Subjective Video Quality Assessment in the H.264/AVC Video Coding Standard

Zoran Miličević and Zoran Bojković

Abstract — **This paper seeks to provide an approach for subjective video quality assessment in the H.264/AVC standard. For this purpose a special software program for the subjective assessment of quality of all the tested video sequences is developed. It was developed in accordance with recommendation ITU-T P.910, since it is suitable for the testing of multimedia applications. The obtained results show that in the proposed selective intra prediction and optimized inter prediction algorithm there is a small difference in picture quality (signal-to-noise ratio) between decoded original and modified video sequences.**

Keywords — **H.264/AVC standard, peak signal-to-noise ratio, subjective video coding quality assessment.**

I. Introduction

DIGITAL video is more prevalent in everyday life through the many video applications such as digital television, digital cinema, internet video, videoconferencing, Youtube, video on demand, home video, etc. Digital video typically goes through several stages during processing, before it is available to end users. Very often, the end user is the observer. One of the effects of many stages in the processing of digital video is the video quality degradation. Accordingly, various methods for the assessment of video quality have an important role in monitoring the quality and to ensure the required quality of service, evaluation of equipment performance to display video, the evaluation of system for video processing, as well as to define the optimal design of perceptual systems for video processing.

Generally speaking, the method for quality assessment can be divided into objective and subjective.

Objective methodology uses mathematical models to depict the behavior of the human visual system, and is based on separating the characteristic parts of the bit sequence.

Subjective assessment of video quality (Subjective Video Quality Assessment, SVQA) presents a methodology for video quality assessement that was received (observed) by observers and give opinion (court) about the video that the observer is watching. In this sense, subjective methodology used by the group entities ("observers") displaying decoded video sequences. They need to assess the quality of these sequences with the use of appropriate assessment methodology.

The sum of their score gives the *Mean Opinion Score* (MOS), which represents a measure of subjective quality assessment.

Because the experiments with the subjective video quality assessment are more complicated from psychological aspects and conditions of observation process, different methodologies for the subjective quality assessment are defined.

In order to provide better compression of video, compared to previous standards, the H.264 MPEG-4 part 10 video coding standard was developed by the Joint Video Team (JVT) [1]-[5]. The H.264 fulfills significant coding efficiency, simple syntax specifications and seamless integration of video coding into all current protocols and multiplex architectures. Thus the H.264 can support various applications like video broadcasting, video streaming, video conferencing over fixed and wireless networks and over different transport protocols [6].

Technically, the design of the H.264/MPEG4-AVC video coding layer is based on the traditional hybrid concept of block-based motion-compensated prediction (MCP) and transform coding. Within this framework, a number of important innovative ideas have been developed [7].

The improvement in coding performance comes mainly from the prediction part. Intra prediction significantly improves the coding performance of H.264/AVC [8] intra frame coder. On the other side, inter prediction is enhanced by motion estimation with quarter-pixel accuracy, variable block sizes, multiple reference frames and improved spatial/temporal direct mode.

H.264/AVC defines a set of Profiles, each supporting a particular set of coding functions and each specifying what is required of an encoder or decoder that complies with the Profile. Also, profiles are defined to cover the various applications from the wireless networks to digital cinema.

In 2004 the Joint Video Team (JVT) added new extensions known as the Fidelity Range Extensions (FRExt), which provide a number of enhanced capabilities relative to the base specification [9]. Also, the Scalable H.264/AVC extension is applied to extend the hybrid video coding approach of H.264/AVC in a way that a wide range of spatial-temporal and quality scalability is achieved [10], [11]. The next major feature added to the standard was Multiview Video Coding (MVC).

This paper is organized as follows. Section II provides motivation for our work. Section III describes a methodology for subjective video quality assessment in the H.264/AVC standard. Section IV contains experimental results and discussion. Section V concludes the work.

Zoran Miličević, Telecommunication and IT Department GS of the SAF in Belgrade, Raska 2, 11000 Belgrade, Republic of Serbia (phone: +381-64-1191125; e-mail: mmilicko@eunet.rs).

