A Novel Selective Frame Discard Method for 3D Video over IP Networks

Young-uk Chung
Department of Electronic Engineering, Kwangwoon University
Seoul - Republic of Korea
[e-mail: yuchung@kw.ac.kr]
*Corresponding author: Young-uk Chung

Received August 16, 2010; revised October 19, 2010; accepted November 8, 2010; published December 23, 2010

Abstract

Three dimensional (3D) video is expected to be an important application for broadcast and IP streaming services. One of the main limitations for the transmission of 3D video over IP networks is network bandwidth mismatch due to the large size of 3D data, which causes fatal decoding errors and mosaic-like damage. This paper presents a novel selective frame discard method to address the problem. The main idea of the proposed method is the symmetrical discard of the two dimensional (2D) video frame and the depth map frame. Also, the frames to be discarded are selected after additional consideration of the playback deadline, the network bandwidth, and the inter-frame dependency relationship within a group of pictures (GOP). It enables the efficient utilization of the network bandwidth and high quality 3D IPTV service. The simulation results demonstrate that the proposed method enhances the media quality of 3D video streaming even in the case of bad network conditions.

Keywords: Selective frame discard, video-plus-depth, 3D, IP network

---

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science and Technology (2010-0016638) and the Research Grant of Kwangwoon University in 2010. We express our thanks to reviewers who checked our manuscript.

DOI: 10.3837/tiis.2010.12.013

1. Introduction

Three-dimensional (3D) video allows viewers to feel a compelling sense of reality and provides powerful immersive experiences. The ultimate goal in 3D video is holography, but a lot of research still needs to be done. Considering the recent advances in technology, stereoscopic/multiview 3D video seems to be the most practical option among the various 3D video representation technologies and a representative application for this in the consumer market is 3D TV over internet protocol (IP) networks.

Recent advances in internet technology make it possible to provide streaming media over the Internet. But it is still hard to implement 3D TV applications over IP networks. The main limitation of 3D video transmission over IP networks is the large size of the data to be transmitted, because at least two synchronized video streams (for the right and left eyes) are required to display a 3D video. To overcome this limitation, various methods have been studied. A popular approach is the video-plus-depth representation which renders a 3D video using a 2D video stream and its associated depth map [1][2]. It has been reported that the depth map can be highly compressed. Two MPEG standards cover this approach: The specification of the depth format is included in ISO/IEC 23002-3 (MPEG-C) and a method of transmitting video-plus-depth within a conventional MPEG-2 transport stream is included in an amendment (Amd. 2) to ISO/IEC 13818-1 (MPEG-2 Systems). The N-view-plus-M-depth representation was proposed as a complementary version of the video-plus-depth representation to support multiview situations [2][3]. In [4], a novel multiview video coding (MVC) scheme based on simulcast coding was introduced to achieve a high compression gain. However, despite these efforts, the bit rates for multiview video are still high: For a 704×480, 30 fps, 8 view sequence with MVC encoding, a bit rate of approximately 5 Mbps with a peak signal noise ratio (PSNR) of 38 dB is commonly required.

Like other streaming media, 3D video requires the properties of real time, continuity, and data dependency. A video frame should be delivered to the receiver before its playback deadline and all video frames must be rendered in a pre-defined order. Before decoding a compressed video frame, all data it refers to should have been received. Otherwise, decoding errors and mosaic-like rendering effects may occur. A good 3D video over IP network should meet all of these requirements. However, since the current IP network is based on a best-effort strategy, its variable and unpredictable behavior in terms of the channel bandwidth, delay, and packet-loss make it difficult to provide good QoS for 3D video delivery.

One of the most challenging problems is the bandwidth mismatch between the network bandwidth and the bit rate of the video stream, which is found in the media server or intermediate node. Due to the heterogeneity of the Internet and its dynamic fluctuating bandwidth, the data rate of an incoming video stream may be greater than the output data rate of a streaming server or intermediate node. This can cause random packet losses in the buffer of the streaming server or intermediate node. This random loss of packets may result in fatal problems in the 3D video decoder. For example, the loss of some packets in an I-frame or P-frame may cause mosaic-like rendering effects in the current picture. Moreover, it also affects all of the following P-frames and B-frames within the current GOP (group of pictures). To resolve these problems, several video source adaptation methods, such as simulcast, transcoding, scalable coding and selective frame discard, have been introduced [5][6][7][8][9][10].

