

The Impact of Packet Loss on Quality of H.264/AVC Video Streaming

Sarunas Paulikas¹, Darius Gursnys¹, Aurimas Anskaitis¹, Arunas Saltis¹

¹*Department of Telecommunications Engineering, Vilnius Gediminas Technical University, Naugarduko St. 41–205, LT-03227 Vilnius, Lithuania
sarunas.paulikas@vgtu.lt*

Abstract— In this paper a simple and robust method for estimation of distorted video quality, which is perceived by human observer in mobile video streaming applications, is proposed and assessed. Increasing bandwidth of mobile communication systems expand the variety of offered multimedia services such as video streaming. However, the quality of these services is very dependent on rapidly varying mobile communication conditions. Most widely used video quality estimation methods, such as Peak Signal to Noise Ratio (PSNR), Structural Similarity (SSIM), and Video Quality Metric (VQM) are based on the presence of full or reduced reference video. These methods could be used to assess video quality of video transmission system only during test stage and in the limited number of scenarios. In order to assess user experienced video quality in real conditions, methods with no reference must be employed. Such existing methods as video quality metric use bit-error rate that has low correlation with human perceived video quality. More precise methods usually are too complex and require too much processing power that cannot be tolerated in handheld mobile devices. In this paper it is shown that developed no reference low complexity video quality estimation method based on H.264/AVC video stream packet structure delivers estimate of received video quality comparable with results of subjective MOS tests.

Index Terms—H.264 coding; mobile video streaming; video quality assessment; VQM.

I. INTRODUCTION

Today mobile telecommunication service providers face the demand to provide more data throughput while maintaining service quality [1], [2]. Thus, quality control of the provided service gets mandatory [3], [4]. Mobile video streaming is one of those fast-developing services, which quality is very noticeable by users and highly influences their satisfaction of service provider.

Video transmission through wireless media to mobile device is a demanding task. It requires high throughput over the wireless channel with time-varying parameters. Currently large number of scientific publications has been dedicated to problems of the end-to-end quality of video transmission through wireless networks.

The video quality transmitted to mobile device is influenced by two distinct types of distortion that result from the lossy compression introduced by the encoder (source distortion), and from the lossy wireless channel (loss distortion) [5].

There are number of measures to evaluate quality of video sequence. Most frequently used measures by engineers and researchers to evaluate the performances of digital video processing systems are based on peak signal to noise ratio (PSNR) [6], [7]. However, these measures have low correlation to human perceived video quality. Generally, to measure video quality in the respect to human perception, the standardised viewing test must be carried out as described in ITU-T P.910 recommendation “Subjective video quality assessment methods for multimedia applications”. As outcome of these test is video quality measured in Mean Opinion Score (MOS). But such tests require a lot of time and resources. To overcome that shortcoming the number of video quality measures that had good correlation to MOS results were designed [8], [9]. To mention a few more popular are Motion Picture Quality Metrics (MPQM), Video Quality Metrics (VQM), and Structural Similarity (SSIM). Nevertheless, these metrics are hardly applicable in real mobile video transmission scenarios. At first, to compute video quality these methods require to compare two video sequences: reference and received. So it is very difficult to make reference sequence available for user mobile device during real service deployment. At second, these methods are very complex and thus computationally extensive. Mobile devices usually have limited computational or/and electrical power resources.

There are proposed several reference free [10], [11] video quality evaluation methods, but they are not yet standardized and have own shortcomings.

So, there still is a need for an efficient video quality estimation method that have good correlation to the human perceived video quality, and at the same time, are simple-to-compute for implementing in mobile devices.

In this paper we will show the analysis of several video quality models that could be used to improve video quality estimation precision using method proposed in [12].

II. VIDEO QUALITY MEASURING METHOD

In [12], authors proposed the reference free streamed video quality estimation method applicable for video clips coded using the base line profile of H.264/AVC codec [13].

