AVS 3D VIDEO STREAMING SYSTEM OVER INTERNET

Baicheng Xin[1], *Ronggang Wang*[1], *Zhenyu Wang*[1], *Wenmin Wang*[1], *Chenchen Gu*[3], *Quanzhan Zheng*[3] *and Wen Gao*[2]

[1]School of Computer & Information Engineering, Peking University Shenzhen Graduate School

[2]National Engineering Laboratory of Video Technology, Peking University

[3] Tecent Technology（ShenZhen）Co., Ltd.

(e-mail: xinbaicheng@sz.pku.edu.cn, {rgwang, wangzhenyu, wangwm}@pkusz.edu.cn, {cicelygu, stevezheng}@tencent.com, wgao@pku.edu.cn)

ABSTRACT

The progresses on PC based 3D video display technology have made it feasible to provide 3D video streaming service over internet. Compared with 2D video streaming, 3D video streaming system demands more network bandwidth, computing power, and more complex transmission mechanism. There are rarely online 3D video services. In this paper, we describe a 3D video streaming services developed by our team cooperated with CNTV and NVIDIA. The main features of the proposed system are high compression performance of 3D video with low complexity both on processing and transmission. The system has been used to distribute 3D content for the 16th Asian Games Guangzhou.

***Index Terms*—** 3D encoder, streaming, AVS, CDN, 3D player

1. INTRODUCTION

Recent advances in 3D video processing, compression technology, strong power devices, and high-speed networks have made it feasible to provide real-time 3D multimedia services over the Internet. Compared with 2D video streaming, 3D video streaming demands more transporting bandwidth, computing power, and more complex transmission mechanism. To address these challenges, extensive researches have been conducted, such as 3D video compression [1], mobile 3D video processing [2], and P2P 3D streaming system [3] etc. But, there's rarely 3D streaming services lunched on the internet. In this paper, we described an online 3D streaming system developed jointly by our team, NVIDIA team and CNTV team. This system has been adopted by China Network Television (CNTV), some trial programs about the 16th Asian Games Guangzhou have been put on the CNTV website.

The main features of this 3D streaming system include: (1) High compression performance. (2) Low codec complexity. (3) Being compatible with AVS standard. (4) Supporting both anaglyph and shutter stereo display modes by detecting the configuration of PC automatically. (5) Supporting bit-rate adaption from 500kbps to 2Mbps. (6) Supporting both "in-window" and "full-screen" stereo display mode. (7) The 3D video depth perception can be adjusted by users. (8) The content distribution architecture is based on CDN, and the element streams of video and audio are packetized into TS format, which is compatible with digital TV.

This paper is organized as follows. The overall 3D streaming system is described in section 2. The 3D video codec design is showed in detail in section 3. The content distribution network and stream protocols are provided in section 4. The 3D video player implementation is presented in section 5. Experimental results on compression performance and the industry application example are provided in section 6, and this paper is concluded in section 7.

2. 3D VIDEO STREAMING SYSTEM

The overall 3D streaming system over internet is composed of 3D Streaming Server, Content Distribution Network and 3D Media Players etc. Fig.1 shows the architecture for 3D video streaming system. Raw left view of video-audio and raw right view of video-audio data are pre-compressed by 3D Media Encoder, and then saved in Content Storage module or pushed to Streaming Server. The Content Management module decides whether the data stream is published in real-time or VOD mode, and then puts this information on the Content Publish module. The Content Publish module puts the program information on the portal. Upon the client's request, if it is a real-time program, Streaming Server fetches the data stream from 3D Media Encode module. Otherwise, Streaming Server retrieves compressed 3D video/audio data from Content Storage module. The Streaming Server module also adapts the 3D video/audio bit-streams according to the network status and QoS requirements. After the adaptation, using the transport protocols to packetize the compressed bit-streams and send the video/audio packets to the Internet. Packets may be dropped or experience excessive delay inside the Internet due to congestion. To improve the quality of video/audio transmission, Content Distribution Network is deployed