Simulcast stores multiple copies of a single video source with different compression rates in the server [5]. The client selects and receives one of them according to its access bandwidth. This method is widely used in the current systems, but is not suitable for intermediate network nodes, because an intermediate node cannot switch to another video stream with a different compression rate. The transcoding method can be employed at both the server and some intermediate network nodes [6]. It can convert the video content to an appropriate format, with the desired quality, form and rate. However, this requires extra computational power, which may not be practical in real time applications. Scalable coding uses a single video source, but the bit stream can be truncated at almost any point [7]. The truncated bit stream can be decoded with any reconstruction quality that corresponds to the number of bits recovered. Scalable coding requires a much smaller computational cost compared with the transcoding method. However, the compression efficiency is low. Selective frame discard (SFD) is regarded as a simplified version of transcoding or scalable coding [8][9][10][11][12]. It is simple but effective. There have been several works about the SFD method. Zhang et al. formulated an optimal SFD problem using a QoS-based cost function, but the complexity of the solution is too high for it to be implemented in real-time [8]. Feng et al. introduced a priority-based frame discard framework. However, it can only support precompressed video streams [9]. Huo et al. proposed a novel SFD algorithm which is adaptive to the network bandwidth [10]. It works fast and requires low computational complexity, but it is not suitable for 3D video. Zhang et al. proposed a tree-aware SFD algorithm for P2P IPTV systems taking into account the location of the set-top box (STB) node in the P2P multicast tree [11]. However, it can only be adopted in the intermediate nodes of P2P networks. Zhou et al. proposed a frame dropping strategy which is aware of the characteristics of the video contents and ensures distortion-fairness sharing among video streams [12]. But they focused on resource allocation among multiple video streams rather than the video adaptation efficiency, and only considered a specific system environment, such as multi-channel multi-radio multi-hop wireless networks.

All of these previous works take into account the traffic source of 2D video. Because 3D video has a larger size of data to be transmitted compared with 2D video, the bandwidth mismatch becomes a more critical problem. In addition, the traffic structure and characteristic s of 3D video are different from those of 2D video. Therefore, these previous methods which are adopted for 2D video may not be optimal solutions for 3D video. In this paper, we propose a noble SFD based video source adaptation method for 3D video, so as to address the bandwidth mismatch problem. The proposed method takes into account the characteristics of depth based 3D video compression standards in its decision to discard overdue frames. Four frame discard policies and the procedure required to implement these policies are developed.

The rest of this paper is organized as follows: In Section II, we describe the system environment in which the proposed method is employed; In Section III, the proposed symmetrical frame discard method is introduced; Section IV evaluates the simulation results, including both subjective and objective performance measures; In Section V, we summarize our results and conclude this paper.

2. System Description

This section gives an overview of the system environment for which the proposed symmetrical frame discard method is adopted. It begins with an introduction to the overall system architecture of 3D video over IP networks. And, a detailed description of the video-plus-depth representation for stereoscopic video is provided.

2.1 3D Video over IP Network

Fig. 1 shows the overall system architecture of 3D video over IP networks which we considered in this paper. The system consists of three parts, a 3D media server, intermediate node, and STB/3DTV, connected through IP networks. A 2D video and its depth maps are obtained from a depth camera system and a 2D-to-3D conversion method. These data are separately encoded as the base and enhancement layers. They are multiplexed as a single stream of media data. Before transmitting them over the IP network, these media data are packetized into individual RTP packets. The format of the RTP header is specified in RFC 3550 [13]. The RTP header contains a timestamp field which is set to the sampling timestamp of the content with a 32 bit value. A 90 kHz clock rate is commonly used to measure the timestamp. The RTP payload header also includes a field which indicates the current frame type, i.e., I-, P-, or B-frame. Each media encoding type has its corresponding companion RTP payload format specifications; RFC 2250 for MPEG-1 and MPEG-2, RFC 3640 for MPEG-4, RFC 3984 for H.264, and Part 8 of the AVS documents.