The H.264/AVC is based on the conventional, defined by to the MPEG standard, block-based motion-compensated video coding. The H.264/AVC standard has eleven profiles and sixteen levels. The profile specifies encoding algorithms and the level presents bit-rate constraints on parameter

values and thus restricts computational complexity. This article will focus on the H.264/AVC baseline profile that is designed for lower-cost applications with limited computing resources. Bit streams conforming to the baseline profile generally have the following main constraints: only I and P frame types may be present in the MPEG stream of group of picture (GOP) and bit rates must be in the range 64 kbps-768 kbps. The abbreviation I frame stands for so-called Intra-frame that can be decoded independently of any other frames. The P frame is an abbreviation for forward Predicted-frame. P frames improve compression by exploiting the temporal redundancy in a video. P frames store only the difference in image from the frame (either an I frame or P frame) immediately preceding it. The difference is calculated using motion vectors that are embedded in the P frame for use by the decoder. If a video drastically changes from one frame to the next, it is more efficient to encode it as an I frame.

The choice of a video codec to investigate was influenced by big amount of currently operating consumer mobile devices that have support for the H.264/AVC.

The main idea of the proposed method was to estimate a video quality of received video stream by using parameters extracted from data of compressed video frames thus avoiding complex and time expensive H.264 decoding. The method implementing algorithm monitors stream of the H.264 video frames, from the frame header extracts information about GOP structure, frame sequence, frame type (I or P frame) and calculates a number of bits used for storing of motion vectors (further we refer to it as motion vectors size). If the algorithm detects the corrupted frame or frames, determines its place and number in the particular GOP and makes the decision about the video quality score using the video quality model.

The video quality model was created after the analysis of the influence of lost frames type, its number and place in GOP and motion vector size to the quality of final video clip. As a reference for the video quality estimation of the final video clip is used the Video Quality Metrics (VQM) [14]. The VQM is a standardized reduced reference method for objectively measuring video quality. It predicts the subjective quality ratings that would be obtained from a panel of human viewers. Four U.S. patents owned by NTIA/ITS cover the technology used in VQM. VQM also showed very good performance in the International Video Quality Experts Group (VQEG) Phase II validation tests and it were adopted by the ANSI as a U.S. national standard (ANSI T1.801.03-2003), and as international ITU Recommendations (ITU-T J.144 and ITU-R BT.1683, both adopted in 2004).

Fig. 1 and Fig. 2 show the experimental dependences of a video quality measured by the VQM of three video clips on place of the lost P frame in the GOP and size of motion vectors in lost frames. As experimental video clips are used three progressive video sequences in the raw format YUV 4:2:0: foreman, hall-monitor and mobile. The sequences are selected so that could be subjectively classified as follows: foreman, classified as a high motion video (talking head, with pan to construction site, geometric shapes, shaking camera), hall-monitor, classified as a low motion video (an example of video supervision, stationary camera, two

moving objects) and mobile, classified as a moderate motion video (a lot of small moving objects). All video clips are coded with two most commonly used resolutions: QCIF and CIF (Quarter Common Intermediate Format, 176×144 pixels and Common Intermediate Format CIF, 352×288 pixels) with 25 fps and with total of 300 frames. Further, these sequences encoded at 15 fps and three different coding rates: 64 kbps, 128 kbps and 192 kbps.

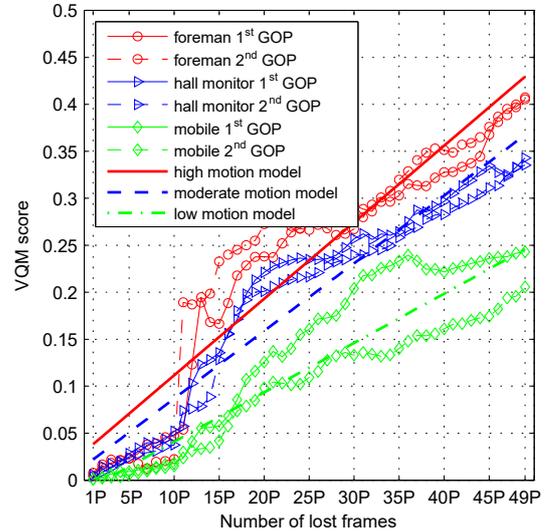


Fig. 1. Experimental dependence of the video quality on the place of lost P frame in the GOP.

