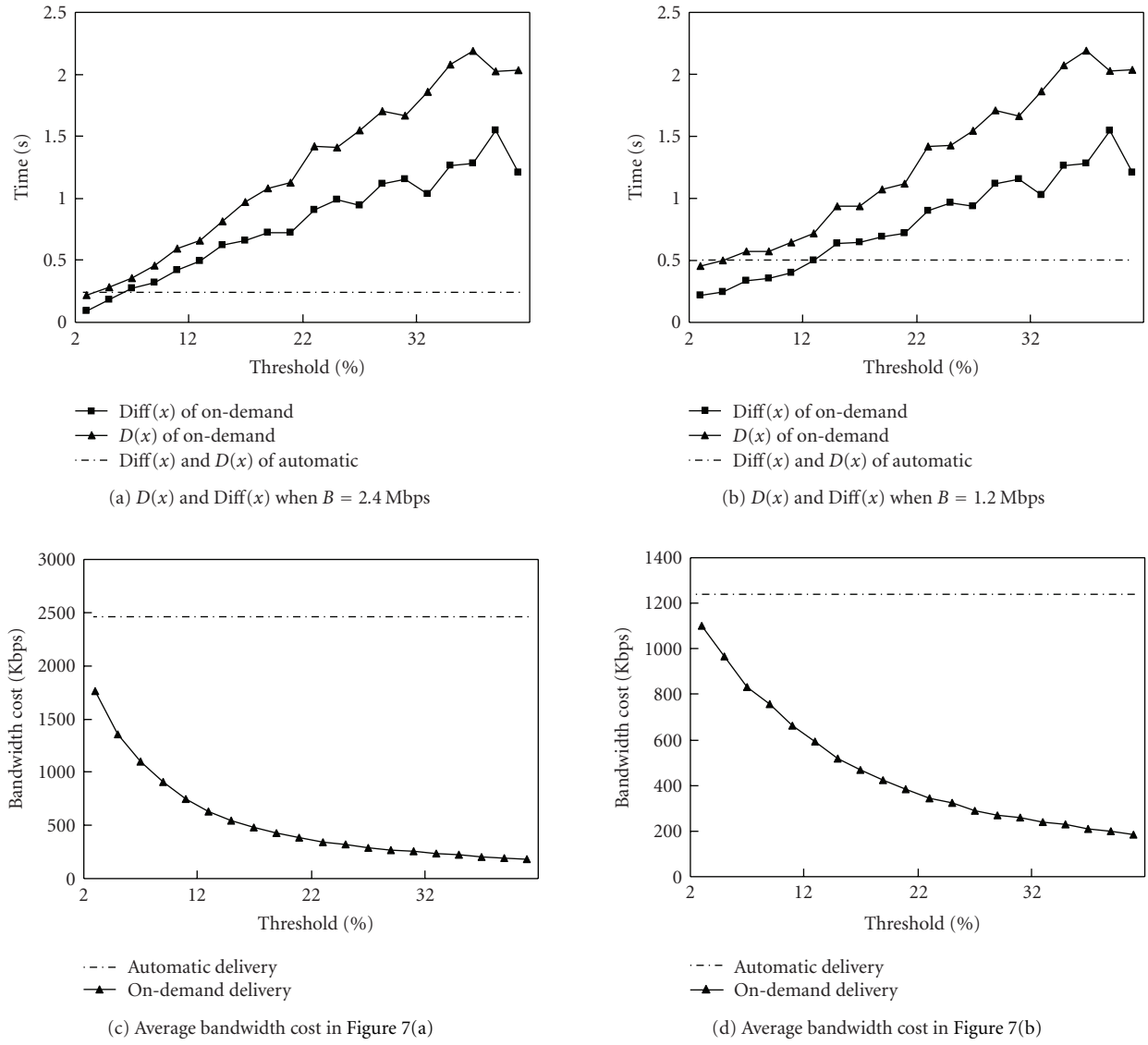
3.3. Experiments on special effects video delivery