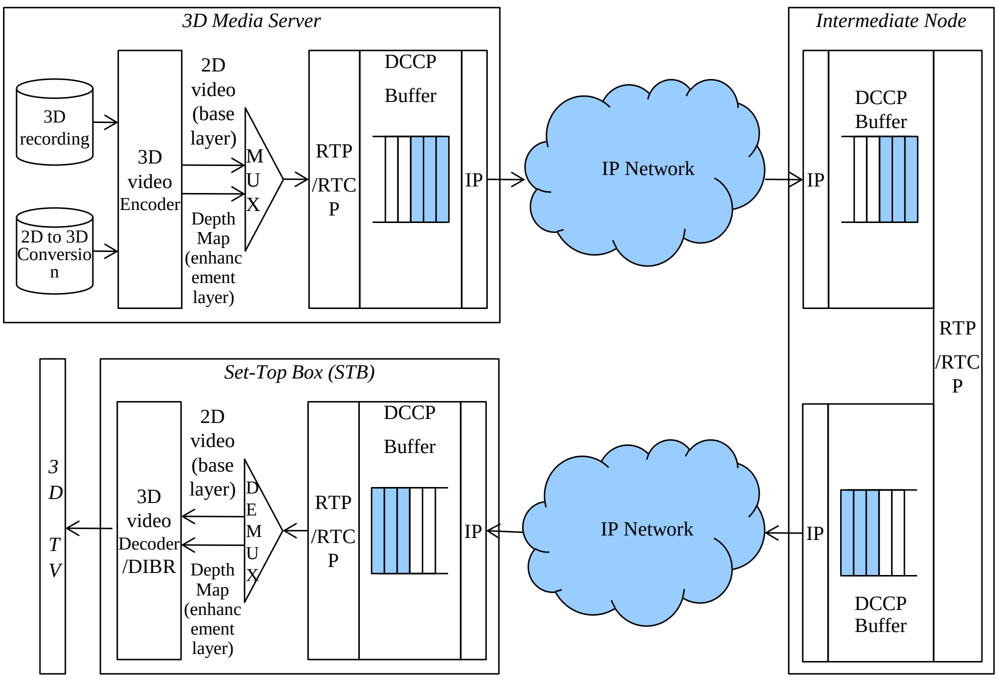


Fig. 1. 3D IPTV system architecture

The media server sends the RTP packets to the STB using the datagram congestion control protocol (DCCP). The DCCP is a transport protocol that was designed as a replacement for the user datagram protocol (UDP). It is appropriate for streaming media applications that prefer to minimize abrupt changes in the sending rate. The congestion control ID 3 (CCID3) profile in DCCP is commonly used as the congestion control mechanism.

The intermediate node repeatedly receives RTP packets through DCCP/IP and then puts them into the DCCP buffer. After processing, the node repeatedly reads the RTP packets from the DCCP buffer and then sends them to the STB or another intermediate node.

Upon receiving the packets, the STB depacketizes the RTP data and divides the single stream of media data into the base and enhancement layer using a demultiplexer. The data in these two layers is decoded so as to recover the depth maps and the 2D video, respectively. Finally, they are converted into the left and right image sequences using the depth image based rendering (DIBR) technique and displayed on a 3D TV.

2.2 Video-Plus-Depth 3D Video

In this paper, we focus on the video-plus-depth representation of stereoscopic video which renders a 3D video using a 2D video stream and its associated depth map. This representation has been widely used due to its flexible representation and compatibility with existing coding and transmission technologies [1][2].

In this format, a depth map has the same spatial resolution as a 2D video and contains the position of the associated 2D video texture in the 3D space. Because the storage and bandwidth requirements of video-plus-depth based stereoscopic video applications are high compared to those of 2D video applications, efficient compression techniques are required for 3D video over IP networks. The layered coding approach can be effectively used for encoding stereoscopic video. In layered coding approaches, the 2D video and depth image sequences are coded as the base and enhancement layers, respectively. Traditional 2D compression techniques are used to encode both the video and the depth map sequences. The base and enhancement layers are transmitted and decoded as the 2D video and depth image sequences, respectively. Before displaying 3D video on the display, the supplied video and depth image sequences are converted into the left and right image sequences using an image warping technique known as DIBR.

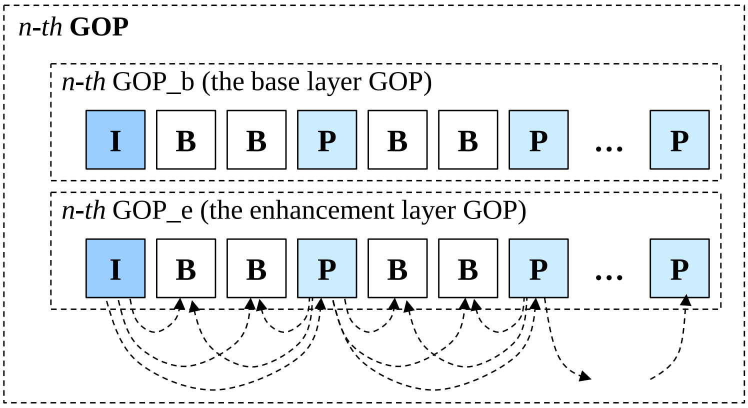


Fig. 2. GOP structure for the depth based 3D video

An example of the layered coding structure with the base and enhancement layers is shown in **Fig. 2**. In this structure, a GOP (group of pictures) consists of two parts: a base layer GOP (GOP_b) and an enhancement layer GOP (GOP_e). They are encoded separately and combined as a single GOP stream. When transmitting, GOP_b is generally transmitted earlier than its related GOP_e. The structure of GOP_b and GOP_e is the same as the coding structure of GOP which is adopted by almost all of the popular video compression standards, such as MPEG-1, MPEG-2, MPEG-4, H.264, and AVS. Both GOP_b and GOP_e consists of an intra-coded frame (I-frame), followed by a series of predictive coded frames (P-frames) and bidirectionally predictive coded frames (B-frames). An I-frame can be decoded by itself, but is coded with only a moderate compression ratio. Compared with an I-frame, a P-frame can be coded more efficiently using motion compensated prediction from one or more past I-frame or P-frames. Therefore, its decoding requires reference to backward I-frame or P-frames. A

B-frame can be highly compressed, but its decoding requires references to both backward and forward I-frame or P-frames. For a GOP_b and a GOP_e with a rendering order of "IBBPBBP…", their transmission and decoding order should be "IPBBPBB…".

3. Proposed Selective Frame Discard Method

This section gives a detailed description of the proposed symmetrical frame discard method which is adaptive to 3D video over IP networks. The key idea of the proposed method is the symmetrical discard of the 2D video frame and the depth frame. Because the 2D video and its depth map are closely related to each other, it is pointless to deliver only one of them. Therefore, the symmetrical discard enables the efficient utilization of the network bandwidth. Besides the consideration of the 3D traffic relationship, discarded frames are selected after additional consideration of the playback deadline, the network bandwidth, and the inter-frame dependency relationship within a GOP. The proposed method consists of two parts; the novel frame discard policies for depth based 3D video and the procedure required to implement these policies.