Zoran Bojković, University of Belgrade, 11000 Belgrade, Republic of Serbia (phone: +381-11-2605-045; e-mail: z.bojkovic@yahoo.com).

II. Motivation

In order to reduce the temporal and spatial redundancy more effectively, motion compensation uses variable block sizes, while directional intra prediction investigates all available coding modes to decide the best one. When the rate-distortion optimization technique is used, an encoder finds the coding mode having a minimum rate-distortion cost. Due to the large number of coding modes combinations, the decision process requires extremely high computation. To reduce the complexity, a selective intra prediction and optimized inter prediction algorithm for Standard Definition and High Definition test sequences is proposed, when P and B pictures are analyzed. Simulation results show that the coding time is reduced by more than 15% on the average, depending on Quantization Parameter (QP) values, when reducing the number of intra and inter candidate modes. However, there is a negligible loss in term signal-to-noise ratio (SNR), until bit rate values slightly fluctuate depending on QP values.

The proposed method is used for P and B pictures processing in the IBBP structure for QP= 24, QP=26 and QP=32, respectively.

Based on the related work, we focused on the fact that the proposed algorithm in different test scenarios gave a negligible loss in term Δ PSNR.

For example, when all analyzed test sequences were in SD resolution, there was a negligible loss in term ΔPSNR for luma component of picture: it was only 0.49, 0.47 and 0.50 dB on the average, depending on QP values.

On the other hand, when all analyzed test sequences were in HD resolution, there was a negligible loss in term ΔPSNR for luma component of picture: it was only 0.40, 0.40 and 0.44 dB on the average depending on QP values.

Finally, when all analyzed test sequences were in full HD resolution, there was a negligible loss in term ΔPSNR for luma component of picture: it was only 0.29, 0.32 and 0.36 dB on the average depending on QP values [12].

This represents the main motivation that dominates in this paper. We would like to receive the subjective video quality metrics from subjective video quality assesment process and make correlation with the results of simulation analysis.

III. Subjective Video Quality Assessment in the H.264/AVC Standard

For subjective video quality assessment, recommendation ITU-T P.910 (04/2008) is used.

Recommendation ITU-T P.910 is intended to define non-interactive subjective assessment methods for evaluating the quality of digital video images coded at bit rates specified in classes for TV3, MM4, MM5 and MM6, for applications such as videotelephony, videoconferencing as well as storage and retrieval applications. The methods can be used for several different purposes including, but not limited to, selection of algorithms, ranking of video system performance and evaluation of the quality level during a video connection [13].

Measurement of the perceived quality of images requires the use of subjective scaling methods. The one of methods which is described in recommendation ITU-T P.910 is *Degradation Category Rating* (DCR). DCR uses explicit references compared to the methods that do not use any explicit reference, e.g., Absolute Category Rating (ACR), Absolute Category Rating with Hidden Reference (ACR-HR), and Pair Comparison Method (PC).

The DCR method should be used when testing the fidelity of transmission with respect to the source signal. This is frequently an important factor in the evaluation of high quality systems. The specific comments of the DCR scale (imperceptible/perceptible) are valuable when the viewer's detection of impairment is an important factor. Thus, when it is important to check the fidelity with respect to the source signal, the DCR method should be used.

DCR should also be applied for high quality system evaluation in the context of multimedia communication. Discrimination of imperceptible/perceptible impairment in the DCR scale supports this, as well as comparison with the reference quality.

The degradation category rating (DCR) implies that the test sequences are presented in pairs: the first stimulus presented in each pair is always the source reference, while the second stimulus is the same source presented through one of the systems under test. This method is also called the double stimulus impairment scale method. The time pattern for the stimulus presentation can be illustrated by Fig. 1.

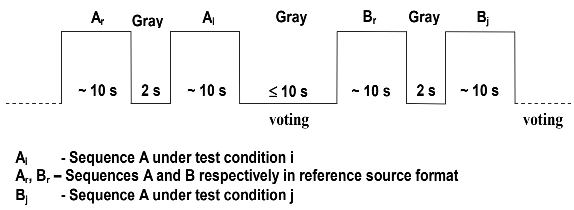


Fig. 1. Stimulus presentation in the DCR method [13].