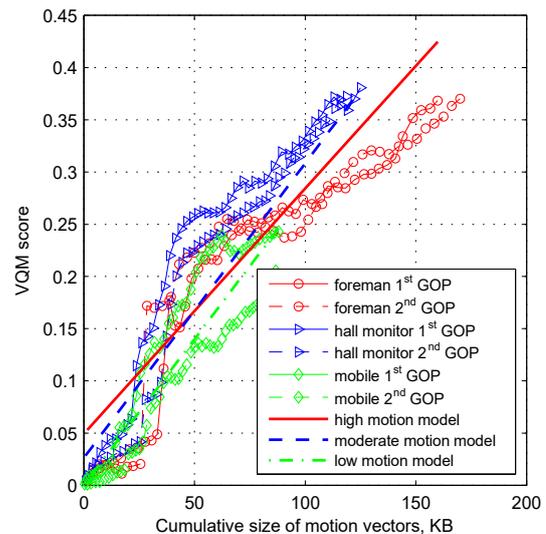


Fig. 2. Experimental dependence of the video quality on cumulative size of motion vectors in lost P frames.

After the analysis of the experimental results, for the video quality estimation is considered to use the linear model (thick lines in Fig. 1 and Fig. 2) that considers place of lost the P frame in the GOP

$$\hat{Q}(N, M) = a(M) \times N + b(M). \quad (1)$$

In (1), $\hat{Q}(N, M)$ represents the video quality estimate: N is a number of lost P frames in a particular GOP; $a(M)$ and $b(M)$ are constants, which values depend on the motion extent M that can be determined by dominating size of a motion vector in the given GOP. The values of constants a and b , were obtained by performing the least-square (LS) fit of particular curve chosen according to M from Fig. 1.

In the paper [12], the cumulative size of motion vectors was not included in the model (1) as it did not show the ability to significantly discriminate of the motion type of video clips.

Analysis of the experimental data indicates that the quality of a degraded video does not depend on a bit-rate of coded video stream and a resolution of video clip. It is very likely result, because of the VQM algorithm determines the quality of video clip by comparing two video clips (reference and degraded) of the same bit-rate and resolution. Such approach lets determine only the influence of impairments in transmission channel but not the effectiveness of the H.264/AVC coding at different bit-rates and resolutions.

III. VIDEO QUALITY ASSESSMENT MODELS

In order to increase the precision of in [12] proposed method the extended experiments on greater variety of a video material using the Video Quality Experts Group (VQEG) test sequences are carried out [15]. Nine video clips in the YUV format with 525 lines per frame and 60 Hz frame rate is chosen. The video sequence consisted of 10 frames (not used) followed by 8 seconds video and appended by 10 frames (not used). The ten frames of the unused video allow enough frames for a codec to stabilize. During experiments these frames are skipped.

Again, with all these video clips the following dependence are obtained: video quality on place of lost P frame in the GOP, video quality on cumulative size of motion vectors in lost P frames and distribution of motion vectors sizes in video clips (Fig. 3).

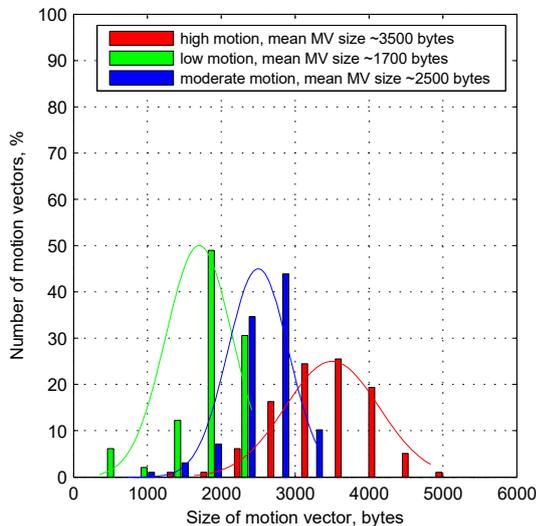


Fig. 3. Distribution of motion vectors sizes depending on the type of video clip.

