Experimental Evaluation of H.264/Multiview Video Coding over IP Networks

Zhao Liu[*], Yuansong Qiao[*], Brian Lee[*], Enda Fallon[**], Karunakar A. K.[*], Chunrong Zhang[*], Shuaijun Zhang[*]

**Software Research Institute, Athlone Institute of Technology, Ireland*

***School of Engineering, Athlone Institute of Technology, Ireland*

email: {zliu, ysqiao}@research.ait.ie, {blee, efallon}@ait.ie, {akkarunakar,crzhang,szhang}@research.ait.ie

***Abstract* — Multiview Video Coding (MVC) is an extension of the H.264/AVC standard. It is designed to encode Multiview videos for immersive multimedia applications, e.g. Tele-immersion. An MVC video contains multiple views. MVC uses inter-view prediction to reduce picture redundancy, which introduces dependency between views. This paper studies the effect of network packet loss on the quality of streamed MVC videos, especially on the quality of different types of MVC view. An experimental platform is designed and implemented to perform the tests. Various packet loss rates are considered during the evaluation. The effect of packetization is also investigated in experiments. The test results reveal the impact of packet loss is different on different views. The quality of bi-predicted view degrades more significantly compared to other types of view. This is not acceptable for many real time applications.**

***Keywords* – H.264/MVC Streaming, Video Quality Evaluation, Multiview Coding, JMVC**.

I. INTRODUCTION

Multiview video coding is one of the key technologies used to provide immersive experience to users. It can support a wide range of applications, e.g. stereoscopic video, free viewpoint television, and multiview 3D television. Multiview video can be divided into two categories, i.e. three-dimensional (3D) video and free-viewpoint video. Both of them have gained significant attention in recent years. 3D video can provide users with an impression of depth of the observed scene. In free-viewpoint video, a user can interactively navigate within a 3D scene while watching a video, i.e. change the viewpoint and the viewing direction [1].

Multiview video applications generate a huge amount of data while capturing videos from multiple cameras. An efficient video compression algorithm is required to reduce complexities in encoding, decoding and transmitting of multiview videos. In order to tackle these problems, the Joint Video Team (JVT) [2] developed H.264/Multiview Video Coding (MVC) [3] [4] based on the widely deployed standard H.264/Advanced Video Coding (AVC) [3].

MVC can achieve high coding gain by using inter-view prediction and intra-view prediction based on a hierarchical B prediction structure [5]. In MVC, views can be divided into 3 types: 1) Base View, which is encoded without referencing to other views and can be decoded independently; 2) Predicted View, which is encoded by referencing to either of the preceding two views; 3) Bi-predicted View, which is encoded by referencing to two adjacent views. This scheme which mixed intra-view and inter-view prediction makes the coded bitstream very vulnerable to transmission errors.

In [6], the authors studied the quality of streamed MVC video under wireless environments, where the packets are corrupted but not dropped. The results show that transmission errors drastically reduce the quality of the reconstructed 3D video. However, the effect of transmission errors on different types of view is not clarified.

This paper studies the MVC video quality streamed over IP networks. For video streaming, network delay, jitter and loss will all affect on the quality of a streamed video. As the effects of delay and jitter can be avoided by an appropriate buffering scheme, this paper focuses on the effect of packet loss. It examines the extent to which packet loss affects different types of view in MVC videos. A number of tests covering various loss rates are performed. Standard video test sequences recommended by Joint Video Team (JVT) representing different scenes are used in the tests. Both effects of packetization and non-packetization are investigated. The quality of the streamed MVC video is measured by Peak Signal-to-Noise Ratio

(PSNR) - a commonly used objective video quality evaluation method.

The rest of the paper is organized as follows: The related work and necessary knowledge are introduced in Section II; Section III describes the testbed design and simulation setup; Section IV presents the test results and analysis; Conclusions and future work are given in Section V.

II. RELATED WORK

Streaming H.264/AVC video over IP networks is studied in [7]. Transmission of H.264/MVC video over wireless networks is evaluated in [6] as introduced in Section I. Currently, there is dearth of error concealment algorithms for stereoscopic or multiview video [8]. To the authors' knowledge, no research has been done to reveal the unequal effects of packet loss on different types of MVC view.