978-1-4673-2193-8/12/$31.00 ©2012 IEEE

over Internet. For packets that are successfully delivered to the

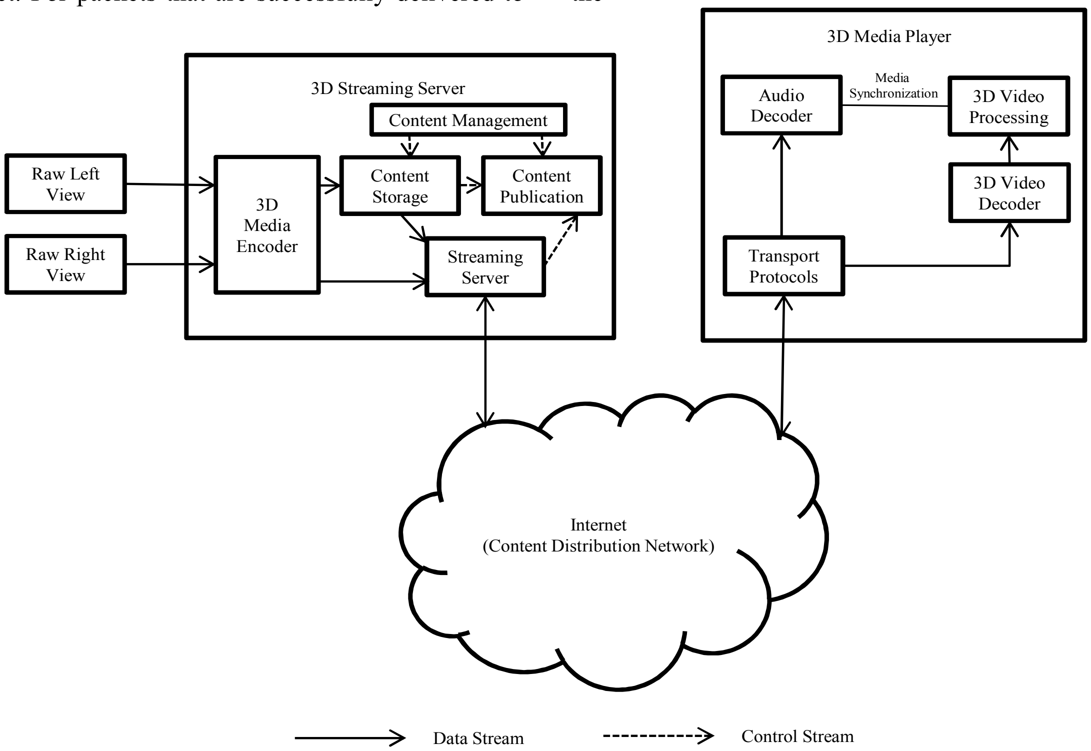

Fig. 1. Proposed architecture for 3D streaming system

3D media player, they first pass through the transport layers and are then processed by the application layer before being decoded at the 3D video/audio decoder, a 3D video Post Processing module is placed behind 3D video decoder to do some work on video up-sampling, quality enhancement, and 3D video depth adjustment etc. . To achieve synchronization between 3D video and audio presentations, media synchronization mechanisms are required. From Fig. 1, it can be seen that the three components are closely related and they are coherent constituents of the 3D video streaming architecture.

3. 3D VIDEO CODEC

The 3D media encoder and decoder are the core modules of the proposed 3D streaming system. Fig. 2 shows the architecture of our proposed 3D media encoder. There are four signal capturing modules to accept the video and audio data from left view and right view respectively. The audio data of the two views are feed to MP3 Encoder, and compressed to double channel audio stream. The video data from left view and right view are first combined into one video data by Stereo Video Composite module, and then compressed by 3D Video Encoder. The advantage of the proposed 3D Media Encoder is that the video data from two views are compressed to just one element stream, which makes it feasible to distribute 3D video stream just like the way of traditional 2D video stream. The compressed video and audio stream are packetized and multiplexed into TS format. The TS packets are forwarded both to Streaming Server and Content Storage.

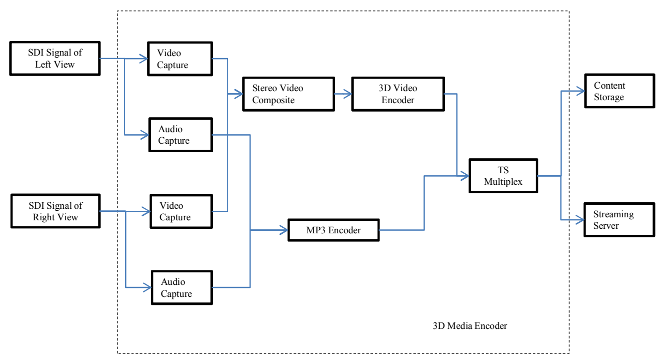

Fig. 2. Proposed architecture of 3D media encoder

To save bandwidth, the video data from left view and right view are first down-sampled in horizontal/vertical direction to half resolution, and then combined into one frame. After being decoded, the decoded frame is then up-sampled to two frames of left view and right view by the interpolation filter. Our 3D video compression algorithm is based on China AVS standard [6] for its high compression performance and low complexity, and the right half of a combined frame can be encoded by referring the left half of the said frame, so the encoding efficiency can be improved further. The proposed 3D codec architecture is showed in Fig. 3.