Figure 7 shows the results of two delivery schemes with different available bandwidth. The experiments were carried out in a LAN with capacity of 100 Mbps. The run trip time (RTT) is less than 10 milliseconds, and can be neglected in our experiments. Although sometimes the available bandwidth of the LAN is large enough, we use 1.2 or 2.4 Mbps to deliver the special effects video in our experiments. In Figures 7(a) and 7(b), two straight dotted lines are the average values of $D(x)$ and $\mathrm{Diff}(x)$ in the automatic delivery scheme, while the two curves are the average values of $D(x)$ (the curve with triangle points) and $\mathrm{Diff}(x)$ (the curve with quadrate points) as the threshold increases in the on-demand delivery scheme. Figures 7(c) and 7(d) are the corresponding average bandwidth costs of the two schemes. As shown in Figure 7(a), the latency and difference of the automatic delivery and the on-demand delivery are very close when we set a very small threshold (e.g., $\tau = 5\%$). Meanwhile, Figure 7(b) shows that the on-demand delivery scheme ($\tau < 12\%$) will have a smaller difference than the automatic one when the bandwidth is relatively small (e.g., $B < 1.2$ Mbps). The reason is that the sent snapshots are selected to meet the subscriptions of most users in the on-demand delivery. Furthermore, from Figures 7(c) and 7(d), we learn that the on-demand delivery scheme can largely reduce the average downlink bandwidth request. This is because that, in the on-demand delivery, snapshots are only sent when there are user requirements in the system. It seems that if a system has large available downlink bandwidth (e.g., $B > 2.4$ Mbps), both schemes are able to meet user requirements, but the automatic scheme is better because the server does not have to manage any user request. On the other hand, the on-demand scheme will be a better choice when the downlink bandwidth is less than 2.4 Mbps, due to lower latency and differences. Finally, we want to point out that, although the results in Figure 7 come from the videos of Chinese Martial Arts, we can draw a similar conclusion from the results of gymnastics videos which are presented in Figure 8.

4. RELATED WORK

Multiview video has attracted lots of interests from both industry and academy. Most of previous research efforts focused on multiview video compression [11–14]. For example, in [15], the authors proposed a coding structure that can facilitate a free viewpoint switching operation. However, only a few works in previous research literature have discussed the multiview video delivery problem.

In [16], Kimata et al. designed a free viewpoint video system based on RTP and RTSP protocols. In their system, multiple video frames are transmitted as a single elementary stream, and several view frames that have the same time stamp are multiplexed into one RTP packet. Viewpoint prediction is adopted to reduce the average action delay. Recently, Liu et al. [17] proposed an interactive dynamic light field transmission system. In their system, each producer PC sends four compressed streams to a streaming server. Based on the subscription of each client, the server selects some streams and transmits them through a transmission channel. Unfortunately, both of the systems are based



(a) $D(x)$ and Diff(x) when $B = 2.4$ Mbps

(b) $D(x)$ and Diff(x) when $B = 1.2$ Mbps

(c) Average bandwidth cost in Figure 7(a)

(d) Average bandwidth cost in Figure 7(b)

Figure 7: Average latency and difference of a Kongfu video.