The following section will introduce necessary knowledge about MVC related to this paper. For detailed explanation, please refer to [3].

a) H.264/Multiview Video Coding

MVC is defined in Annex H of H.264 /AVC standard [3], which is backward compatible to H.264/AVC. An MVC sequence consists of many views captured from multiple cameras to cover different aspects of a scene. These cameras are normally placed very close to support 3D applications. Therefore, spatial similarities exist between these views. MVC uses hierarchical B prediction structure [4] to remove redundancy between views to enhance multiview video compression efficiency. In MVC, all types of view uses intra-view prediction as defined in H.264/AVC. The difference is that some types of view use inter-view prediction.

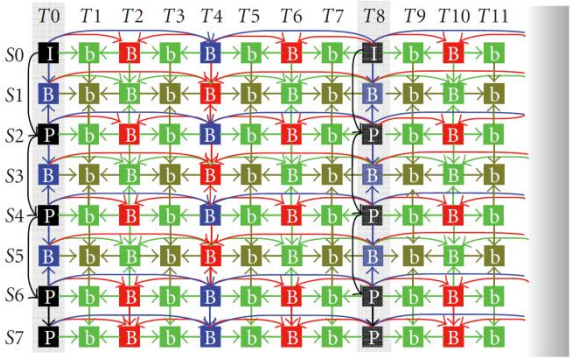


Figure 1: MVC Hierarchical B Prediction Structure [4]

Three types of view are defined in MVC, as illustrated in Figure 1.

Base View: It is coded independently using only intra-view prediction E.g. View S0 in Figure 1.

Predicted View: It is encoded based on a previous reference view. However, only the picture at the boundary of a Group of Picture (GOP) is predicted from either of the preceding two views. The pictures between the boundaries of a GOP are not predicted from other views. E.g. View S2, S4, S6 and S7 in Figure 1.

Bi-predicted View: It is predicted based on both the previous view and the next view. All pictures are coded using inter-view prediction and intra-view prediction. E.g. View S1, S3 and S5 in Figure 1.

All the views are coded as an MVC video bit-stream which is organized into Network Abstraction Layer (NAL) Units. 32 types of NAL unit are defined in the MVC Standard. They are classified into Video Coding Layer (VCL) and non-VCL NAL units. The coded picture information is contained in VCL NAL units. The associated additional information for enhancing usability of the decoded video signal is contained in non-VCL NAL units.

b) Video Quality Evaluation - PSNR

This paper uses PSNR to evaluate the quality of streamed video. PSNR is a well-known objective measurement to evaluate video quality. I think it is sufficient to estimate the impacts of packet loss on different MVC views. The PSNR is defined as:

$$\text{PSNR} = 10 \times \log_{10}\left(\frac{MAX_I^2}{MSE}\right) \qquad (1)s$$

Mean Squared Error (MSE) is between two $m \times n$ monochrome image I and K where one of images is considered a noisy approximation of the other. MSE is defined as:

$$\text{MSE} = \frac{1}{mn}\sum_{i=0}^{m}\sum_{j=0}^{n}[I(i,j)-K(i,j)]^2 \qquad (2)$$

MAX_I is the maximum possible pixel value of the image, e.g. the MAX_I value is 255 for images with 8–bit colour depth.

c) Mapping from PSNR to MOS

Mean Opinion Score (MOS) is a subjective measurement to evaluate the video quality. A general mapping of PSNR to MOS is shown as below in Table 1.

Table 1: PSNR to MOS Mapping Table [1]

PSNR (dB)	MOS	Perceived Quality	Impairment
> 37	5	Excellent	Imperceptible
31 - 37	4	Good	Perceptible, but not annoying
25 - 30	3	Fair	Slightly annoying
20 - 24	2	Poor	Annoying
< 20	1	Bad	Very annoying

III SIMULATION PLATFORM & SETUP

a) Simulation Platform Design

The simulation testbed is designed based on VLC media player (VLC) [10] and the MVC reference software – JMVC version 8.3.1 [11].

VLC is a free and open source cross-platform multimedia player and framework that supports most multimedia formats as well as various streaming protocols. It is very convenient for programmers to modify and add new functions depending on their requirements. JMVC software is the reference software for the MVC project of the JVT. Among six tools provided by JMVC, this paper uses the encoder, decoder and PSNR tools.