If a constant voting time is used, then it should be less than or equal to 10 s. The presentation time may be reduced or increased according to the content of the test material.

The subjects are asked to rate the impairment of the second stimulus in relation to the reference one. The following five-level scale for rating the impairment should be used: 5 - *Imperceptible*, 4 - *Perceptible but not annoying*, 3 - *Slightly annoying*, 2 – *Annoying*, 1 - *Very annoying*. The necessary number of replications could be obtained for the DCR method by repeating the same test conditions at different points of time in the test.

IV. Experimental Results and Discussion

In order to evaluate the objective PSNR results in [12] on the subjective way, a software program is developed. It was developed in accordance with recommendation ITU-T P.910, since it is suitable for the testing of multimedia applications [13].

In the process of subjective video quality assessment, the following general conditions are taken into account [14]:

- The DCR method was used;

- Four different test sequences (*Blue sky, Pedestrian area, Riverbed and Rush hour*) in the SD (Standard Definitions), HD (High Definitions) and full HD test sequences were used in the YUV 4:2:0 video format;
- The sequence in a reference source format and sequence under the test condition are shown simultaneously on the projection board;
- The sequences are displayed in the two window display and put side by side within a 50% grey background;
- The sequences were perfectly synchronized which means that they both were started and stopped at the same frame;
- Each video sequence was repeated twice;
- The sequence in a reference source format was placed always on the left side, while the sequence under the test condition on the right;
- 20 observers belonging to the population aged from 24 to 27 participated in the experiment;
- Evaluation of image quality was not part of the work of observers, so they were not experienced estimators;
- Observers had normal visual acuity and normal colour recognition;
- In order to reduce the eye movement to switch the attention between the two windows, the viewing distance was 8H, where H indicated the picture height;
- Observers were used to assess the scale with five levels;
- Before starting the experiment scenario, applications that were found in the system under test are presented to observers;
- The descriptions of the evaluation type, as well as the assessment scale, together with simulation presentation are given in a written form.

User interface for the subjective assessment of quality of test sequences has been developed in accordance with the rules defined for the DCR test method (Fig. 2).

Before the testing procedure was carried out, first it was necessary to fulfill all conditions relating to the observers, and then they have accessed the personal evaluation. The testing included two phases. The first phase entailed the entry of sequence name and serial number of the observers and then entering (loading) the sequence in a reference source format as well as the sequence under the test condition in an appropriate format. It is shown in Fig. 3 for *Blue sky* test sequence as an example.

After that, the second phase was applied - processing and rating in subjective video quality assessment for *Blue sky* test sequence in Standard, High and full-High resolution for QP=28, as shown in Fig. 4.

Also, the second phase for *Rush hour* test sequence in all three different resolutions is shown in Fig. 5.

All ratings given by observers were stored in a *Microsoft Access* database.

All video test sequences which are used in objective and subjective video quality assessment process can be divided into two groups, depending on the camera position.

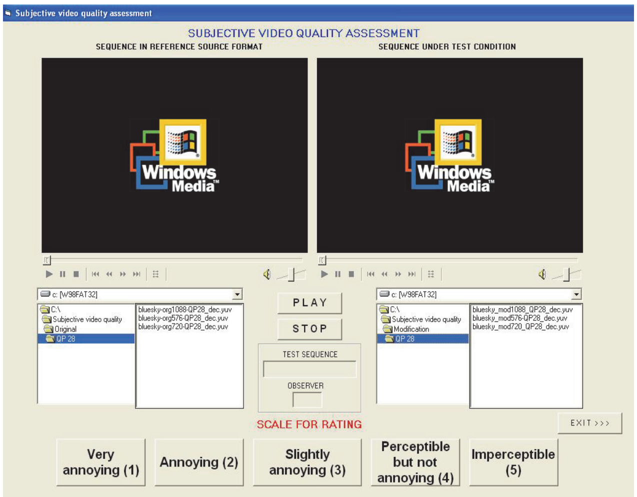


Fig. 2. User interface for subjective video quality assessment [14], [15].