3.1 Frame Discard Policy

When the bit rate of a 3D video stream is greater than the available network bandwidth, the node selects frames not to be sent, so as to save bandwidth and ensure the delivery of the stream before its playback deadline. The frame discard should be performed according to the predefined policy. Our proposed frame discard policy is as follows:

Policy 1: If a frame within a base layer GOP is decided to be discarded, the frame at the same position within the related enhancement layer GOP should be discarded first, because a frame of the enhancement layer (depth data) is worthless without the related frame of the base layer (2D video data), according to the concept of video-plus-depth representation. In this paper, we assume that the base layer GOP is generally transmitted earlier than its related enhancement layer GOP. In the opposite case, the relation of the frame discard is also reversed.

Policy 2: The 2D video data is more important than the depth map data in a video-plus-depth 3D video; because the depth map data can be estimated from the 2D video data by several techniques, but not vice versa [14].

Policy 3: The frames which are decided to be discarded should have as little dependency on the frames selected to be sent as possible; once a frame is selected to be discarded, it is preferable to discard all frames within the same GOP which refer to this frame. Otherwise, unexpected errors may occur in the decoding of these frames.

Policy 4: The distance between adjacent discarded frames should be as far as possible to obtain a smooth video rendering quality.

From the above frame discard policy and the inter-frame dependency relationships within a GOP, several typical cases will occur, as follows: B-frames will be discarded with the highest priority. I-frames will be discarded with the lowest priority. When a P-frame is discarded, all of the subsequent P-frames and B-frames within the same GOP should also be discarded. When an I-frame is discarded, we do not have to send all other frames within the same GOP.

3.2 Frame Discard Procedure

In 3D video over IP networks, the frames within a GOP should be delivered to the end system (such as the STB) before it plays the GOP. The playback deadline of a video frame is included in the timestamp field of the RTP packets containing the video frame. Let *ts(n)* be the timestamp value of the first RTP packet of the n-th GOP and $W_{th}(n)$ be the sending time interval of the n-th GOP. Then, $W_{th}(n)$ can be calculated as

$$W_{th}(n) = ts(n+1) - ts(n). \tag{1}$$

We define $Ws(n)$ as the sending window of the n-th GOP. The duration of the sending window of the n-th GOP is $Wth(n)$.

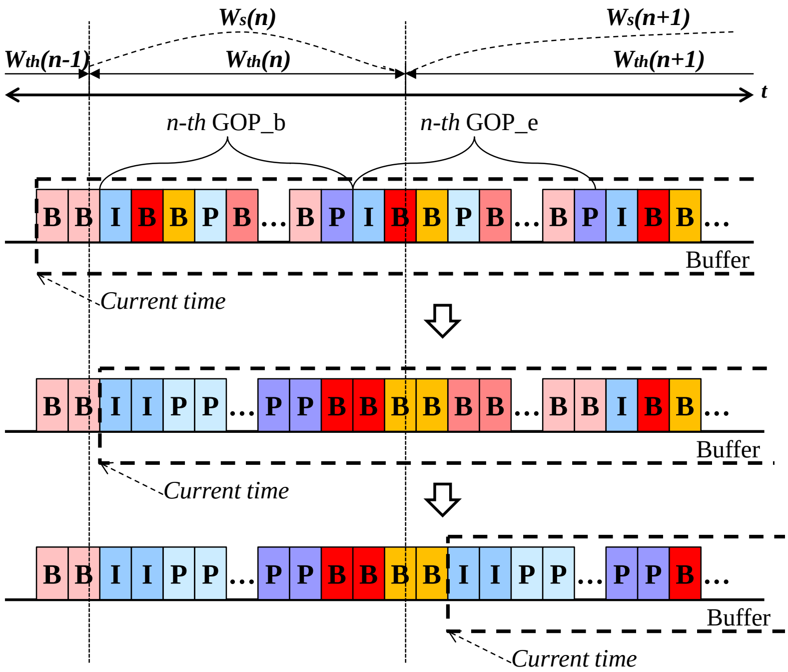


Fig. 3. Example of the proposed frame discard procedure

The detailed frame discard procedure is described as follows: For the initialization of the transmission, the proposed method sets the current time as the reference point, T. Then, it calculates $Wth(n)$ for all GOPs, and sets the start point and end point of $Ws(n)$ for all GOPs using $Wth(n)$ and T. From the GOP structure for the depth based 3D video in Section II, the n-th GOP consists of the n-th GOP_b and its related n-th GOP_e; the n-th GOP_b and n-th GOP_e are multiplexed and transmitted as a single GOP stream, as shown in **Fig. 1**. Before sending a GOP, the proposed method first rearranges all I/P/B frames in the GOP according to their importance, priority, and relationship. As a result of the rearrangement, the frames in the GOP are lined up in order of importance, i.e., first the I-frames, then the P-frames and finally the B-frames, and each frame of GOP_b is paired up with a frame of GOP_e in accordance with their relations. Then, the proposed method starts sending the first RTP packet of the first frame in the current GOP. After all RTP packets of each pair of frames have been sent, the proposed method checks whether the current time is beyond the end time point of the current GOP's sending window. If not, it starts sending the RTP packets of the next pair of frames. Otherwise, it discards all of the following unsent frames contained in the current GOP, and directly switches to the transmission of the next GOP.

