Reliable Real-Time Transport of Stereo Video for Immersive Media Communication

Hyeyoung Chang[1], Sehchan Oh[1], JongWon Kim[1*],
Woontak Woo[1], and Jaesung Kwak[2]

[1] Department of Information and Communications
Kwang-Ju Institute of Science & Technology (K-JIST)
Gwangju 500-712, Korea
{tarilove,soh,jongwon,wwoo}@kjist.ac.kr
[2] Supercomputing Infra Development Lab.
Korea Institute of Science & Technology Information (KISTI)
Daejeon, 305-701, Korea
jskwak@kisti.re.kr

Abstract. Emerging high-speed next-generation Internet is enabling immersive media communication systems and applications to realize geographically distributed team collaborations, overcoming the limit of distance and time. Focusing on the reliable real-time delivery of 3D (i.e., stereo) video among corresponding parties, in this paper, key schemes for stereo video processing/display and reliable transport of stereo video packets over high-speed Internet are designed and implemented. The performance of proposed stereo video delivery system is evaluated both by emulating various network situations for quantitative comparison and by transmitting over real-world Internet up to the speed of around 100 Mbps. The results demonstrate the feasibility of the proposed system in supporting the desired immersive communication.

1 Introduction

Recent advancement towards high-quality media equipments and high-speed backbones is rapidly enabling the immersive communication between remote users. Video conferencing systems over Internet such as Access Grid (AG) [6] and Virtual Rooms Video Conferencing Service (VRVS) provide the environment where the remote conferees cooperate with each other face-to-face and share on-line materials with natural interaction. So far, due to insufficient bandwidth and power-limit in end systems, most existing video conferencing systems are struggling with the less-than-satisfactory and plain 2D video quality at rates between 100 ~ 300 kbps. These systems are still limited in making the involved users feel natural and immersed. In the context of advanced

* Corresponding author.

C.-W. Chung et al. (Eds.): HSI 2003, LNCS 2713, pp. 362-371, 2003.
© Springer-Verlag Berlin Heidelberg 2003

collaboration environment (ACE) [1], there have been continuous efforts to improve the tele-presence of visual communication by adding 3D immersiveness and tangibility. For example, in [2], shared virtual table environment is introduced to provide the impression of sitting together around a table. However, most of these 3D prototype systems are focusing on 3D video processing and display part while assuming kind of idealized transport of massive high-resolution 3D video over the network. That is, the issue of how to delivery the massive 3D video contents reliably over the best-effort Internet has not been addressed in sufficient depth.

Thus, in this paper, we discuss our trial on the immersive media communication system that can reliable transport stereo video in real-time. The designed system deals with the acquisition, corresponding processing, and transport of high-quality stereo video at the range of 100Mbps. Several display modes are supported in order to customize stereo display according to end user's environment. We also consider reliable high-speed media transport to successfully deliver massive stereo video contents on time. More specifically, captured left/right frames from stereo camera are processed to meet the available bandwidth of network link. They are then transported by high-speed transport module where the packet losses are selectively handled by automatic repeat request (ARQ) or forward error correction (FEC) based on the latency requirement. The performance of proposed stereo video delivery system is first evaluated by transporting over local area network, where we emulate several network situations for quantitative comparison. The stereo video is then transmitted over the real-world Internet delivering over 155Mbps maximum bandwidth KOREN (Korea Advanced Research Network)/KREONET (Korea Research Environment Open Network).

The outline of the paper is as follows. After discussing the system overview in Section 2, we will detail about the acquisition and processing module in Section 3. The reliable transport module with real-time support follows in Section 4. Section 5 provides the performance analysis and throughput demonstration. Finally, Section 6 concludes the paper.