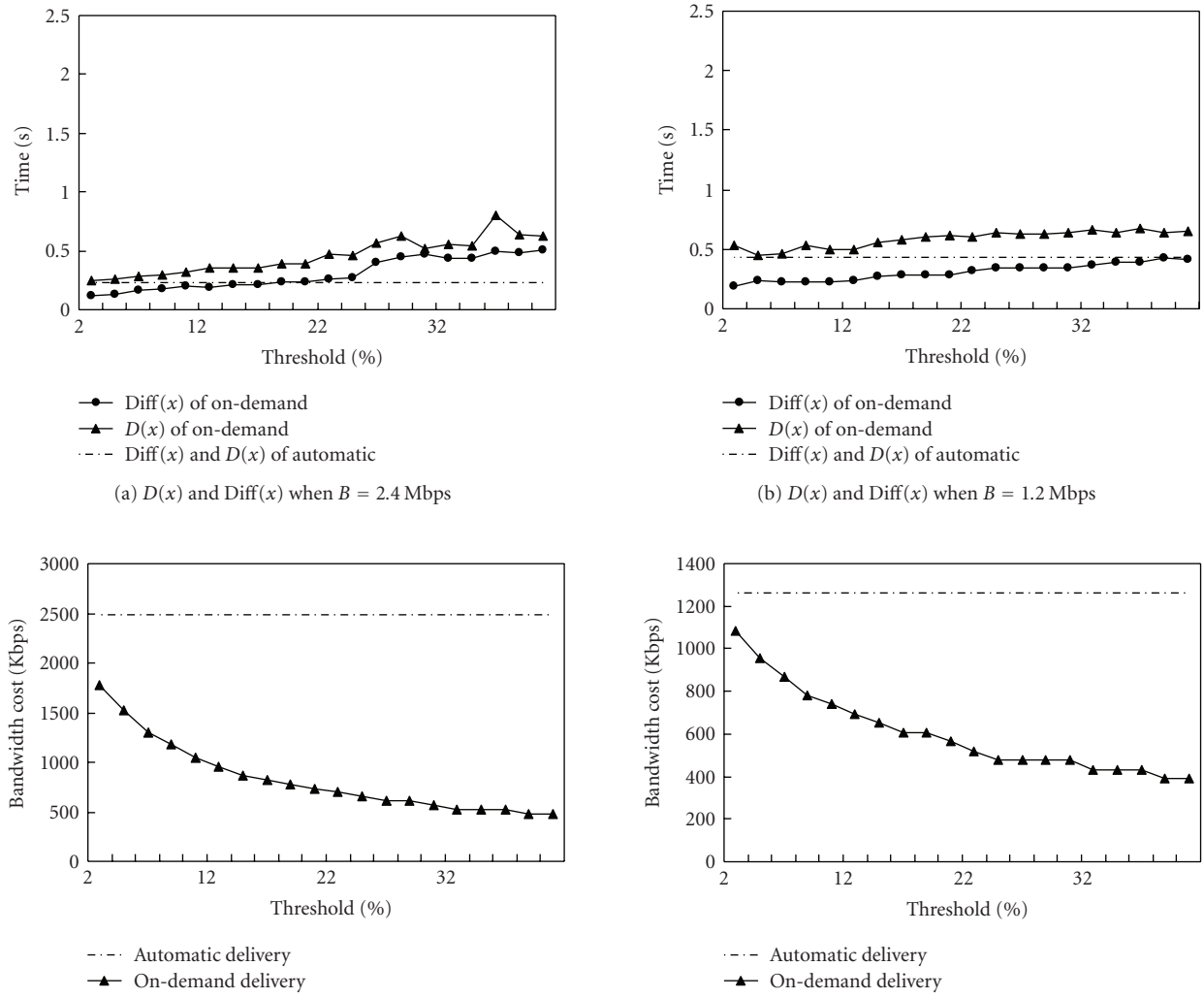
on unicast transmission schemes. Although they can provide a free viewpoint video service with high interactivity, such systems cannot scale to support a large number of users.

Matusik et al. [18] have designed a scalable system for real-time acquisition, transmission, and display of dynamic 3D scenes. In their system, each video stream is compressed individually using an MPEG2 encoder, and then is broadcasted to consumers. Unlike our real-time interactive multiview system, their system does not support users' interactivity. In their system, users cannot enjoy special visual effects.

Video streaming based on multicast has been proved useful for supporting large numbers of users. Kurutepe et al. [19] proposed a multiview video streaming system based on an application level multicast protocol, NICE [20]. Both original views and depth videos are compressed using the H.264 video codec. Clients selectively join different streams based on their available bandwidth and viewing angles. However, their system does not support the frozen moment and view sweeping effects.

5. CONCLUSION

In this paper, we propose a multiview video streaming system based on IP multicast. The multiview videos are transmitted through $M + N$ video channels. This multiple-channel scheme can support various users who have different available bandwidth. Furthermore, two multicast delivery strategies, automatic delivery and on-demand delivery, are presented and evaluated in this paper. Based on the proposed streaming schemes, our system can serve users on a large scale, and provide satisfying interactivity for most of them. Moreover, the analysis can also facilitate selecting proper streaming strategy for different multiview video applications.



(a) $D(x)$ and Diff(x) when $B = 2.4$ Mbps

(b) $D(x)$ and Diff(x) when $B = 1.2$ Mbps

(c) Average bandwidth cost in Figure 8(a)

(d) Average bandwidth cost in Figure 8(b)

Figure 8: Average latency and difference of a gymnastics video.

ACKNOWLEDGMENTS

We thank Li Zuo for his efforts during the implementation and experimental testing of the idea in this paper, and our colleague Dwight Daniels for his editing of the paper.

REFERENCES

[1] "EyeVision project," 2001, http://www.ri.cmu.edu/projects/project_449.html.

[2] "Digital movia camera systems," http://www.digitalair.com/techniques/.

[3] S. Wurmlin, E. Lamboray, O. G. Staadt, and M. H. Gross, "3d video recoder," in *Proceedings of the 10th Pacific Conference on Computer Graphics and Applications (PG '02)*, pp. 325–334, Beijing, China, October 2002.