Fig. 3 shows three snapshots of the DCCP buffer which adopts the proposed method. The first snapshot shows the situation just before transmitting the last pair of frames in the $(n\text{-}1)$-th

GOP. The second snapshot shows the situation when the rearrangement of the *n*-th GOP has just finished. All of the frames in the *n*-th GOP are rearranged according to the proposed policies. The third snapshot shows the situation when the rearrangement of the *(n+1)*-th GOP has just finished. In this snapshot, the transmission of the former yellow B-frame (B-frame of the *n*-th GOP_b) exceeds the sending window, but the proposed method still transmits the latter yellow B-frame (B-frame of the *n*-th GOP_e), because it checks whether the time allocated to the sending window has expired or not after the transmission of a pair of frames. After the transmission of the pair of red B-frames, the proposed method checks if the current time is beyond the end time point of the current GOP's sending window. Because, the time does not exceed the time allowed for sending, the next pair of frames (yellow B-frames) is transmitted. Because a pair of frames is transmitted together, two yellow B-frames are transmitted, despite the fact that the time exceeds the sending window. After the transmission of a pair of yellow B-frames, the proposed method checks if the current time is beyond the end time point of the current GOP's sending window. Because the time exceeds the end time point of the current GOP's sending window, all of the following unsent frames contained in the *n*-th GOP are discarded and the transmission of the *(n+1)*-th GOP is started.

4. Simulation Results

To evaluate the performance of the proposed method, a simulation is carried out. In this section, we first describe the simulation environment and then, we evaluate the simulation results using both objective and subjective performance measures.

4.1 Simulation Environments

The simulation environment is emulated using Network Simulator 2 (NS2). We modified the tool-set of NS2 which was developed for simulating the delivery quality of MPEG video transmissions over IP networks [15]. We adopt the DCCP module instead of UDP in this tool-set. Also, we adopt the video-plus-depth based 3D video source in this tool-set. In the simulation, a 3D video sequence namely, interview, whose resolution is 720x576 pixels, is used as the test video. The Joint Scalable Video Model (JSVM) reference software version 9.15 is used to encode both the 2D video and the depth map simultaneously as a scalable video bit stream. The average bit rate of the encoded bit stream is about 740kbps and the GOP length is 12 with a GOP structure of "IBBPBBP..." for both GOP_b and GOP_e.

In our simulation, we consider a one hop network between a media server and an end-system (STB). The network bandwidth is adjusted to evaluate the performance of the proposed method. Both objective and subjective performance measures, such as the mean opinion score, average PSNR, the percentage of I/P frames, and the received GOP structures, are considered.

4.2 Performance Evaluation

The subjective rendering performance observed under different network bandwidth conditions is rated using the mean opinion score (MOS) which was defined in [16]. The MOS is obtained by averaging the opinion scores which scale the difference in subjective ratings between the delivered 3D image sequence and the original 3D image sequence. The opinion score has a value ranging from 0 (excellent) to 100 (bad), where 0 indicates the same quality as the original sequence and 100 indicates the worst perceived image quality compared to the original video sequence. Ten volunteer observers participated in this evaluation. **Fig. 4** shows the resultant MOS score for the overall image quality of the test video sequence.

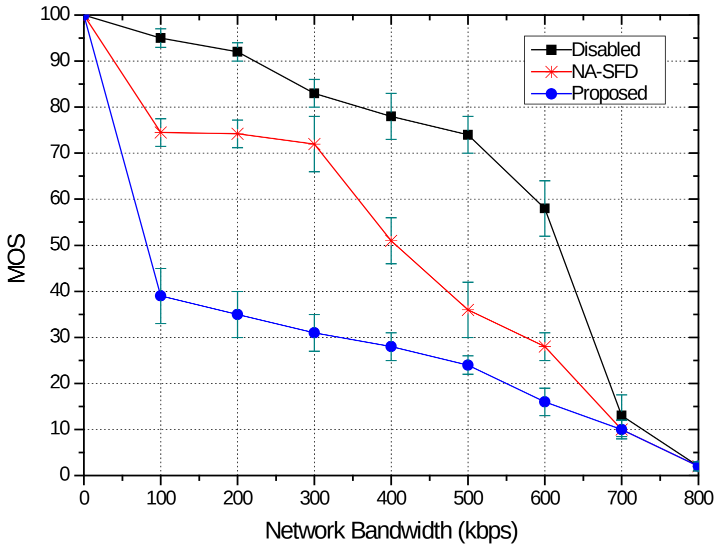
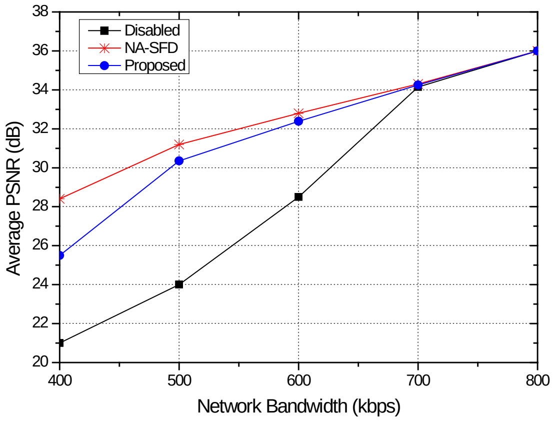


