ON THE PERFORMANCE OF H.264/MVC OVER LOSSY IP-BASED NETWORKS

Athanasios Kordelas[1, 2], Tasos Dagiuklas[2,3], Ilias Politis[1,2]

1 Dept. of Electrical and Computer Engineering, University of Patras, Rion, 26500, Patras, Greece
2 Dept. of Telecommunication System and Networks, TEI of Mesolonghi, Nafpaktos, 30300, Greece
3 Open University of Cyprus, Latsia, Nicosia, 2220, Cyprus
athankord@tesyd.teimes.gr, ntan@teimes.gr, ilpolitis@gmail.com

ABSTRACT

This paper studies the performance of different packetization schemes (single NAL Unit, aggregation packet and Fragmentation Unit) for the emerging H.264/MVC standard, in terms of overhead, number of decoded frames, error propagation and PSNR using different network conditions (MTU size, packet loss). The experimentation test-bed platform, utilizes the Multi Session Transmission approach for various video packetization options in terms of the number of NAL Units per frame. Extensive test-bed experiments indicate that the fragmentation of frames in more than one NAL Units results in significantly higher perceived video quality, in terms of PSNR, for both base and non-base view, than the fragmentation of the NAL Unit at the RTP layer.

Index Terms — 3D Video**,** H.264/MVC, RTP, Video Quality Evaluation

1. INTRODUCTION

The delivery of 3D media to individual users remains a highly challenging problem due to the large amount of data involved, diverse network characteristics, user terminal requirements, as well as, users' context such as their preferences and location. As the number of visual views increases, current systems will struggle to meet the demanding requirements in terms of delivery of consistent video quality to fixed and mobile users.

Several problems occur during the transmission of H.264 3D video sequences. The most important one is that the non base view video quality after the transmission in an IP-based network may lead to reduced perceived video quality compared to base view due to inter-camera prediction [1]. Recently, 3D video transmission over IP networks has received particular attention from the research community. In particular, [2] studies the quality reduction of multi-view coded (MVC) [3] video due to wireless losses, however the impact of these loses on the base and non-base views is not considered. Moreover, [4] investigates the performance of two different packetization modes of the H.264/MVC standard (Single NAL Unit (SNU) and Fragmentation Unit (FU) - 1 NALU per frame) under different network.

Opposite to previous studies, this paper aims to assess the performance of 3D video streaming over IP based networks using various video packetization modes in the H.264/MVC standard. Particular emphasis has been given to the impact of the number of Network Abstraction Layer Units (NALUs) per frame on the video performance as well as, to the evaluation of different packetization in terms of PSNR and overhead in the base and non-base view. Moreover, an error concealment scheme for recovering lost NALU headers is implemented. Through experimentation it has been found that the best solution in terms of error propagation and PSNR is the frame fragmentation in multiple NAL Units of both the base and non-base views.

The rest of the paper is organized as follows. Section 2 includes an overview of the H.264/MVC and RTP standards, while in Section 3 the proposed error resiliency scheme is described. The experimentation setup is presented in Section 4 and Section 5 presents the performance evaluation results. Finally, Section 6 concludes the paper.

2. H.264/MVC OVERVIEW

H.264/MVC standard is an extension of the H.264/Advance Video Coding (AVC) and H.264/Scalable Video Coding standards [5], [6]. According to these standards, the Video Coding Layer (VLC) produces a coded representation of the video, while the NAL Unit encapsulates the video data in a prepared way for transmission. Fig.1 illustrates the inter-camera prediction introduced by MVC between the base and non-base view providing higher compression efficiency.

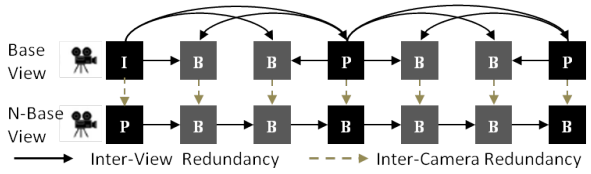

Fig.1. Typical H.264/MVC stereo prediction structure

For the encapsulation of the MVC video data in NALUs, a new MVC extension header is introduced by H.264/MVC standard (4 octets), as shown in Fig.2. In base view, a Prefix

NAL Unit (type 14) preceded each H.264/AVC base view NALU utilizing the extended H.264/AVC NALU header. In the same manner, the extended H.264/AVC NALU header is used for the encapsulation of the non-base view bitstream which is named coded slice of non-base view (type 20).