Obtained experimental results let introduce the video quality model that relates the number of lost frames and the size of motion vector of lost frames to the quality of video clip

$$\hat{Q} = F(P, M). \quad (2)$$

In (2), \hat{Q} represents the estimated video quality: P is number of lost P frames and M is cumulative size of motion vectors of lost frames.

Different approaches of the approximation of the experimental data leads to several possibilities to build a mathematical model for assessing of a video quality:

- Model I – based only on a number and position in GOP of lost P frames (as illustrated in Fig. 1).
- Model II – based on a number and position of lost P frames in GOP and a cumulative size of lost motion vectors (as illustrated in Fig. 2).
- Model III – based on a number and position of lost P frames in the GOP and grouping video clips according their contents dynamics (as illustrated in Fig. 3).

Models I and III are the simplest one dimensional approach by the least-square error (LSE) approximation of the experimental data while minimizing

$$\mathbf{J} = [\hat{\mathbf{Q}}_{\text{VQM}} - \mathbf{Q}_{\text{VQM}}]^T [\hat{\mathbf{Q}}_{\text{VQM}} - \mathbf{Q}_{\text{VQM}}]. \quad (3)$$

In (3), \mathbf{J} represents the target function: $\hat{\mathbf{Q}}_{\text{VQM}}$ is a video quality estimate in VQM scores and \mathbf{Q}_{VQM} is a measured video quality estimate VQM scores using the VQM algorithm.

The data presented in Fig. 2 shows that loss of a P frame with bigger motion vector will degrade the quality more rapidly.

The improvement of the model could be expected by employing the weighed LSE approximation of the experimental data by minimizing

$$\mathbf{J}_W = [\hat{\mathbf{Q}}_{\text{VQM}} - \mathbf{Q}_{\text{VQM}}]^T \mathbf{W} [\hat{\mathbf{Q}}_{\text{VQM}} - \mathbf{Q}_{\text{VQM}}]. \quad (4)$$

In (4), \mathbf{J}_W represents the weighted target function: \mathbf{W} is a weights matrix that is composed from sizes of lost motion vectors.

For the approximation of the data (Fig. 2) a linear and quadratic polynomials are employed. The quality of a least squares fitting is determined by calculating determination coefficient R^2 and root mean square error:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{Q}_{\text{VQM}} - Q_{\text{VQM}})^2}{n}}, \quad (5)$$

$$R^2 = \frac{SSR}{SST}. \quad (6)$$

where SSR stands for the residual sum of squares and SST denotes the total sum of squares.

In the Model II is employed the two-dimensional LSE approximation based on a number and position of lost P frames and the cumulative size of lost motion vectors. Summary of tested models is presented in Table I. It shows that the Linear LSE model based only on a number of lost P frames performs quite well. However, the greater precision shows the Quadratic LSE and the Linear 2D LSE. The best results guarantee most the complex Quadratic 2D LSE model. Increasing the approximation order and incorporating knowledge about video dynamics the approximation precision increased up to approx. 10 %.

TABLE I. SUMMARY OF VIDEO QUALITY ESTIMATION MODELS BASED ON MSE APPROXIMATION OF EXPERIMENTAL DATA.

Model	Approximation method	Expression	R ²	RMSE
I	Linear LSE	$\hat{Q}_{VQM} = 0.007P + 0.014$	0.7985	0.0495
	Linear WLSE	$\hat{Q}_{VQM}^w = 0.006P + 0.055$	0.6617	0.1407
	Quadratic LSE	$\hat{Q}_{VQM} = 0.005P + 0.007M + 0.012$	0.8333	0.0451
	Quadratic WLSE	$\hat{Q}_{VQM} = -0.00008P^2 - 0.001M^2 + 0.0003PM + 0.021M + 0.006P - 0.037$	0.6950	0.1337
II	Linear 2D LSE	$\hat{Q}_{VQM} = 0.005P + 0.007M + 0.012$	0.8218	0.0466
	Quadratic 2D LSE	$\hat{Q}_{VQM} = -0.00008P^2 - 0.001M^2 + 0.0003PM + 0.021M + 0.006P - 0.037$	0.8628	0.0412
III	Linear LSE (low motion)	$\hat{Q}_{VQM} = 0.007P + 0.006$	0.7688	0.0516
	Linear WLSE (low motion)	$\hat{Q}_{VQM}^w = 0.006P + 0.017$	0.5848	0.1395
	Linear LSE (moderate motion)	$\hat{Q}_{VQM} = 0.007P + 0.013$	0.8823	0.0380
	Linear WLSE (moderate motion)	$\hat{Q}_{VQM}^w = 0.006P + 0.049$	0.8019	0.1025
	Linear LSE (high motion)	$\hat{Q}_{VQM} = 0.007P + 0.030$	0.8201	0.0461
	Linear WLSE (high motion)	$\hat{Q}_{VQM}^w = 0.006P + 0.075$	0.6965	0.1406