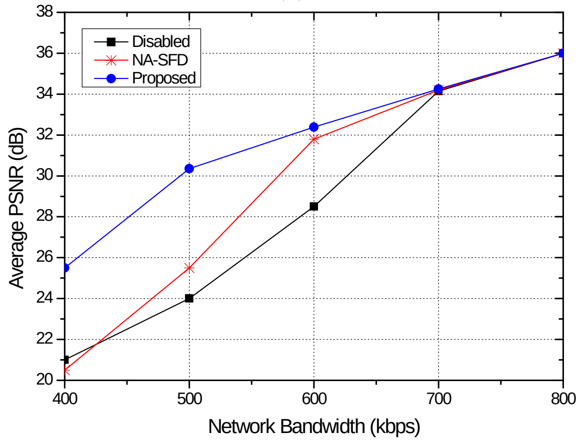
Fig. 4. Relationship between the MOS scores for the perceived overall 3D image quality and the network bandwidth

In **Fig. 4**, the label "Disabled" indicates the case where no SFD method is adopted, and the label "NA-SFD" indicates the ordinary selective frame discarding scheme without considering the relationship between the 2D video and the depth map [10]. The label "Proposed" indicates the case where the proposed selective frame discard method is adopted. When the SFD method is disabled, **Fig. 4** shows that the delivered 3D video has good quality only when the network bandwidth is similar to the average bit rate of the original 3D video bit stream. In the case of the "Disabled" method, there is no rearrangement of I/P/B frames in the GOP and no SFD procedure. When the bandwidth is insufficient, random packet losses occur. This random loss of packets may result in decoding errors and mosaic-like rendering effects on the Set-Top Box side. Therefore, viewers may be subjected to mosaic-like pictures, unmeaningful images or 2D/3D mixed images. Even worse, the media player may stop and not be able to continue to display the video in a low bandwidth situation. However, when the proposed method is enabled, the delivered video exhibits relatively good quality even in the case where the network bandwidth is much lower than the average bit rate of the original 3D video stream. It is for this reason that the proposed method can transmit at least all of the I-frames and some selected P-frames of both the 2D video and the depth map in low bandwidth environments. Therefore, the delivered 3D image sequence is intact and clear, even though pauses may sometimes occur.

Compared with the NA-SFD method, the proposed method also has a better 3D video quality, especially in the mid-range of the network bandwidth. The reason for this is that asymmetrical frame discards can occur between the 2D video and related depth map in the conventional NA-SFD method. The NA-SFD method discards more frames of the depth map compared with the 2D video, because the depth map data is at the back of a GOP. This asymmetrical frame discard leads to the deterioration in the 3D image quality.



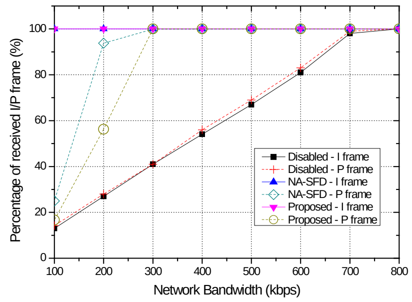
(a)



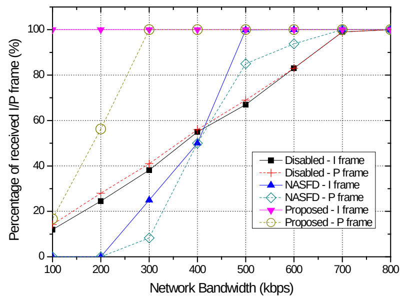
(b)

Fig. 5. Relationship between the average PSNR and the network bandwidth **(a)** the base layer (2D video) and **(b)** the enhancement layer (depth map)