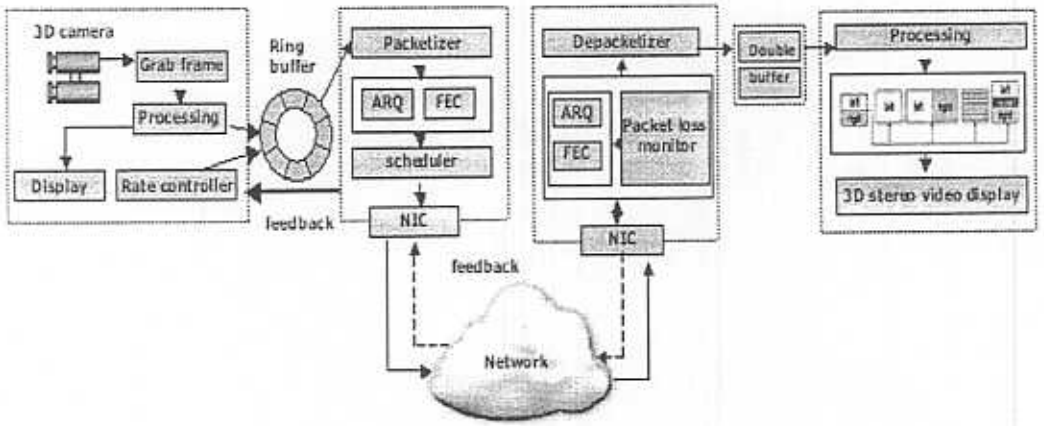

Fig. 1. Building blocks of stereo video delivery system

2 Stereo Video Delivery System Framework

In Fig. 1, the building blocks of the proposed stereo video delivery system are depicted. From the sender side, modules are linked in the order of frame capture, processing, packetization, and reliable transport. The captured left/right stereo video frames from IEEE 1394 stereo camera (maximum 400 Mbps) are first processed to match the target network and system limitation. To maintain high-quality video while keeping the involved processing delay minimum, we choose to use uncompressed video in several scalable forms and avoid expensive hardware compression interfaces. At present, simple processing including pre-filtering and sub-sampling is adopted to match the bandwidth limit of given network.

The processing and network transport modules, running in separate threads, share the video frames through the ring buffer located in middle. Once processed media stream is delivered to the network transport module, it is fragmented and packetized into UDP/IP packets with maximum transfer unit (MTU) size limit and customized packet header with sequence number[1]. The required transport should better be aware of the characteristics of media contents. That is, we may want to exploit that the video in general does not require 100% reliable transport. Although depending on the involved video format, it is generally much more important to provide in-time delivery, especially for interactive communication between users. In this paper, we adopt two categories of reliable transport: ARQ and packet-level FEC [7][8]. Note that the packets are scheduled and inter-packet space is controlled to avoiding burst transmission.

At the receiver side, it checks the sequence numbers to detect gap due to packet loss. Reaction to the network variations, especially to the packet loss, is performed via the feedback. Successfully received packets are then processed and combined into video frames. After passing through reconstruction, the stereo video is rendered in various formats. In handling the stereo video, the immersiveness to users largely depends on the appropriate setup of 3D display. Considering the diversity of user display ranging from single CRT monitor to silver screen with 3D glasses and projectors, we provide diverse display modes for compatibility with user environment. Special attention has also been paid to keep the synchronization between the left/right eye frames.

3 Stereo Image Acquisition/Processing and Display

Digital stereo camera with IEEE 1394 interface is utilized to generate stereo video frames corresponding to left and right human eyes. Figure 2 depicts camera configurations for capturing stereo image, i.e., parallel and intersectional alignment of two cameras. We adopt the parallel alignment to minimize the 3D distortion effect. To synchronize a pair of stereo video frames, we can represent the images in various ways as shown in Fig. 3.

---

[1] We are planning to upgrade the packetization based on RTP/UDP/IP so that we can take advantage of RTP/RTCP standard suite.