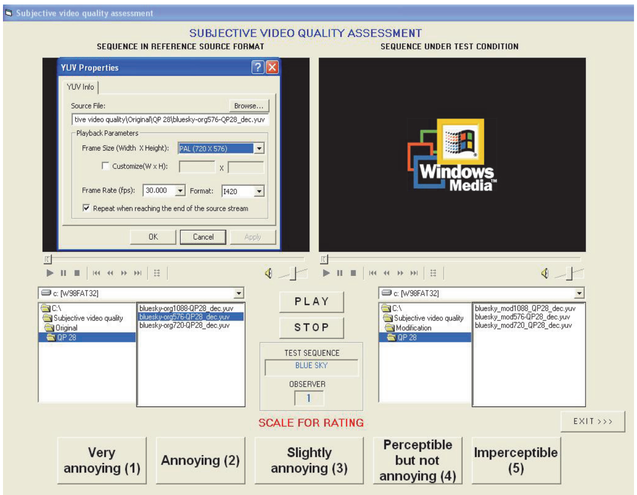


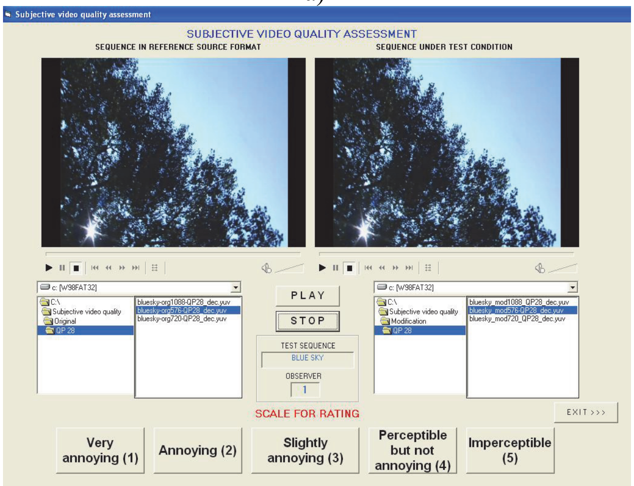
Fig. 3. An example: The first phase for *Blue sky* test sequence in an appropriate format [14].

For example, we show *Blue sky* and *Rush hour* video test sequences, because they have differences in camera motion. *Blue sky* represents cycle camera motions which show blue sky and top of trees, while *Rush hour* represents a still camera which shows traffic jam.

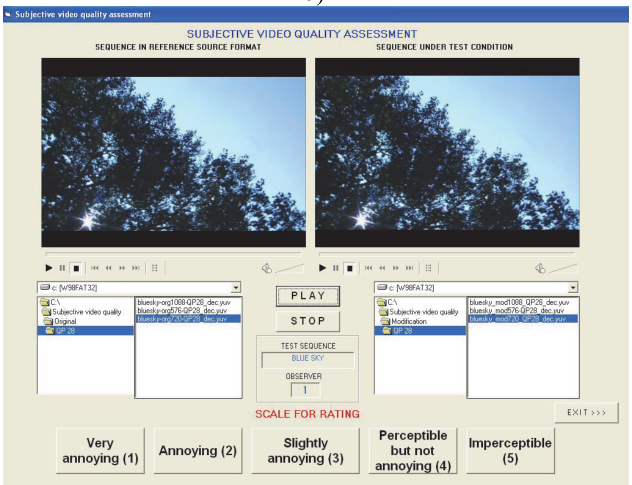
In the proposed algorithm for selective intra- and optimized inter-prediction, the obtained results for the subjective video quality evaluation between sequence in a reference source format and sequence under the test condition, show that there exists a small difference in the signal-to-noise ratio (SNR), i.e. the image quality. To prove this, there are mean estimation values given by the viewers, varying in the range from 4.00 to 4.95 for each video sequence in a different video resolution and with a different quantization parameter (QP).