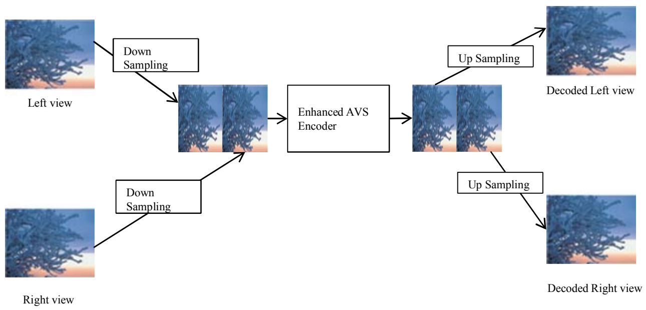

Fig.3. Proposed 3D video codec architecture

4. CONTENT DISTRIBUTION NETWORK

The 3D Media is distributed by Content Distribution Network, which is provided by CNTV. The Content Distribution Network is composed of a series of streaming servers, which are deployed in Beijing, Shanghai, Guangzhou and Shenzhen etc. There are also number of streaming servers distributed in different networks such as China Telecom network, China Unicom network and China Education and Research network. The CDN structure is illustrated in Fig. 4.

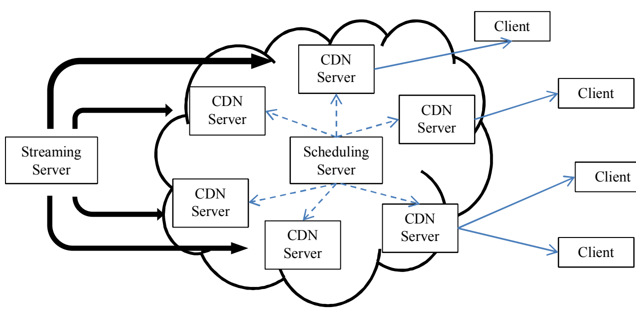

Fig. 4. Content Distribution Network

The playback flow of one program is in the following:

1) The user finds program information from the Content Publish module.
2) The request from user is redirected to a nearest CDN server by Content Publish module in Fig.1. If the requested program is in the cache of the nearest CDN server, then goto 4), otherwise goto 3).
3) A proper CDN server or Streaming Server is selected by Scheduling Sever to provide content to the nearest CDN server.
4) The data stream of the program is sent to the client by the nearest CDN server.

The handshake protocols between Streaming Server and CDN Server, CDN Sever and clients are all based on HTTP. When the requested program is a VOD program, the data stream path is through Content Storage, CDN Server and Clients, there's no data loss. But when the requested program is a real-time program, the data is generated at constant speed in spite of network conditions. The cache size in CDN server is limited, for one cache, when data receiving speed is faster than data sending speed, the cache tends to be overflow, and data loss is occurred. How to handle this kind of data loss is essential to user experiences on the 3D video quality. There're at least two ways to handle the data loss when the cache is overflow. One way is to drop the new arrived data packet; the other way is to drop the data packets in the cache. For the first way, if the data receiving speed is faster than the data sending speed for one period of time, data loss will be occurred frequently as illustrated in Fig. 5(a). Since there're data dependencies between data packets in P/B frames and data packets in their reference frames, the decoded frames will be destroyed for a long time. So we adopted the second way, when the cache is overflow, we empty the whole cache, and put the new arrived data packet in the cache, so the data loss will be occurred in a short time, the influenced frames will be less as illustrated in Fig.5 (b).

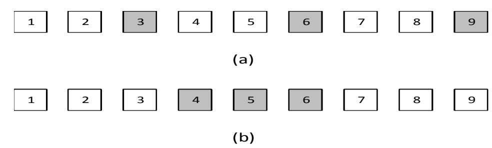

Fig.5. Data loss pattern by the two different cache overflow handling ways (grey packets are lost)

5. 3D MEDIA PLAYER

Two 3D display modes of anaglyph and shutter are supported by our 3D Media Player. The 3D Media Plugin will detect the user's PC configuration automatically, if the NVIDIA®3D Vision™ [4] is detected, the shutter display mode is active, otherwise, the anaglyph display mode is active. So almost all PCs with windows OS can enjoy 3D videos provided by our system.

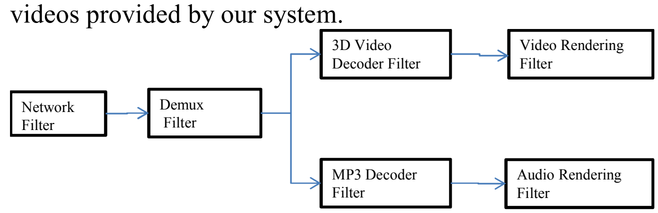

Fig. 6. 3D media player based on Direct Show framework

The 3D Media Player is developed based on Direct Show framework, which is showed in Fig.6. The data packets are fetched by Network Filter, and then the Demux Filter recovers the element stream from the TS packets. The 3D video element stream and audio element stream are feed to 3D Video Decoder Filter and MP3 Decoder Filter respectively. In the video Rendering Filter, Each decoded frame of video is divided into two sub-frames with half resolution, the sub-frames are up-sampled to full resolution, and displayed by anaglyph mode or shutter mode. The decoded audio data is playback and synchronized with video frames in the Audio Rendering Filter.