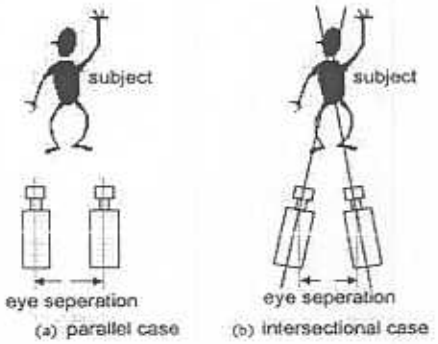

Fig. 2. Stereo camera configurations

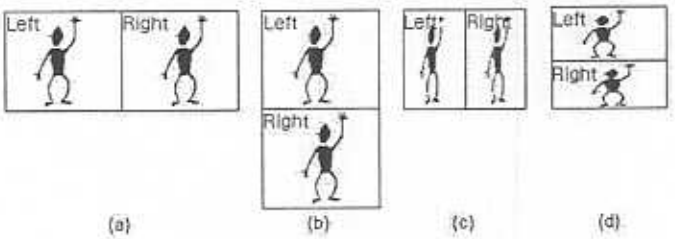

Fig. 3. Methodologies of stereo video frame representation

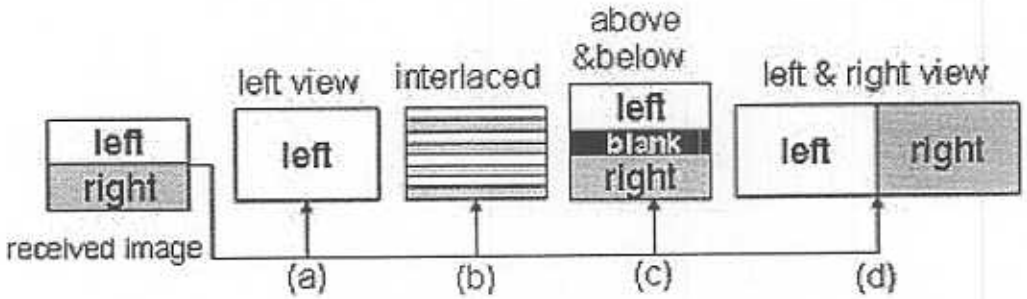

Fig. 4. Options for stereo video frame representations in the receiver

Once stereo frames are generated, processing (i.e., pre-filtering and down-sampling) is applied to adapt to limited network bandwidth. For example, we can perform 2:1 down sampling in combination with low-pass decimation filtering. The received data are up-sampled and applied to a low-pass interpolation filter according to the status of the display system in the receiver.

At the receiver, we provide several stereo displays for flexibility. For example, if the receiver does not have any stereo display, the receiver system only displays up-sampled left frame as shown in Fig. 4(a). If the receiver has a stereo-enabled monitor, the data directly transformed to interlaced formats as shown in Fig. 4(b) and Fig. 4(c). In case of Fig 4(c), left frame is placed to the top and right one is placed to the bottom. Blank lines must be inserted between these two images for delay adjustment. The value of the blank shift depends mainly on the resolution of the graphic card used as

well as the monitor refresh rate used. If the receiver has dual monitor or a stereo-enabled HMD (head mount display), the format in Fig. 4(d) is used.

4 Reliable Transport of Stereo Video

To provide the required reliability, we exploit the Reliable Blast UDP suggested in [3] for the ARQ. For the FEC side, we adopt fast realization of packet-level FEC taking after Rizzo's implementation [7].

High Speed Transport Mode Based on ARQ. RB_UDP [3] is a reliable data transmission scheme using augmented acknowledgement over QoS enabled network. It takes the blast approach, where single retransmission request is made for missing packets of each blast (instead of a packet). In RB_UDP, the delivery on UDP-channel data is assisted by TCP feedback channel. Assuming low-loss QoS-enabled network, it exhibits very high throughput and targets close-loop mechanism to provide 100% reliability. However, if the underlying network fails to provide the guaranteed level of loss and delay, its throughput is subject to rapid deterioration due to the burden of retransmission.

High Speed Transport Mode Based on FEC. FEC is attractive for time-constrained media stream over long-distance network since it provides minimum amount of latency and high degree of reliability. We adopt fast realization of packet-level FEC modified from Rizzo's implementation using Vandermonde matrices [7]. Packet-level FEC generates redundant packet by performing bit-wise XOR operation by aggregating adjacent packets(typically 2 or 3)[10]. Despite its advantage for low-delay situation, it pays the penalty of computational overhead caused by encoding/decoding processing of redundant packets. It also pays the bandwidth overhead of redundant packets.

5 Experimental Performance

The prototype stereo video delivery system is implemented using high-performance PC's, 1394 stereo camera, and 3D display. Dell Workstation 530MT with dual Xeon™ 1.7GHz CPU, 1G memory, WildCat6110 graphic card is used. For stereo video capture, we use IEEE 1394 stereo camera (PointGrey's Bumblebee™) to support maximum resolution of 640 x 480 pixels at 25 frames per second.

5.1 Stereo Image Acquisition/Processing and Display

Stereo video is captured at 10 frames per second (i.e., 640 x 480 x 3 x 2 x 10 x 8 = 147.5 Mbps). To meet the available network bandwidth (100Mbps), captured video frames are down-sampled with the corresponding filtering and decimation.

Figure 5 shows stereo display options at the receiver. With polarized glasses, the user can enjoy stereo video with less eye fatigue than other options with shutter

glasses. Also, with shutter glasses, the user can see half resolution version only. Experimental heuristic for Fig. 4(c)'s above/below format is about 2.3 percent blank lines of its vertical resolution (10 blanks lines for 480 vertical resolution) [5].