Fig. 6 contains the results for the subjective evaluation of the tested video sequences quality in SD, HD and *full* HD resolution for QP=28. As it can be seen, the observer's ratings are mainly 4 or 5.

a)



b)



c)

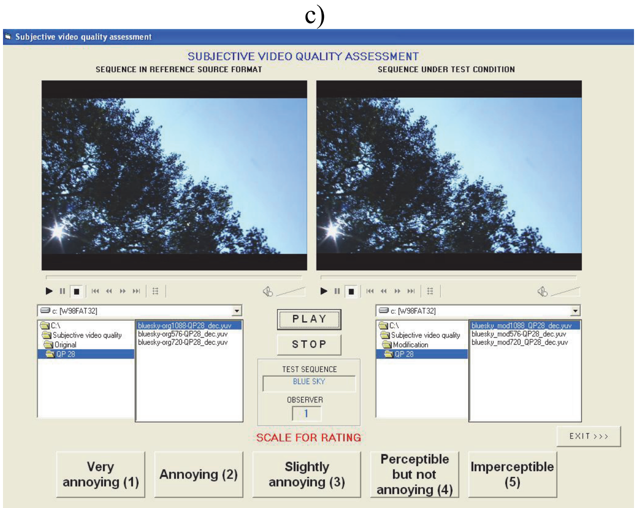
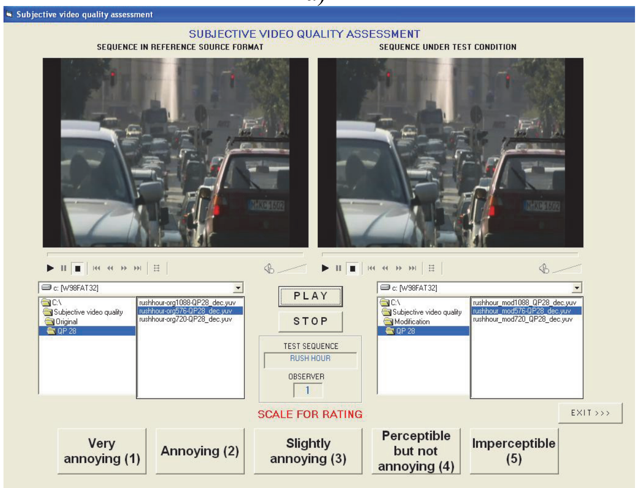
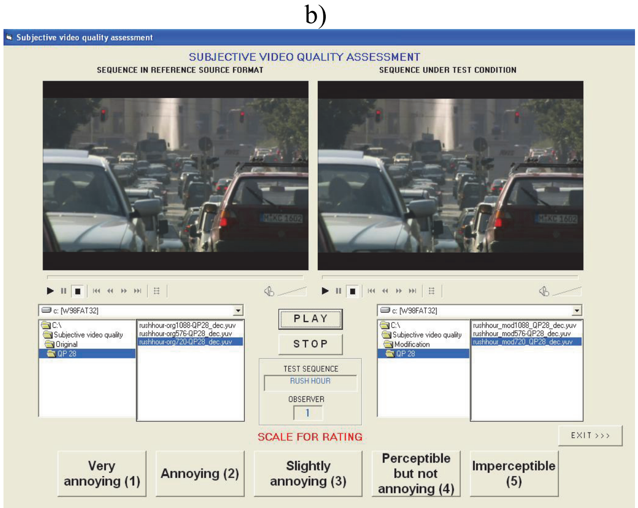


Fig. 4. An example: the second phase-processing and rating for *Blue sky* test sequence in (a) Standard, (b) High resolution and (c) full-High resolution [14].

a)



b)



c)