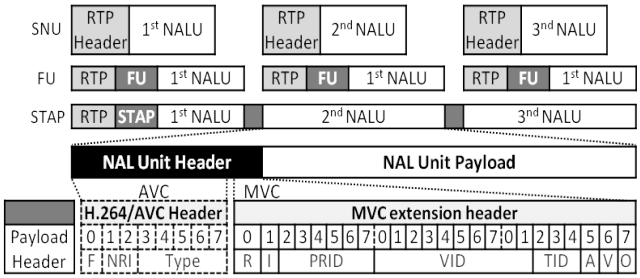

Fig.2. RTP MVC Payload structures

In parallel with view scalability, H.264/MVC inherits temporal scalability from H.264/SVC. The use of both scalability options offers the ability of the extraction, transmission and playback of the desired views, under a specific frame rate. The scalability information exists within the new NALU headers (prefix NAL header, coded slice of non-base view) at the fields TID (Temporal_id) and VID (View_id).

For the encapsulation of NALUs in RTP packets, three NALU payload structures are specified in the RTP specification of MVC [7], as shown in Fig.2. The first one is the "Single NAL unit" (SNU) according to which each RTP packet encapsulates a whole NALU. The second one is the "Aggregation Packets", which specifies that multiple NALUs are encapsulated in one RTP packet and includes five versions, STAP-A, STAP-B, MTAP-16, MTAP-24 and NI-MTAP. The latter is known as "Fragmentation Unit (FU)" and allows the fragmentation of one NALU into smaller RTP packets. In turn, FU includes two versions, FU-A and FU-B.

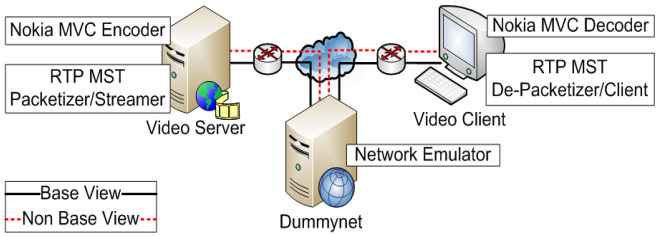

Fig.3. 3D Video Streaming Testbed

MVC incorporates three transmission modes namely, Single-Session Transmission (SST), Multi-Session Transmission (MST) and Media-Aware Network Element (MANE) based transmission [8]. In the case of SST mode, all the MVC information is transmitted over a single RTP session, utilizing one transport address (unicast). In MST mode, the number of the RTP sessions is equal to the number of the used transport addressed (multicast), as illustrated in Fig.3. Each RTP session in MST may carry only the base view, or a combination of the base view with a number of non-base views, or only non-base views. Furthermore, in MANE mode, the server utilizes MST transmission. RTP packets are collected from an intermediate entity (MANE) and using an Adaptation Decision Taking Engine (ADTE) RTP packets may be dropped by taking into account network conditions and client's characteristics.

3. ERROR RESILIENCE IN RTP LAYER

Several approaches for concealing errors due to lost frames have been proposed [9]. In order to deal with the loss of important headers, a RTP-aware error recovery algorithm based on FU payload structure is implemented at the RTP de-packetizer. This scheme is able to reconstruct in the application layer all the lost H.264/AVC NALU headers, allowing the decoder to recognize and decode all the video frames (NALUs) of both views. Frame losses occur more frequently when an entire encoded video frame is encapsulated in a single NALU. If the IP datagram that encapsulates the NALU header is lost, then the decoder discards the entire NALU (frame) as it is unable to recognize the NALU type.

According to the RTP standard [10], a FU-A always encapsulates part of a NALU, using the RTP header and two additional headers namely, "FU Indicator" and "FU Header", as shown in Fig.4. The first two fields (3 bits) of the "FU Indicator" denoted by *F* and *NRI* and the last field (5 bits) of the "FU Header" denoted by *Type* obtain their values from the corresponding fields of the H.264/AVC NALU header. The remaining fields describe the first/last fragment of the frame and the type of the RTP packet. Following the fragmentation of a NAL unit, the H.264/AVC NALU header is erased during packetization and re-constructed by the two FU headers during de-packetization.