5.2 Stereo Video Transport over IP Networks

The performance of proposed stereo video delivery system is first evaluated by transporting over local area network, where we emulate several network situations for quantitative comparison of selected module (Case 1 of Fig. 6). Divert Socket [4] mediated between sender and receiver is used to support various packet erasure and delay model on this emulated LAN testbed. Unfortunately, due to its low-profile PC hardware, it can only support rates less than 15Mbps at present. The stereo video is also transmitted over the real-world Internet (Case 2 of Fig. 6) by delivering over KOREN/KREONET. WAN path between K-JIST (Kwangju) and KISTI (Daejeon) includes six hops in each direction and has an RTT of approximately 5 ms[2]. With these network configurations, packet loss-rates (initial L_i and retransmission L_r) and effective throughput of transport, Th_f, is measured by changing the sending rate. Note that these are average values calculated over whole transmission period.

Transport Performance over Lan Without Loss Insertion. In this case, the transport rate is increased up to 100 Mbps limit[3]. Test has been performed at FEC rate of $(n,k) = (4,3)$. Effective throughput Th_f is depicted in Fig.7 and initial loss L_i in Fig. 8, respectively.

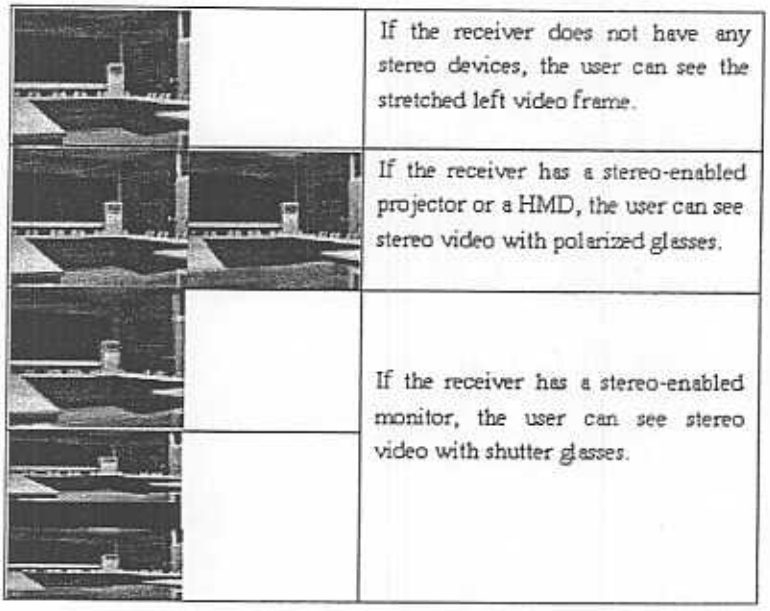

Fig. 5. Comparison of different stereo representations at the receiver

[2] Running iperf [9] between the sender at K-JIST and the receiver at KISTI, we get record of 73.4Mbps UDP stream with around 23% packet loss.

[3] Running iperf [9] over LAN gives 84.7Mbps UDP stream with around 11% packet loss.