To ensure the reliablity of the subjective performance, we evaluate several objective performance measures. First, a comparison of the average PSNR is performed for different network bandwidths. **Fig. 5** illustrates the comparisons of the average PSNR that we performed in our simulation. In **Fig. 5-(a)**, we compare the average PSNR of the 2D video decoded from the received base layer data. In **Fig. 5-(b)**, we compare the average PSNR of the depth map decoded from the received enhancement layer data. As shown in **Fig. 5**, the proposed method has a similar average PSNR performance level for both the 2D video and the depth data. It is for this reason that the proposed method discards frames of the 2D video data and the depth map data symmetrically according to frame discarding policy 1. The SFD-disabled method also has a similar average PSNR performance level for both the 2D video and the depth data, because it discards frames of the 2D video and the depth map with the same probability. The NA-SFD method has good PSNR performance for the 2D video, but shows poor performance for the depth map in low network bandwidth environments. The reason for this is that this method discards more frames of the depth map, which is located at the back of a GOP, than those of the 2D video, according to its algorithm. These results are consistent with the MOS result which is shown in **Fig. 4**. Because NA-SFD produces a 2D video with good quality and a depth map with relatively poor quality on the receiver side, the decoded 3D video quality decreases as the network bandwidth decreases. The SFD-disabled method shows the worst performance, because frames are discarded asymmetrically and important frames such as I-frames and P-frames can be randomly discarded. On the other hand, the proposed method realizes the best 3D image quality, because it prevents I/P-frames from being discarded and transmits frames symmetrically.



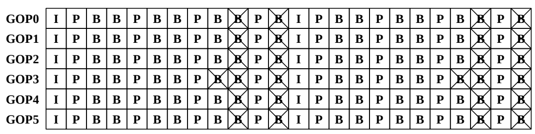
(a)



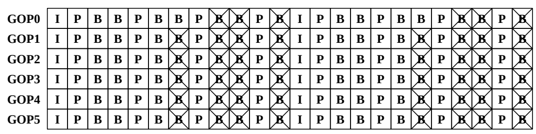
(b)

Fig. 6. Percentage of received I-frames and P-frames **(a)** the base layer (2D video) and **(b)** the enhancement layer (depth map).

The second objective performance measure is the percentage of received I-frames and P-frames according to the network bandwidth. **Fig. 6** presents the relationship between the percentage of received I/P-frames and the variation of the network bandwidth. In **Fig. 6-(a)**, we compare the percentage of received I/P frames of the base layer data. In **Fig. 6-(b)**, we compare the percentage of received I/P frame of the enhancement layer data. In **Fig. 6**, we find that the proposed method can transmit all of the I-frames of both the base layer and the enhancement layer data, regardless of the allocated network bandwidth. Most of the P-frames can also be sent by the proposed method. NA-SFD can transmit all of the I-frames and almost all of the P-frames for the base layer traffic. However, for the enhancement layer traffic, it transmits all of the I/P-frames only when the allocated network bandwidth is over 500 kbps. It is also shown that the SFD-disabled method can transmit all of the I/P-frames only when the allocated network bandwidth is over 700 kbps. These results are consistent with those shown in **Fig. 4** and **Fig. 5**. From these results, we can see that the proposed method outperforms the previous methods.



(a)



(b)

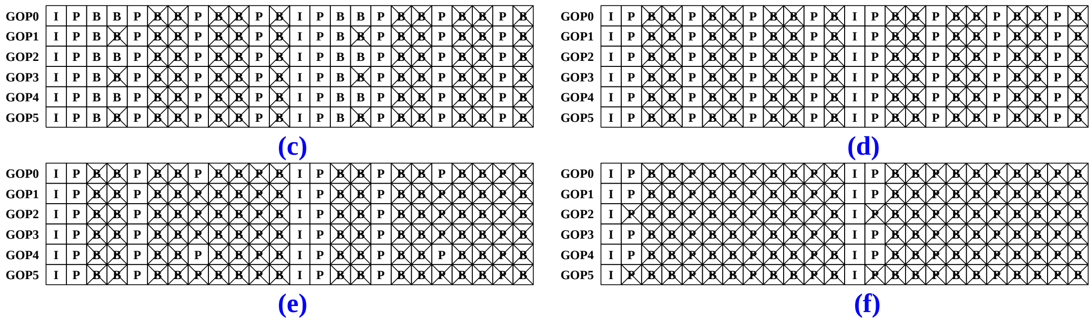


Fig. 7. Successfully received frames observed at the end system in different network bandwidth conditions. The crossed symbol indicates the dropped frame. (a) bandwidth = 600kbps; (b) bandwidth = 500kbps; (c) bandwidth = 400kbps; (d) bandwidth = 300kbps; (e) bandwidth = 200kbps; (f) bandwidth = 100kbps.

Finally, to check if the proposed method works well, we record successfully delivered video frames within 10 consecutive GOPs for different network bandwidth conditions. **Fig. 7** shows an example of the received frames observed at the end system. In this figure, we can see that in each case, the GOP_b and GOP_e frames are discarded symmetrically. Also, we can confirm that the B-frames are discarded first, followed by the P-frames located near the end of the GOP, as the network bandwidth decreases. None of the discarded frames prevented the correct decoding of the successfully received frames. These results exactly correspond with the policy of our proposed method and other objective and subjective results.