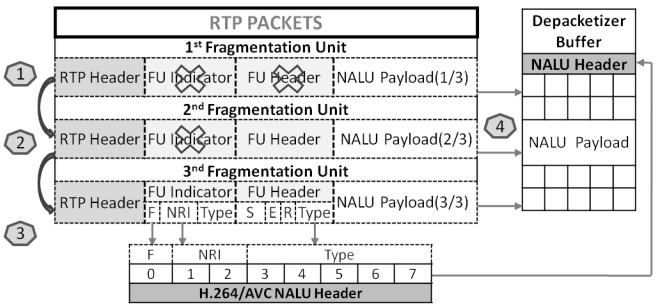

Fig.4. Proposed RTP error-resilience algorithm

The proposed error resilience scheme is based on the fact that the necessary fields of the H.264/AVC header are present in all FU's of a NALU while the missing pixel values included in lost fragmentation units are replaced by default values that the decoder interprets as distortion of the original frame and conceals the errors. The proposed

algorithm is illustrated in Fig.4 and comprises the following steps:

1. The de-packetizer discovers that the first fragment of the NALU is missing.
2. It searches the following fragments of the same NALU to discover a received fragment with no errors.
3. Upon discovering such a fragment, it copies the needed values of the bit-fields in the "FU Indicator" and the "FU Header" and reconstructs the necessary for the decoding H.264/AVC NALU header.
4. The de-packetizer rejects the other RTP headers and stores the NALU payload.

4. EXPERIMENTATION SETUP

4.1. Test-bed platform

MVC video transmission over IP is achieved by using several tools, as shown in Fig.3. Nokia's MVC Encoder/Decoder is used for the encoding and decoding of different view video sequences [11], while packet losses are emulated using Dummynet [12]. Furthermore, a number of tools have been implemented based on the MVC, RTP and UDP standards:

1. RTP packetizer – able to create RTP packets from a coded MVC sequence regardless of the fact that NALUs may contain an entire frame or a part of it. It is also able to create RTP packets using all the payload structures as defined by the standard (SNU, STAP-A, FU-A).
2. RTP de-packetizer – able to de-packetize RTP packets and reconstruct a MVC sequence. It is tolerant to bit errors and it is able to recognize the payload type used during the packetization process, as well as, the coding parameters used (one or multiple NALUs per frame). The de-packetizer incorporates the proposed error resilience scheme.
3. MVC streamer – that encapsulates RTP packets into IP datagram's and creates concurrent UDP/IP connections to the client for multicast transmission of both views. Without loss of generality, vital information for the decoding process, including the Parameter Sets, is transmitted reliably over TCP [13].
4. MVC client – through a Graphical User Interface (GUI) transmits a request (based on transmission information including the number of views, the payload type and the MTU size) to the video streamer over TCP/IP connection. The client establishes UDP/IP connections (one connection in the case of AVC) with the streamer in order to receive IP datagram's of both views.

4.2. Experimentation Parameters

In order to obtain results that can be compared to those of previous studies, we selected similar set of coding parameters used in [4]. The codec characteristics, packetization and network parameters are summarized in Table 1. Each video transmission is repeated for 15 times to obtain the average PSNR values. STAP-A payload structure is not used in conjunction with 1 NALU per frame, since this would cause multiple frames to be encapsulated into a single RTP packet. Additionally, FU-A structure is not applied in the case of multiple NALUs per frame, since further fragmentation of NALUs into RTP packets is redundant.

Table 1. Simulation parameters

Coding Parameters				
Video Sequence	Flamenco		Objects	
No of Frames	1000		624	
Intra period	5 frames		5 frame	
Frame rate	25 fps		25 fps	
Resolution	640x480 pixels		640x480 pixels	
Packetization Parameters				
Video packetization options	1 NALU / Frame		Multiple NALUs/Frame	
RTP packetization options	SNU	FU-A	SNU	STAP-A
Network Parameters				
MTU Size (Bytes)	1024		512	
Packet loss rate	0% , 1% , 2% , 5%			
Error Model	Both Views		Base View	

5. EXPERIMENTAL RESULTS

5.1. Overhead

Fig.5a illustrates the overhead reduction for STAP-A, compared to the SNU mode as the number of the aggregated NALU's increases. Fig.5b shows the overhead introduced by FU-A versus SNU for different number of NALUs. It is shown that aggregating multiple NALUs in one RTP packet rather than performing RTP fragmentation optimizes overhead reduction.