In the implemented testbed, the MVC decoder is extracted from the JMVC software and integrated into VLC player as a plug-in. A network simulator is added into VLC to simulate network packet loss. The simulation architecture and simulation process are shown in Figure 2.

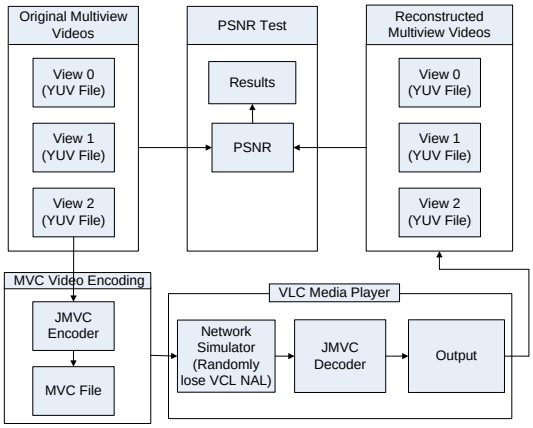


Figure 2: Simulation Architecture

Three original YUV videos representing different views of an MVC sequence are encoded into a MVC file. The VLC player reads the MVC bit stream and parses it into NALs. The Network Simulator discards the NALs (or packetized NALs) with a specified packet loss rate. The remaining NALs are passed to the MVC decoder. The reconstructed pictures are stored into YUV files by the Output module. If a picture is totally lost and could not be decoded, a blank YUV picture is inserted into the corresponding YUV file in correct order. Finally, the streamed video quality is evaluated by using the PSNR tool to compare the original YUV videos with the reconstructed YUV videos.

b) Simulation Parameters

In the tests, only VCL NALs are dropped and non-VCL NALs are kept intact. Some important non-VCL NALs, e.g. sequence parameter set and picture parameter set, could be transmitted reliably by a different transport mechanism in advance of sending the VCL NALs that refer to them. Other non-VCL NALs are not necessary for decoding video pictures.

Five standard MVC test sequences are used in the simulation in order to cover a wide range of scene types. The video sequence parameters are listed in Table 2. These video sequences are encoded by JMVC reference software version 8.3.1 with the same encoder parameters as shown in Table 2.

To represent the three types of MVC view, the first three views (View 0, 1, 2) of each video sequence are selected to be encoded as Base View, Bi-predicted View and Predicted View respectively as shown in Figure 3.

In the tests, the packet loss rate varies from 0% to 10%. VCL NALs of different views in the test video sequences are dropped randomly. For each loss rate, 20 tests are conducted, and finally the mean PSNR value of each view is calculated.

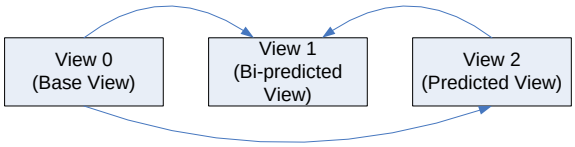


Figure 3: View Dependency

Table 2: Picture Feature and Encoding Parameter

Sequence	Object Motion	Camera Motion	Background Motion	Frame Size	Frame Rate (fps)	GOP Length	No. of Frame
Ballroom	Medium	No	No	640x480	25	4	250
Exit	High	No	No	640x480	25	4	250
Vassar	Low	No	No	640x480	25	4	250
Flamenco2	Medium	No	Yes	640x480	25	4	250
Racel	Medium	Yes	Yes	640x480	25	4	250

IV EXPERIMENTAL RESULTS & ANALYSIS

Three groups of test are performed. The first two groups of tests show the effect of non-packetization and packetization. The third test shows the effect of errors due to dependency between views.

a) Test 1: Transmitting MVC without Fragmentation

The NAL is specified for network transmission according to the MVC standard. An MVC video bit-stream is split and encapsulated in NALs, each one of which, can be carried in a separate packet. In the test group, for each test, the NALs are dropped randomly according to the specified loss rate. No error concealment tools are used in the tests.