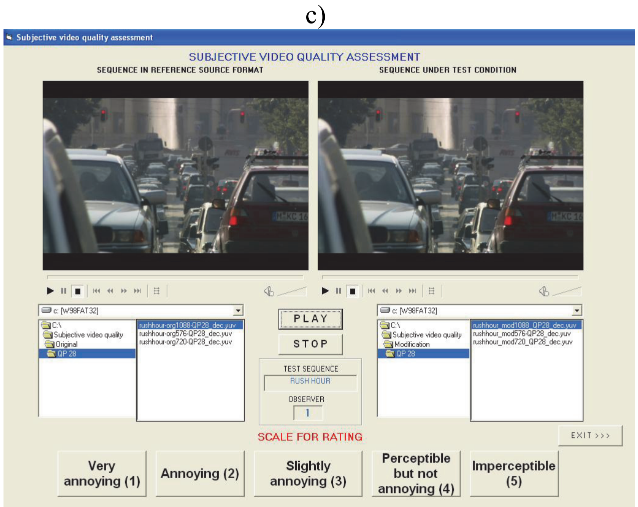


Fig. 5. An example: the second phase-processing and rating for *Rush hour* test sequence in (a) Standard, (b) High resolution and (c) full-High resolution [14], [15].

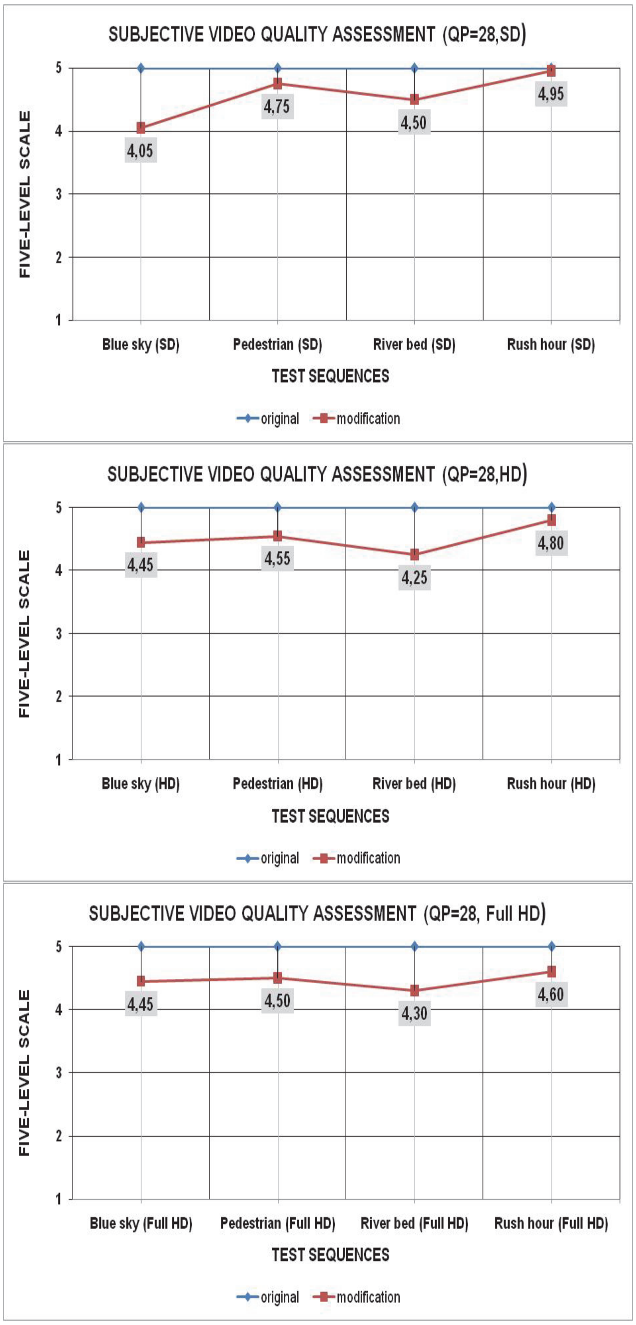


Fig. 6. Subjective video quality assessment results for different resolutions and QP=28 [14], [15].

V. Conclusion

This paper seeks to provide an approach for subjective video quality assessment in the H.264/AVC standard for proposed selective intra prediction and optimized inter prediction algorithm. For this purpose a special software packet for subjective video quality assessment is developed. It is developed in accordance with recommendation ITU-T P.910, which is suitable for testing of various multimedia applications.