6. EXPERIMENTS AND APPLICATION

High compression performance is one of key advantages for our proposed 3D Video Encoder. We compared our 3D video encoding method with simulcast encoding mode on AVS platform RM09 [5]. Six typical 3D test sequences from AVS 3D workgroup are tested, GOP structure of "IBBPBBP…" is adopted, the cycle of random access is one second. The performance gain by our method over simulcast scheme is showed in Table 1. The bitrate can be saved more than 20% in average by our method. Two example RD curve comparisons between our method and the simulcast scheme are shown in Fig. 7.

Table 1. Performance gain of our proposed method over simulcast

Sequences	Frame rate	Avg. bit-rate increased	Avg. dB gain
Alt_Moabit	16.67	-25.37%	1.59
Book_Arrival	16.67	-21.91%	1.077
Door_Flower	16.67	-22.00%	0.842
Poznan_Street	25	-24.73%	1.260
Poznan_CarPark	25	-27.15%	1.394
Tsinghua_Classroom	25	-15.95%	0.884
Average		-22.85%	1.17

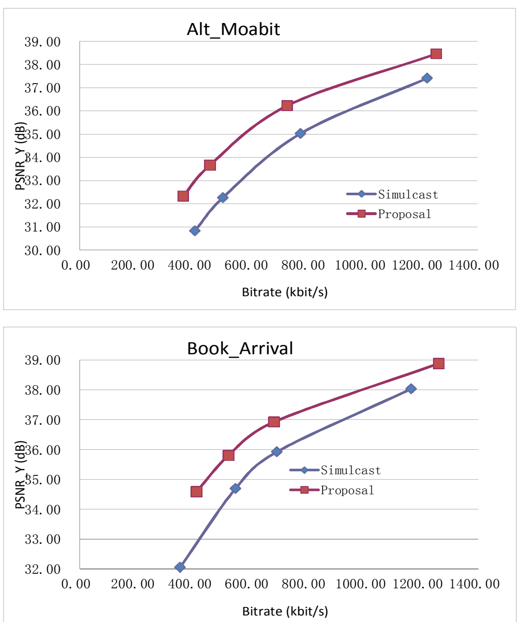

Fig.7. RD curve comparisons between simulcast and our poposal

Our proposed 3D video streaming system has been adopted by China Network Television. A 3D video streaming service is online at the website of CNTV. The 3D contents related to 16th Asian Games Guangzhou are available for VOD. The URL of the website is http://yayun.cntv.cn/3D/index.shtml. Fig.8. gives the web portal for this service. Two copies of 500kpbs stream and 2Mbps stream are provided for each program.

Fig.8. The web portal of 3D streaming service on CNTV

7. CONCLUSIONS

A 3D video streaming services has been developed by our team cooperated with CNTV and NVIDIA. A high performance and low complexity 3D encoder is proposed. More than 20% bitrate has been saved compared with simulcast coding. This system has been online at the website of CNTV. Numerous internet users from all over the world have enjoyed the 3D video content on PC.

8. ACKNOWLEDGE

This work was supported by the grant of National Basic Research Program of China (973 Program) 2009CB320907, the grant of Key Projects in the National Science & Technology Pillar Program 2011BAH08B03, Tecent and the grant of Shenzhen Basic Research Program of JC201104210117A.

9. REFERENCES

[1] A. Vetro, T. Wiegand, and G. J. Sullivan, "Overview of the stereo and multiview video coding extensions of the H.264/AVC standard", Proc. IEEE, Special Issue on 3D Media and Displays, 2011.
[2] P. Merkle, Y. Wang, K. Müller, A. Smolic, and T. Wiegand, "Video plus Depth Compression for Mobile 3D Services", Proc. IEEE 3DTV Conference, Potsdam, Germany, Mai 2009.
[3] S.-Y. Hu et al., "Flod: A framework for peer-to-peer 3d streaming," in Proc. INFOCOM, 2008.
[4] http://www.nvidia.com/object/3d-vision-surround-technology.html.
[5] AVS reference software, ftp://159.226.42.57/public/avs_docs/avs_software.
[6] W. GAO et al. "AVS - The Chinese next-generation video coding standard" NAB, Las Vegas, 2004.