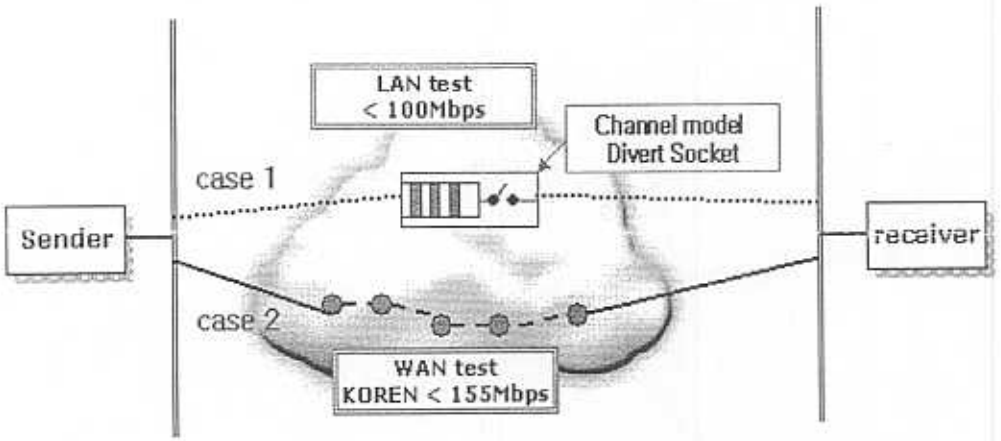

Fig. 6. Testbed configurations: Network emulation over LAN and real-world WAN

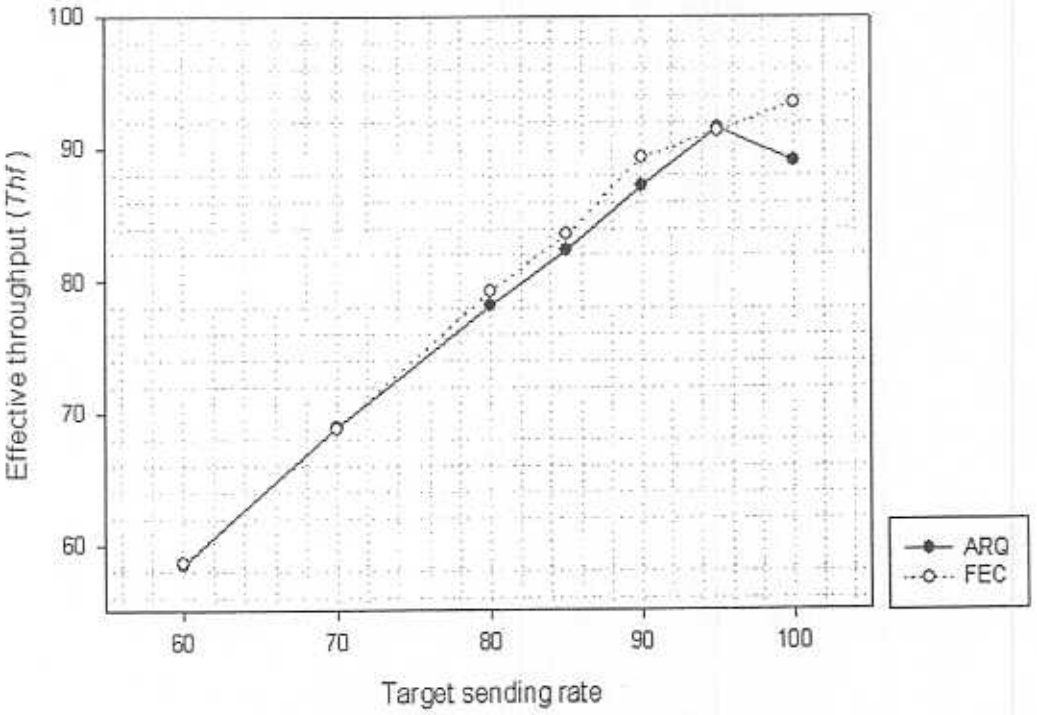

Fig. 7. Effective throughput variation (LAN without loss)

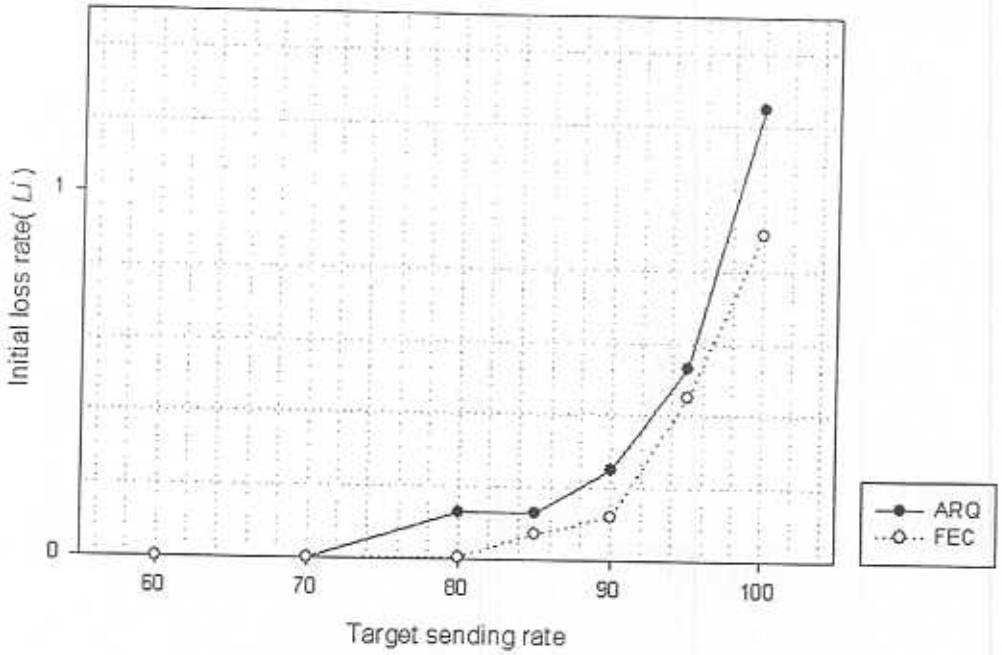

Fig. 8. Initial packet loss-rate variation (LAN without loss)