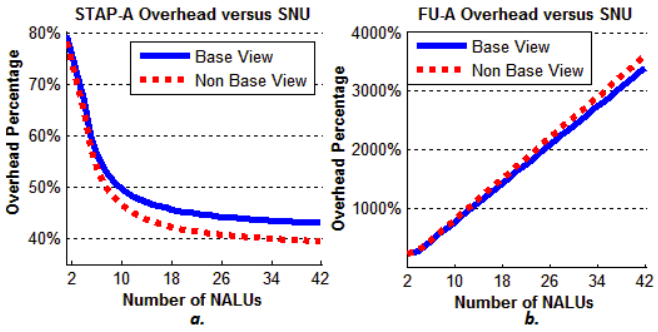

Fig.5. Overhead comparison between base and non-base views for different number of NALUs

5.2. Lost Frames

In the context of this study, a frame loss occurs when each RTP packet contains one NALU per frame using SNU mode. Fig.6 shows the number of frames that were decoded successfully after the transmission of two stereo video sequences, under various network conditions.

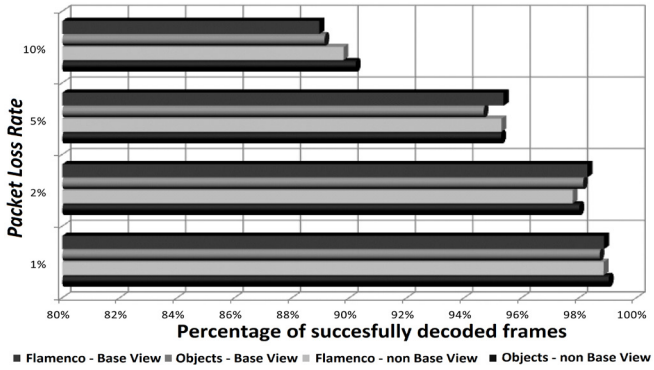

Fig.6.Decoded frames of base and non-base views in SNU mode under different packet loss rates

5.3. Quality with lost packets in both views

Fig.7 and Fig.8 illustrate PSNR of two stereo video sequences under different network conditions. During video transmission, when packet loss occurs in both views, additional error propagation to the non- base view exists due to its dependency with the base view. It can be seen that for both video sequences, the perceived quality in terms of PSNR of the base view is significantly better than the PSNR in the non-base view. This difference in the PSNR could only be eliminated if the non-base view was encoded independently, at the cost of coding efficiency. Moreover, PSNR increases as MTU size increases, since error resilience is better handled at the decoder. For the same packet loss rate, it is more preferable for the decoder to handle single packet loss (large MTU size) from error bursts (consecutive loss of smaller packets).

In particular, Fig.7a and Fig.8a illustrate PSNR by SNU mode and one NALU per frame packetization for both base and non-base view for two MTU sizes (512 bytes and 1024 bytes respectively). It can be seen that one frame per RTP encapsulation results in the lowest PSNR, among all packetization schemes. In this case, a lost packet that includes a NALU header causes an entire frame to be lost and results into increasing the received video distortion. The decoder implements frame copy concealment technique in the case of lost frames. Fig.7b and Fig.8b show PSNR in FU-A mode. The fragmentation of a NALU in parallel with the proposed error-resiliency scheme results in increased PSNR, especially in the cases of *1%* and *2%* packet loss. In this case, all frames are successfully decoded, thus the perceived video quality is significantly increased. However, the encapsulation of a frame into multiple NALUs at both modes, reduces the distortion effect due to packet loss in the received video as shown in Fig.7c,d and Fig.8c,d. Finally, when frame fragmentation into multiple NALUs is used, then the MTU size has only a limited impact on the PSNR.