5. Conclusions

In this paper, we propose a novel selective frame discard method for the transmission of 3D video over IP network systems to address the bandwidth mismatch problem. Especially, the proposed method takes into account the relationship between the 2D video and the depth map, in its decision to discard overdue frames. The performance evaluations show that our method outperforms the existing methods. The MOS of the received 3D stream is tested subjectively by having viewers watch the generated 3D streams on a stereoscopic display. The subjective tests show that the 3D streams received by our method provide the viewers with a superior 3D experience. Moreover, in terms of several objective performance measures, such as the average PSNR, the percentage of I/P frames, and the received GOP structures, our method also outperforms the existing methods. Moreover, the proposed method can be easily deployed in large-scale media streaming systems, such as P2P based live 3D IPTV broadcasting systems, because it has lower computational complexity and requires no extra modifications to the target system.

References

[1] C. Fehn, K. Hopf and Q. Quante, “Key technologies for an advanced 3D-TV system,” in *Proc. of SPIE 3-D TV, Video Disp. III*, pp. 66-80, 2004. Article (CrossRef Link)

[2] A. M. Tekalp, E. Kurutepe and M. R. Civanlar, “3DTV over IP,” *IEEE Signal Processing Magazine*, vol. 24, no. 6, pp. 77-87, Nov. 2007. Article (CrossRef Link)

[3] O. Schreer, C. Fehn, N. Atzpadin, M. Muller, A. Smolic, R. Tanger and P. Kauff, “A flexible 3D TV system for different multi-baseline geometries,” in *Proc. of IEEE ICME*, pp. 1877-1880, 2006. Article (CrossRef Link)

[4] H. M. Ozaktas and L. Onural, *Three Dimensional Television: Capture, Transmission, and Display*, Springer, Heidelberg, 2007.
[5] B. Furht, R. Westwater and J. Ice, "Multimedia broadcasting over the Internet: Part II - Video compression," *IEEE multimedia*, vol. 6, no. 1, pp.85-89, 1999. Article (CrossRef Link)
[6] A. Vetro, C. Christopoulos and H.Sun, "Video transcoding architectures and techniques: an overview," *IEEE Signal Processing Magazine*, vol. 20, no. 2, pp. 18-29, Mar. 2003. Article (CrossRef Link)
[7] Y. Wang, J. Ostermann and Y. Zhang, *Video Processing and Communications*, Prentice-Hall, New Jersey, 2002. Article (CrossRef Link)
[8] Z. L. Zhang, S. Nelakuditi, R. Aggarwa and R. P. Tsang, "Efficient server selective frame discard algorithms for stored video delivery over resource constrained networks," in *Proc. of IEEE INFOCOM'99*, pp. 472-479, Mar. 1999. Article (CrossRef Link)
[9] W. Feng, M. Liu, B. Krishnaswami and A. Prabhudev, "A priority-based technique for the best-effort delivery of stored video," in *Proc. of SPIE/IS&T Multimedia Computing and Networking*, Jan. 1999. Article (CrossRef Link)
[10] L. Huo, Q. Fu, Y. Zou and W. Gao, "Network Adapted Selective Frame-Dropping Algorithm for Streaming Media," *IEEE Transactions on Consumer Electronics*, vol. 53, no. 2, pp. 417-423, May, 2007. Article (CrossRef Link)
[11] W. Zhang, Q. Zheng and Y. Lian, "Tree-aware selective frame discard for P2P IPTV system on set-top boxes," *IEEE Transactions on Consumer Electronics*, vol.55, no.4, pp. 1982-1987, Nov. 2009. Article (CrossRef Link)
[12] L. Zhou, X. Wang, W. Tu, G.-M. Muntean and B. Geller, "Distributed scheduling scheme for video streaming over multi-channel multi-radio multi-hop wireless networks," *IEEE Journal of Selective Areas on Communications*, vol. 28, no. 3, pp. 409-419, April, 2010. Article (CrossRef Link)
[13] H. Schulzrinne, S. Casner, R. Frederick and V. Jacobson, "RTP: A transport protocol for real-time applications," RFC 3550, July, 2003. Article (CrossRef Link)
[14] M. T. Pourazad, P. Nasiopoulos and R. K. Ward, "An H.264-based scheme for 2D to 3D video conversion," *IEEE Transactions on Consumer Electronics*, vol. 55, no. 2, pp. 742-748, May, 2009. Article (CrossRef Link)
[15] C. Ke, C. Shieh, W. Hwang and A. Ziviani, "An evaluation framework for more realistic simulations of MPEG video transmission," *Journal of Information Science and Engineering*, vol.24, no.2, pp.425-440, Mar. 2008. Article (CrossRef Link)
[16] C. T. E. R. Hewage, S. T. Worrall, S. Dogan, S. Villette and A. M. Kondoz, "Quality evaluation of color plus depth map-based stereoscopic video," *IEEE Journal of Selective Topics on Signal Processing*, vol. 3, no. 2, pp. 304-318, Apr. 2009. Article (CrossRef Link)

Young-uk Chung received the B.S., M.S. and Ph.D. degrees in electrical engineering and computer science from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea, in 1997, 1999 and 2003, respectively. From 2003 to 2005, he was a manager in Hynix Semiconductor Inc., Ltd. He joined the Faculty of Kwangwoon University, Seoul, Republic of Korea, in 2005, where he is currently an Assistant Professor with the Department of Electronic Engineering. His research interests include multimedia communication, wired/wireless communication networks, protocols and services.