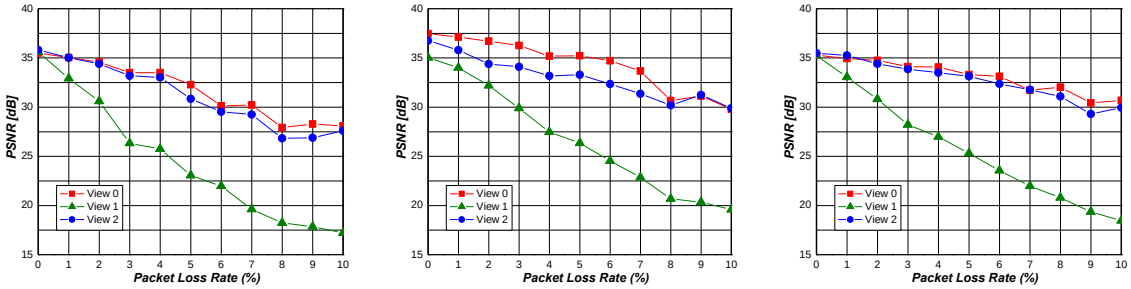
The experiment results are given in Figure 4. The PSNR results of the three views for each video sequence are shown in a separate diagram. Two phenomena can be observed from the diagrams.

1) The PSNR values for all views decrease while the packet loss rate increases. However, the PSNR values of View 1 in all the sequence diagrams fall dramatically compared to View 0 and View 2. For example in Figure 4 (d) – Flameco2 Sequence, when loss rate is 10%, the perceived quality (as per Table 1) of View 1 becomes "very annoying" (MOS=1, PSNR<20dB), whereas the qualities of View 0 and View 2 are both in "good" condition

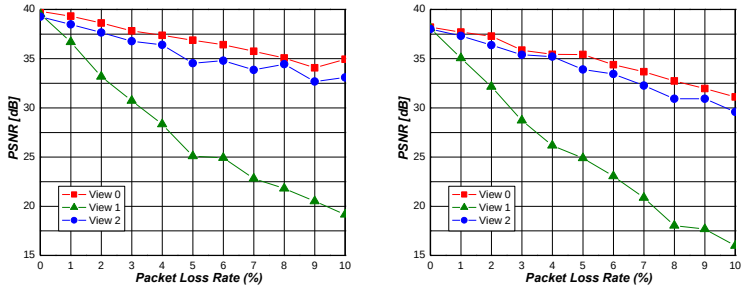
(MOS=4, PSNR>31). The mapping between MOS and PSNR is shown in Table 1.

This creates a problem for free-viewpoint applications. When the user switches from one view to another, the perceived video quality may fluctuate drastically.

2) High background motions in the videos cause larger quality difference between View 1 and the other two views. In the five test sequences, the first three videos (Figure 4 a, b, c) have static backgrounds, whereas the last two videos (Figure 4 d, e) contain background motion. The difference between the PSNR values of View 1 and that of View 0 and 2 in Figure 4 (d and e) is significantly larger than the difference in Figure 4 (a, b, c).