[4] J.-G. Lou, H. Cai, and J. Li, "A real-time interactive multi-view video system," in *Proceedings of the 13th ACM International Conference on Multimedia (ACM MULTIMEDIA '05)*, pp. 161–170, Singapore, November 2005.

[5] H. Ma and K. G. Shin, "Multicast video-on-demand services," *ACM SIGCOMM Computer Communication Review*, vol. 32, no. 1, pp. 31–43, 2002.

[6] W. F. Poon and K. T. Lo,, "New batching policy for providing true video-on-demand (T-VoD) in multicast system," in *Proceedings of IEEE International Conference on Communications (ICC '99)*, vol. 2, pp. 983–987, Vancouver, BC, Canada, June 1999.

[7] A. Dan, D. Sitaram, and P. Shahabuddin, "Scheduling policies for an on-demand video server with batching," in *Proceedings of the 2nd ACM International Conference on Multimedia (ACM MULTIMEDIA '94)*, pp. 15–23, San Francisco, Calif, USA, October 1994.

[8] Y. W. Wong and J. Y. B. Lee, "Recursive patching—an efficient technique for multicast video streaming," in *Proceedings of the 5th International Conference on Enterprise Information Systems (ICEIS '03)*, vol. 1, pp. 306–312, Angers, France, April 2003.

[9] K. A. Hua, Y. Cai, and S. Sheu, "Patching: a multicast technique for true video-on-demand services," in *Proceedings of the 6th ACM International Conference on Multimedia (ACM MULTIMEDIA '98)*, pp. 191–200, Bristol, UK, September 1998.

[10] S. P. Lloyd, "Least squares quantization in PCM," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–137, 1982.

[11] M. Siegel, S. Sethuraman, J. S. McVeigh, and A. G. Jordan, "Compression and interpolation of 3D stereoscopic and multi-view video," in *Stereoscopic Displays and Virtual Reality Systems IV*, vol. 3012 of *Proceedings of SPIE*, pp. 227–238, San Jose, Calif, USA, February 1997.

[12] X. Tong and R. M. Gray, "Coding of multi-view images for immersive viewing," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '00)*, vol. 6, pp. 1879–1882, Istanbul, Turkey, June 2000.

[13] G. Chen, K. Ng, and H. Wang, "A multi-view video compression scheme based on direct view synthesis," in *ISO/IEC JTC1/SC29/WG11, MPEG2003/M10025*, Brisbane, Australia, October 2003.

[14] J. Lu, H. Cai, J.-G. Lou, and J. Li, "An effective epipolar geometry assisted motion estimation technique for multi-view image and video coding," in *Proceedings of IEEE International Conference on Image Processing (ICIP '06)*, Atlanta, Ga, USA, October 2006.

[15] X. Guo, Y. Lu, W. Gao, and Q. Huang, "Viewpoint switching in multiview video streaming," in *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS '05)*, vol. 4, pp. 3471–3474, Kobe, Japan, May 2005.

[16] H. Kimata, M. Kitahara, K. Kamikura, Y. Yashima, T. Fujii, and M. Tanimoto, "System design of free viewpoint video communication," in *Proceedings of the 4th International Conference on Computer and Information Technology (CIT '04)*, pp. 52–59, Wuhan, China, September 2004.

[17] Y. Liu, Q. Dai, and W. Xu, "A real time interactive dynamic light field transmission system," in *Proceedings of IEEE International Conference on Multimedia Expo (ICME '06)*, pp. 2173–2176, Toronto, Ontario, Canada, July 2006.

[18] W. Matusik and H. Pfister, "3D TV: a scalable system for real-time acquisition, transmission, and autostereoscopic display of dynamic scenes," *ACM Transactions on Graphics*, vol. 23, no. 3, pp. 814–824, 2004.

[19] E. Kurutepe, M. R. Civanlar, and A. M. Tekalp, "Interactive transport of multi-view videos for 3DTV applications," *Journal of Zhejiang University: Science A*, vol. 7, no. 5, pp. 830–836, 2006.

[20] S. Banerjee, B. Bhattacharjee, and C. Kommareddy, "Scalable application layer multicast," in *Proceedings of the ACM SIGCOMM Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication (ACM SIGCOMM '02)*, vol. 32, pp. 205–217, Pittsburgh, Pa, USA, August 2002.