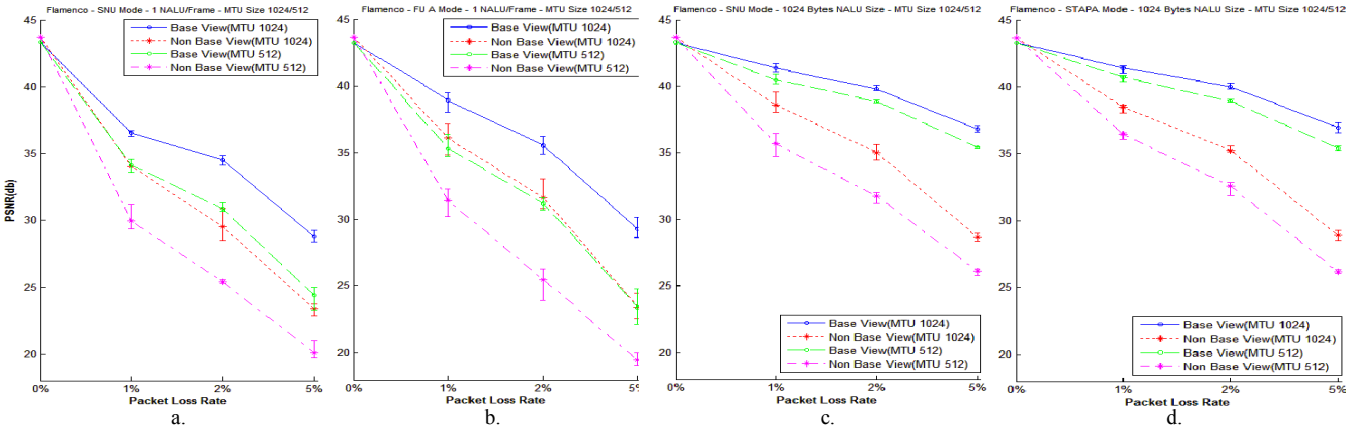

Fig.7. PSNR versus packet loss under various packetization schemes and MTU sizes for both base and non-base views (flamenco)

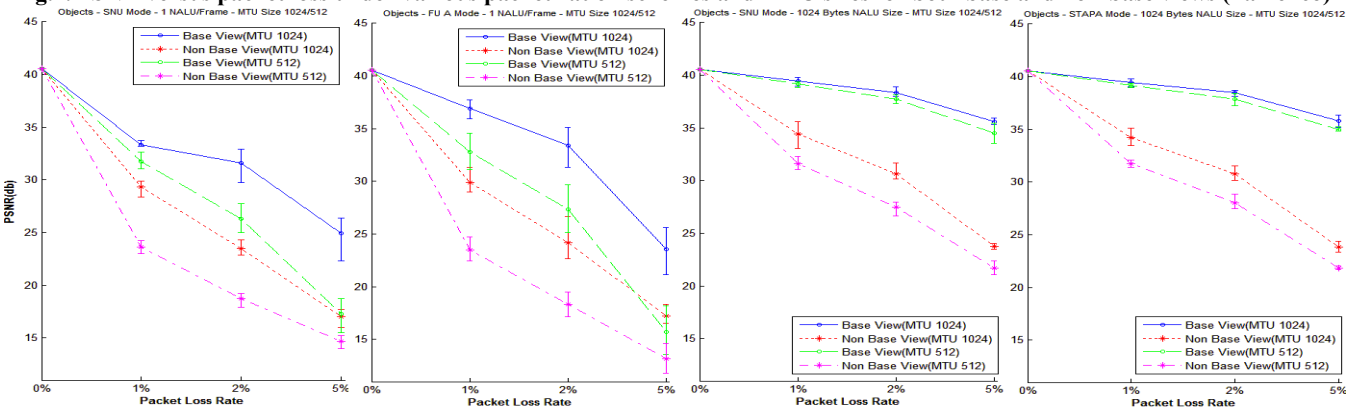

Fig.8. PSNR versus packet loss under various packetization schemes and MTU sizes for both base and non-base views (Objects)

5.4. Quality with lost packets in base view

Table 2 includes the PSNR when errors occur only in the base view. As shown in the following table, *1%* of packet loss in the base view leads to *0.65dB* and *1.03dB* drop of PSNR at the non-base view at FLAMENCO and OBJECTS video sequences respectively (case 1024 MTU, 1NALU/ frame, SNU). This means that although there are no packet losses in the non-base view, PSNR quality in the non-base view deteriorates due to the coding dependencies from the base view. Similar to the previous case, PSNR is optimized by considering more NALUs/frame in both RTP modes.

Table 2. Average PSNR for base and non-base view versus Packet Loss in base view

		FLAMENCO				OBJECTS			
		1024 MTU		512 MTU		1024 MTU		512 MTU	
		Base View	n-base View	Base View	n-base View	Base View	n-base View	Base View	n-base View
	PLR	PSNR							
1 NAL per Frame + SNU	0%	43.31	43.71	43.30	43.70	40.57	40.56	40.57	40.56
	1%	36.72	43.06	34.32	42.94	33.71	39.53	32.63	38.98
	2%	34.88	42.88	31.34	42.30	32.96	38.85	27.75	37.61
	5%	29.23	40.97	24.00	39.98	26.38	35.75	18.73	33.87
1 NAL per Frame + FU-A	0%	43.31	43.71	43.31	43.70	40.57	40.56	40.57	40.56
	1%	39.50	43.59	35.90	43.44	37.77	39.98	34.54	39.31
	2%	35.72	43.41	31.49	43.22	37.06	39.88	29.53	38.07
	5%	29.96	42.93	24.67	42.35	26.04	37.60	18.06	33.65
more NAL per Frame + SNU	0%	43.28	43.68	43.28	43.68	40.55	40.55	40.55	40.55
	1%	41.61	43.63	40.94	43.63	39.77	40.42	39.02	40.27
	2%	39.93	43.54	38.88	43.59	38.11	40.08	38.14	40.04
	5%	36.76	43.49	35.43	43.41	35.95	39.50	34.74	39.26
more NAL per Frame + STAP-A	0%	43.28	43.68	43.28	43.68	40.55	40.55	40.55	40.55
	1%	41.61	43.64	40.75	43.63	39.73	40.42	39.41	40.27
	2%	40.27	43.62	38.98	43.59	38.69	40.08	38.21	40.08
	5%	37.14	43.50	35.43	43.41	36.15	39.58	35.22	39.38