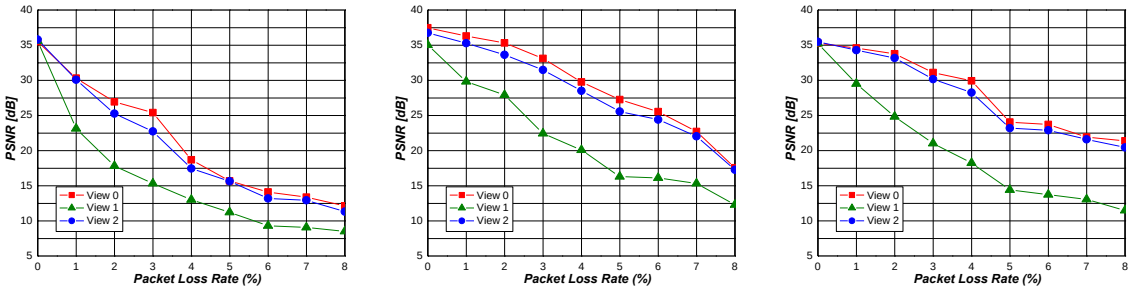


(a) Ballroom Sequence (b) Exit Sequence (c) Vassar Sequence

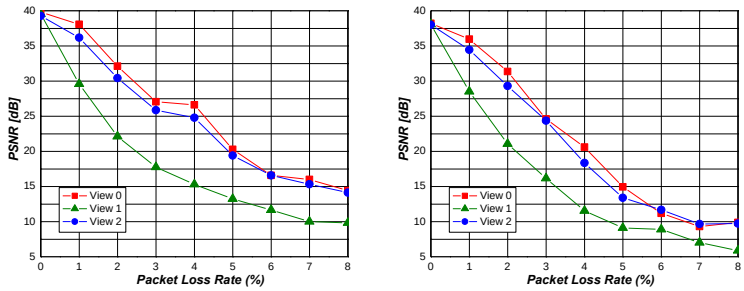


(d) Flameco2 Sequence (e) Race1 Sequence

Figure 4: Test Results for Transmission without Fragmentation

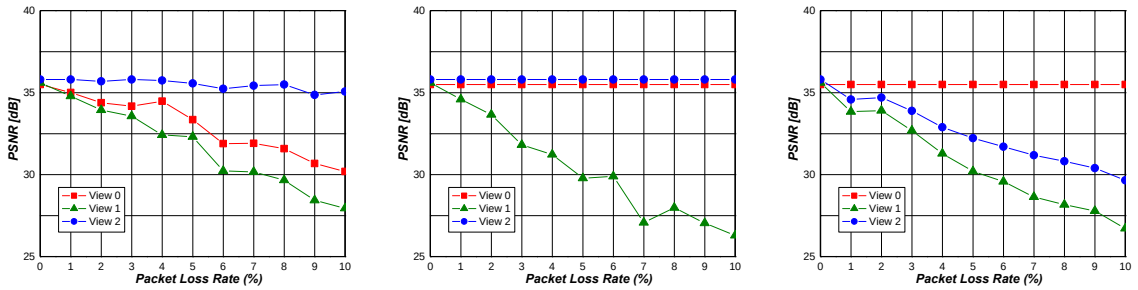


(a) Ballroom Sequence (b) Exit Sequence (c) Vassar Sequence



(d) Flamenco2 sequence (e) Race1 Sequence

Figure 5: Test Results for Transmission with Fragmentation



(a) Losses in View 0 only (b) Losses in View 1 only (c) Losses in View 2 only

Figure 6: View Dependency Test for Ballroom Sequence

b) Test 2: Transmitting MVC with Fragmentation

In the previous section, packetization is not considered in the tests. When a video sequence is transmitted over the Internet, video data is segmented into packets subjected to the size of network Maximum Transmission Unit (MTU).

In this test group, the MTU is set as 1500 Bytes. The NALs retrieved from the MVC file (Figure 2) are fragmented into packets with maximum size of 1500 Bytes. The Network Simulator module (Figure 2) performs network loss simulations on the packets and whenever any packet of a NAL is lost, the NAL can not be recovered.

Ideally, a picture should be encoded into slices (not exceeding MTU). Consequently, each slice can be encapsulated into a separate NAL which will be further packetized without fragmentation. However, the current JMVC implementation only supports a single slice per picture. Therefore the above fragmentation scheme is chosen.

The experimental results are shown in Figure 5. In the tests, the loss rates range from 0 to 8%. The results for 9% and 10% are not shown here because the current implementation of JMVC software crashes very frequently at such loss rates.

As the results shown in the previous section, the quality of View 1 is much lower than that of View 0 and View 2. However, there are 2 differences:

1) The PSNR values for all views are significantly smaller than that in the first test group. This is caused by NAL fragmentation. Any losses of fragmentation will result in a NAL loss. Suppose a NAL is divided into N packets. For a given loss rate M%, the probability for losing this NAL is (M*N) %.

2) The quality gap between View 1 and View 0 & 2 increases first and then decreases. This is caused by the PSNR curve property (Equation 1) that PSNR decays exponentially. Therefore, when the qualities of the three views are all low enough, the PSNR curves will be closer to each other.

For example, in Figure 5 (a), the PSNR value of View 1 declines to 8 dB when the packet loss rate is 6%. Afterwards, the PSNR value decreases very slowly at packet loss rate of 7% and 8%.

c) Test 3: View Dependency Test

In the two test groups above, the quality of View 1 is significantly lower than that of View 0 and 2, which is caused by the dependency between views. This section studies how much a view depends on other views in the case of network losses.

In the test group, three tests are executed. For each test, only one view experiences packet loss. No direct packet loss happens in the other two views. The test results are shown in Figure 6.

In Figure 6 (a), packet loss only happens in View 0. Consequently the quality of View 0 decreases while the loss rate increases. However, the quality of View 1 degrades simultaneously and is dramatically lower than that of View 0. On the contrary, the quality of View 2 is slightly affected by the loss on View 1.