Note: Model I: Based only on number and position of lost P frames.

Model II: Based on number and position of lost P frames and cumulative size of lost motion vectors.

Model III: Based on number and position of lost P frames and grouping video clips according contents dynamics.

IV. TESTING OF MODELS

For tests of the proposed models for estimating the video quality, is used another video clip (*bowing*) that subjectively can be classified as a moderate motion and again is simulated the artificial loss of P frames.

The summary of video quality estimation results using all proposed methods in Table I are shown in Table II.

From Table II can be seen that all models performed quite well, determination coefficients are greater than 0.85. The best results were obtained using the Quadratic 2D LSE and the Linear LSE for the moderate motion video clips. However, the second one is suitable for predefined type in the sense of motion extent video clips and thus can't be used in more general case. Also, it can be stated that the quality of the degraded video under described conditions could be estimated within expected interval of 85 %, using quite low complexity, easy to compute models. These models are based only on a number and motion vector size of the lost P frames, thus not requires computationally complex and power consuming decoders.

TABLE II. PERFORMANCE OF VIDEO QUALITY ESTIMATION MODELS.

Model	Approximation method	R ²	RMSE
I	Linear LSE	0.8661	0.0430
	Linear WLSE	0.8510	0.0453
	Quadratic LSE	0.8994	0.0372
	Quadratic WLSE	0.9134	0.0345
II	Linear 2D LSE	0.8580	0.0442
	Quadratic 2D LSE	0.9167	0.0339
III	Linear LSE (moderate motion)	0.8997	0.0372
	Linear WLSE (moderate motion)	0.8788	0.0409

V. CONCLUSIONS

The described approach permits to determine the quality of a received video that was influence only by frame loss in

transmission channel and excluding performance of the H.264/AVC codec.

The dependence of the video quality on frame loss when a place of the lost P frame is close to the following I frame is almost linear and practically does not depend on video content. However, increasing distance of the lost P frame until the next successful I frame increases dispersion of individual quality estimates.

The video quality also depends on the content of the video clip. Less influence to the video quality under frame loss have video clips with a static background and small amount of moving objects.

The obtained experimental results indicate that it is possible to construct the method to predict the quality of a video clip of known content using only parameters that can be easily obtained from a coded video stream: size of motion vectors, place and type of lost frame.

Proposed low calculation complexity models let estimate quality of the video clip with a precision of 10 %–15 %, thus comparable with subjective MOS results of evaluation of video quality presented in the study [16], [17] that are in the range of approximately 10 % precision.

REFERENCES

- [1] S. Wang, W. Guo, C. Khirallah, D. Vukobratovic, J. Thompson, "Interference allocation scheduler for green multimedia delivery", *IEEE Trans. Vehicular Technology*, vol. 63, no. 5, pp. 2059–2070, 2014. [Online]. Available: <http://dx.doi.org/10.1109/TVT.2014.2312373>
- [2] Yang Liu, Zhipeng Yang, Ting Ning, Hongyi Wu, "Efficient quality-of-service (QoS) support in mobile opportunistic networks", *IEEE Trans. Vehicular Technology*, vol. 63, no. 9, pp. 4574–4584, 2014. [Online]. Available: <http://dx.doi.org/10.1109/TVT.2014.2311450>
- [3] Y. Baia, Y. Chub, M. R. Itoc, "Dynamic end-to-end QoS support for video over the Internet", *Int. J. Electron. Commun. (AEUE)*, vol. 65, no. 5, pp. 385–391, 2011. [Online]. Available: <http://dx.doi.org/10.1016/j.aeue.2010.07.002>
- [4] P. Rengaraju, Chung-Horng Lung, F. R. Yu, A. Srinivasan, "On QoS monitoring and E2E service assurance in 4G wireless networks", *IEEE Wireless Communications*, vol. 19, no. 4, pp. 89–96, 2012.