User interface for the subjective assessment of quality of test sequences has been developed in accordance with the rules defined for the DCR test method.

Before the testing procedure was carried out, first it was necessary to fulfill all conditions relating to the observers, and then they have accessed the personal evaluation.

The testing included two phases. In the first phase a user interface was prepared. After that, in the next phase video test sequences were processed through the used interface and presented to the observers.

Finally, observers gave their subjective assessments of the quality of video content that was presented.

The obtained results show that in the proposed selective intra prediction and optimized inter prediction algorithm there is a small difference in picture quality (signal-to-noise ratio) between decoded original and modified video sequences.

References

[1] I. E. G. Richardson, *H.264 and MPEG-4 video compression video coding for next-generation multimedia*, John Viley & Sons Ltd., The Atrium, Southern Gate, Chichester, West Sussex PO 19 8SQ, England, 2003.

[2] H.264: International Telecommunication Union, Recommendation ITU-T H.264: Advanced Video Coding for Generic Audiovisual Services, ITU-T, 2003.

[3] T. Weigand, G. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Transactions on Circuits and System for Video Technology*, vol. 13, no. 7, pp. 560-576, July 2003.

[4] A. Puri, H. Chen and A. Luthra, "Video coding using the H.264/MPEG-4 AVC compression standard," *Signal Processing: Image Communication*, vol. 19, is. 9, pp. 793-846, October 2004.

[5] S. Kwon, A. Tamhankar, and K. R. Rao, "Overview of the H.264/MPEG-4 part 10," *Journal of Visual Communication and Image Representation*, vol. 17, is. 9, pp. 186-215, April 2006.

[6] K. R. Rao, Z. S. Bojkovic, and D. A. Milovanović, *WirelessMultimedia Communications: Convergence, DSP, QoS and Security*, CRC Press, Boca Raton, Florida, USA, 2009.

[7] D. Marple, T. Wiegand and G. J. Sullivan, "The H.264/MPEG4 Advanced Video Coding Standard and its Applications," *IEEE Communication Magazine*, vol. 44, is. 8, pp. 134-143, August 2006.

[8] J. Ostermann, J. Bormans, P. List, D. Marple, M. Narroschke, F. Pereira, T. Stockhammer, and T. Wedi, "Video coding with H.264/AVC: tools, performances, and complexity," *IEEE Circuits and System Magazine*, pp. 7-28, first quarter 2004.

[9] G. Sullivan, P. Topiwala, and A. Luthra, "Overview of the H.264/AVC Advanced Video Coding Standard: Overview and introduction to the Fidelity Range Extensions," *SPIE Conference on Applications of Digital Image Processing XXVII, Special Session on Advances in the New Emerging Standard: H.264/AVC*, pp. 1-22, August 2004.

[10] J.Reichel, H.Schwarz, and M. Wien (eds.), *Scalable video coding WD 1, Doc. JVT-N020*, Hong Kong, January 2005.

[11] H.C. Huang, W.H Peng, T. Chiang, and H-M. Hang, "Advances in the Scalable Amendment of H.264/AVC", *IEEE Communication Magazine*, vol. 45, is. 1, pp. 68-76, January 2007.

[12] Z. Milicevic and Z. Bojkovic, „H.264/AVC Standard: A Proposal for Selective Intra and Optimized Inter Prediction," *Journal of Network and Computer Applications,* volume 34, issue 2, pp. 686-691, March 2011.

[13] Recommendation ITU-T P.910: Subjective video quality assessment methods for multimedia applications, ITU-T, 2008.

[14] Z. Miličević, „New approch to video coding performance improvement using multimedia H.264/AVC standard," Ph.D. dissertation, Faculty of transport and traffic engineering, University of Belgrade, Jun 2011.

[15] Z. Miličević, Z. Bojković „Subjective video quality assessment in H.264/AVC video coding standard", *Proceeding of XIX Telecommunication forum TELFOR 2011*, pp. 1183-1186, Belgrade, Serbia, 22-24. November 2011.