In Figure 6 (b), it shows that loss on View 1 has no effect on View 0 and 2.

Figure 6 (c) is similar to Figure 6 (a). The loss on View 2 significantly affects on View 1, but has no effect on View 0.

This is due to the hierarchal B prediction structure of MVC. A detailed explanation is given below:

1) View 0 is the Base View. It uses intra-view prediction only. Therefore the loss on View 1 and 2 do not affect on View 0.

2) View 1 is a Bi-predicted View. It uses inter-view prediction and intra-view prediction at the same time to compress video. Any loss in View 0, View 2 and View 1 will affect on the quality of View 1.

3) View 2 is a Predicted View. It uses inter-view prediction and intra-view prediction to compress video. However, inter-view prediction is only used in the boundary of a GOP. Only when the loss of View 0 happens on the boundary pictures, it affects on the quality of View 2. Therefore, the loss in View 0 has light effect on the quality of View 1.

V CONCLUSIONS & FUTURE WORK

This paper studied MVC video streaming in IP environments. It focuses on the effect of network packet loss on MVC hierarchical B prediction structure and investigates the quality difference

between the streamed views. An evaluation testbed based on VLC and JMVC is developed. The effects of loss rate and packetization are both considered in the tests.

The results show that the quality of bi-predicted views degrades significantly faster than that of base views and predicted views when the loss rate increases. For example in the Ballroom test of Test Group 1, the PSNR value for view 1 (bi-predicted view) is lower than those of view 0 (based view) and view 2 (predicted view) about 10 dB in the case of 10% loss rate (As shown in Figure 4). It is also shown that greater background motion in a video sequence causes larger difference between the bi-predicted view and the two other views.

This paper identifies that the above problem is caused by the MVC design rationale, i.e. applying the single view design philosophy directly to multiview video coding. However, in many multiview scenarios, the videos captured from multiple cameras are equally important. Consequently, the views should be reconstructed uniformly in the display side.

The problems found in this paper could be tackled from different perspectives, e.g. improving the coding scheme, using appropriate error concealment algorithms, and designing packet dropping algorithms for MVC streams inside the network elements. The results of this paper could be applicable to the transmission of H264/MVC video over network when the packets are corrupted but not discarded. In future work, we will verify this modification and also evaluate and propose possible solutions for improving MVC streaming quality.

ACKNOWLEDGEMENTS

This Research is supported by AIT President's Seed Fund 2010 and by Enterprise Ireland through its Applied Research Enhancement Fund –SUNAT.

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 International Conference on Multimedia and Expo (ICME)*, Jul 2006.

[2] http://www.itu.int/en/ITU-T/studygroups/com16/video/Pages/jvt.aspx

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

[4] Y.Chen, Y.K.Wang, K.Ugur, M.M.Hannuksela, J.Lainema, and M.Gabbouj, "The emerging MVC standard for 3D video services", *EURASIP Journal on Applied Signal Processing - 3DTV: Capture, Transmission, and Display of 3D Video*, vol. 2009, Jan 2009.

[5] P.Merkle, K.Müller, A.Smolic, and T.Wiegand, "Efficient Compression of Multi-View Video Exploiting Inter-View Dependencies Based on H.264/MPEG4-AVC", *IEEE International Conference on Multimedia and Expo (ICME)*, Jul 2006.

[6] B.W.Micallef, C.J.Debono, "An analysis on the effect of transmission errors in real-time H.264-MVC Bit-streams", *15th IEEE Mediterranean Electrotechnical Conference (MELECON)*, Apr 2010.

[7] S.Wenger, "H.264/AVC Over IP", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, JUL 2003.

[8] K.Song, T.Chung, Y.Ohb, C.S.Kim, "Error concealment of multi-view video sequences using inter-view and intra-view correlations", *Journal of Visual Communication and Image Representation*, vol. 20 Issue 4, May 2009.

[9] J. Klaue, B. Rathke and A. Wolisz, "EvalVid – A Framework for Video Transmission and Quality Evaluation", *Proceedings of the 13th International Conference on Modeling, Techniques and Tools for Computer Performance Evaluation*, Urbana, Illinois, 2003.

[10] VLC, http://www.videolan.org/vlc/

[11] JMVC 8.3.1 Reference Software, 2010.