6. CONCLUSIONS

This papers aims to assess the performance of 3D video streaming over lossy IP based networks using various video packetization modes according to the H.264/MVC standard. Particular interest was given to frame fragmentation into multiple NAL units and FUs. In order to recover missing header information, an error resilient scheme is proposed that uses the FU-A mode to reconstruct the H.264/AVC NALU Header. It is shown that the best MVC packetization mode in terms of error propagation elimination and PSNR is the frame fragmentation in multiple NALUs at the encoder's process for both SNU and STAP-A modes. The performance evaluation considers parameters like the overhead caused by different packetization schemes, the number of decoded frames and the resulted video quality in terms of PSNR for both base and non-base views.

7. ACKNOWLEGMENT

This work has been supported in part by ROMEO Integrated Project (www.ict-romeo.eu), funded under the European Commission 7th Framework Programme.

8. REFERENCES

[1] A.Smolic, K.Mueller, P.Merkle, C.Fehn, P.Kauff, P.Eisert, and T.Wiegand, "3D Video and Free Viewpoint Video - Technologies, Applications and MPEG Standards", *IEEE ICME*, Jul 2006.

[2] B.W.Micallef, C.J.Debono, "An analysis on the effect of transmission errors in real-time H.264-MVC Bit-streams", MELECON, Apr 2010.

[3] Y. Chen, Ye-Kui Wang, K. Ugur, M. M. Hannuksela, J. Lainema , M. Gabbouj, "The Emerging MVC Standard for 3D Video Services", Eurasip Journal on Advances in Signal Processing, No. 8, 2009, DOI: 10.1155/2009/786015.

[4] Z. Liu, Y. Qiao, B. Lee, E. Fallon, Karunakar A. K., C. Zhang, S. Zhang, "Experimental Evaluation of H.264/Multiview Video Coding over IP Networks", ISSC, Trinity College Dublin, June 23-24, 2011.

[5] ITU-T Recommendation, "H.264 - Advanced video coding for generic audiovisual services", Mar 2010.

[6] H. Schwarz, D. Marpe and T. Wiegand, "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard", IEEE Circuits and Systems for Video Technology, Vo. 17, No. 9, September 2007.

[7] Y. Wang and T. Schierl, "RTP payload format for MVC video", IETF Internet Draft, (accessed March 2011).

[8] K.-D. Seo, J.-S. Kim, S.-H. Jung, and J.-J. Yoo, "A Practical RTP Packetization Scheme for SVC Video Transport over IP Networks", ETRI Journal, No. 2, April 2010, DOI: 10.4218/etrij.10.1409.0031.

[9] S. Liu, Y. Chen, Y. Wang, M. Gabbouj, M. M. Hannuksela, H. Li, "Frame Loss Error Concealment For Multiview Video Coding", June 2008, DOI: 10.1109/ISCAS.2008.4542206 .

[10] Y. Wang, R. Even, T. Kristensen, R. Jesup, "RTP payload format for H.264 video",IETF Internet Draft, (accessed March 2011).

[11] A. Hallapuro, M. Hannuksela, J. Lainema, M. Salmimaa, K. Ugur, K. Willner, Y. Wang, Y.Chen, "Nokia "3D Video & The MVC Standard", Nokia research center , May 2009.

[12] L., Rizzo, "Dummynet: a simple approach to the evaluation of network protocols," ACM SIGCOMM Computer Communication Review, 27 (1), 31-41, 1997.

[13] C.G. Gurler, B. Gorkemli, G. Saygili, M. Tekalp, Flexible transport of 3D video over network , Proceedings of the IEEE, 99 (4) (April 2011).