Table 1. Performance comparison of reliable transport (loss insertion: 10%)

Target Sending Rate		ARQ	FEC (33% redun.)
10 Mbps	L_i (initial loss)	4.8 %	0.78 %
	L_r (loss after retransmission)	0 %	0.78 %
	Th_f	8.96 M	9.56 M

Table 2. Performance comparison of reliable transport over WAN

Target Sending Rate		ARQ	FEC (33% redun.)
100 Mbps	L_i (initial loss)	2.9 %	0.52 %
	L_r (loss after retransmission)	0 %	0.52 %
	Th_f	72.6 M	84.74 M

In case of ARQ, as the sending rate reaches limit, the effective throughput starts decreasing due to the severe increase in retransmission cost and delay. Similar (but minor than ARQ) effect is occurring to FEC case due to processing and bandwidth overhead. As expected, initial loss-rate is increasing with the sending rate. Note that initial loss-rate of FEC is better than ARQ owing to the reconstruction by redundant parity data packet.

Transport Performance over LAN with 10% Loss Insertion. In order to have fair comparison, we insert the randomized packet loss of 10% using Divert Socket. Due to the reason mentioned above, the sending rate is limited to 10Mbps. Again the rate of FEC is $(n, k) = (4,3)$. As shown Table 1, the effective throughput of ARQ is lower than that of FEC due to the increased retransmission overhead. On the contrary, FEC shows marginal loss only with single transport. However, note that, in certain situation, remaining packet loss can cause significant deterioration of video quality.

Transport Performance over WAN. Table 2 shows measurements of throughput and loss. Tendency of loss and throughput are similar to that observed in the local test with 10% loss insertion.

6 Conclusion

We implement reliable real time transport system for stereo video over high-speed network. The system provides reliable transport with ARQ and FEC. It also provides immersiveness to the user by displaying high-quality stereo video in various modes. The performance evaluated by delivering real Internet demonstrates the feasibility of the system in supporting the immersive communication.

Acknowledgement

This research is funded in part by Korea Institute of Science and Technology Information (KISTI) and by Korea Research Foundation (KRF). The authors would like to thank Dr. C. Park, Y. Lee, J. Jeong, and J. Park at K-JIST for their help in the system setup and evaluation.

References

[1] ACE homepage, http://calder.ncsa.uiuc.edu/ACE-grid/.

[2] O. Schreer and P. Kauff, "An immersive 3D video-conferencing system using shared virtual team user environments," in *Proc. ACM Collaborative Environments CVE 2002*, Bonn, Germany, Sept./Oct. 2002.

[3] E. He, J. Leigh, O. Yu, and T. A. DeFanti, "Reliable blast UDP: Predictable high performance bulk data transfer," in *Proc. IEEE Cluster Computing 2002*, Chicago, IL, Sept. 2002.

[4] W. Kellerer, E. Steinbach, P. Eisert, and B. Girod, "A real-time internet streaming media testbed," in *Proc. IEEE Inter. Conf. on Multimedia and Expo (ICME`2002)*, 2002.

[5] M. Husak, "Guide to making your own digital stereo-video movies in DVD quality for playing on computers," http://staffold.vscht.cz/~husakm/stereopcvideo.html.

[6] Access Grid Homepage, http://www-fp.mcs.anl.gov/fl/accessgrid/.
[7] L. Rizzo, "Effective erasure codes for reliable computer communication protocols," *Computer Communication Review,* vol. 27, April 1997.
[8] W. Kumwilaisak, J. Kim, and C.-C. J. Kuo, "Video transmission over wireless fading channels with adaptive FEC," in *Proc. Picture Coding Symposium `2001,* Apr. 2001.
[9] Iperf, http://dast.nlanr.net/projects/Iperf.
[10] J. Rosenberg and H. Schulzrinne. An RTP payload format for generic forward error correction, December 1999. RFC 2733.