- [Online]. Available: <http://dx.doi.org/10.1109/MWC.2012.6272428>
- [5] F. D. Simonea, M. Naccarib, M. Tagliasacchic, F. Dufauxd, S. Tubaroc, T. Ebrahimia, "Subjective quality assessment of H.264/AVC video streaming with packet losses", *EURASIP Journal on Image and Video Processing*, vol. 2011, 2011.
- [6] Z. He, H. Xiong, "Transmission distortion analysis for real-time video encoding and streaming over wireless networks", *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 6, no. 9, pp. 1051–1062, 2006. [Online]. Available: <http://dx.doi.org/10.1109/TCSVT.2006.881198>
- [7] M. Vranjes, S. Rimac-Drlje, D. Zagar, "Subjective and objective quality evaluation of the H.264/AVC coded video", in *Proc. of 15th Int. Conf. on Systems, Signals and Image Processing*, 2008, vol. 1, pp. 287–290. [Online]. Available: <http://dx.doi.org/10.1109/iwssip.2008.4604423>
- [8] M. Ries, O. Nemethova, M. Rupp, "Video quality estimation for mobile H.264/AVC video streaming", *Journal of Communications*, vol. 3, no. 1, pp. 41–50, 2008. [Online]. Available: <http://dx.doi.org/10.4304/jcm.3.1.41-50>
- [9] S. Wolf, M. Pinson, "A no reference (NR) and reduced reference (RR) metric for detecting dropped video frames", in *Proc. Fourth Int. Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM-09)*, 2009, vol. 1, pp. 1–4.
- [10] A. Eden, "No-reference estimation of the coding PSNR for H.264-coded sequences", *IEEE Trans. Consumer Electronics*, vol. 53, no. 2, pp. 667–674, 2007. [Online]. Available: <http://dx.doi.org/10.1109/TCE.2007.381744>
- [11] N. Liao, Z. Chen, "A packet-layer video quality assessment model with spatiotemporal complexity estimation", *EURASIP Journal on Image and Video Processing*, vol. 1, no. 5, p. 13, 2011. [Online]. Available: <http://dx.doi.org/10.1186/1687-5281-2011-5>
- [12] S. Paulikas, "Estimation of video quality of H.264/AVC video streaming", in *Proc. of EUROCON, Zagreb, Croatia, 2013*, pp. 694–700. [Online]. Available: <http://dx.doi.org/10.1109/eurocon.2013.6625056>
- [13] *ITU-T Recommendation H.264, Advanced video coding for generic audiovisual services*. Series H: Audiovisual and Multimedia Systems: Infrastructure of audiovisual services Coding of moving video. 2014.
- [14] S. Wolf, "Video Quality Measurement techniques", NTIA, 2009.
- [15] Video test sequences. [Online] Available: <https://media.xiph.org/vqeg/TestSequences/>
- [16] E. P. Ong, W. Lin, S. Z. L. Yao, M. H. Loke, "Perceptual quality metric for H.264 low bit rate videos", in *Proc. of Int. Conf. Multimedia and Expo, 2006*, vol. 1, pp. 677–680. [Online]. Available: <http://dx.doi.org/10.1109/icme.2006.262536>
- [17] J. Nightingale, Qi Wang, C. Grecos, S. Goma, "The impact of network impairment on quality of experience (QoE) in H.265/HEVC video streaming", *IEEE Trans. on Consumer Electronics*, vol. 60, no. 2, pp. 242–250, 2014. [Online]. Available: <http://dx.doi.org/10.1109/TCE.2014.6852000>