Low-Complexity No-Reference PSNR Estimation for H.264/AVC Encoded Video

Damien Schroeder, Ali El Essaili, Eckehard Steinbach
Institute for Media Technology
Technische Universität München
Munich, Germany

Dirk Staehle, Mohammed Shehada
DOCOMO Communications Laboratories Europe GmbH
Munich, Germany

***Abstract*—We present a no-reference (NR) PSNR estimation method which is based on only two bitstream features (average bitrate and mean quantization parameter of the I-frames). The low computational complexity of the proposed method makes it suitable for in-network real-time applications. The NR metric achieves a Pearson correlation of 0.99 for individual videos and a RMSE of approximately 1 dB PSNR on average. We additionally investigate the effect of various encoding configurations on the PSNR and show the robustness of our method towards these. Finally, we incorporate the proposed metric into an example application and demonstrate that only a minor performance loss is observed compared to the reference scheme which assumes the availability of true PSNR information.**

I. Introduction

Nowadays, video applications are responsible for more than half of the internet data traffic [1], and this proportion is expected to further increase, especially in mobile networks as a result of the growing number of smartphones and tablets. Due to this large amount of video traffic and the highly dynamic video characteristics, internet service providers and mobile network operators need to plan their networks carefully in order to keep up with the increasing traffic and provide the best possible experience to the users. To this end, in-network video quality monitoring is required, which needs to be of particularly low-complexity in order to be scalable to a large number of concurrent video streams and to be applicable in real-time.

Subjective quality evaluation relies on the ratings of a pool of human viewers and is the reference for video quality assessment but is not feasible in the context of real-time in-network quality monitoring due to cost and complexity issues. On the other hand, objective quality assessment is designed to estimate the video quality without the involvement of human observers. Depending on the amount of information about the original unencoded video available, objective metrics are classified into full-reference, reduced-reference and no-reference (NR) metrics. As no information about the original video is generally available during in-network monitoring, the objective quality assessment has to rely on a NR metric.

The large majority of the video traffic is now HTTP/TCP based [1], due to the pragmatic shift over the last decade from RTP/UDP based video streaming to HTTP/TCP based video delivery. Since TCP ensures reliable delivery of the video packets, no additional distortion is introduced in the network and the video distortion at the receiver is the same as at the encoder/sender. To quantify the encoding distortion, the Peak Signal-to-Noise Ratio (PSNR) is still widely used for its simplicity, although it is known to be not perfectly correlated with the human visual system [2], due to saturation effects in very low and very high PSNR regions. Still, the PSNR constitutes the basis of some state-of-the-art perceptual video quality metrics (e.g. $VQMTQ$ in [3] or $STVQM$ in [4]).

The goal of this work is to design a low-complexity NR PSNR estimation method which can be used for real-time in-network assessment of H.264/AVC encoded video streaming. NR metrics either operate in the pixel-domain or on the bitstream of the video [5]. Pixel-domain approaches have the drawback that the video needs to be fully decoded, which implicates a high computational complexity and are thus not considered here. In the case of bitstream-based metrics, almost all of the existing No-Reference PSNR estimation methods rely on the statistical properties of the transform coefficients ([6], [7], [8] and [9]), where the distortion is estimated by comparing the quantized coefficients to the estimated distribution of the original transform coefficients. Recently, [10] extended this method by taking into account de-blocking filtering. Computing the distribution of the transform coefficients is, however, computationally expensive, which limits the applicability and scalability of such methods for real-time quality monitoring.

Bitstream information has also been considered in the closely related field of NR perceptual video metrics. Recently, the ITU has standardized NR methods [11] that rely on the bitstream as well as on the packet headers of video streams to assess the quality. These complex methods are intended for RTP/UDP-based streaming and take into account effects due to packet losses at the client. In [12], the authors extract 64 different H.264/AVC bitstream features and use a partial least squares regression to fit a model to their subjective data. The high number of features leads to a complex model and to a lack of understanding of the effect of the different features on the video quality. The authors in [13] use 7 bitstream features, among them Quantization Parameter (QP) and bitrate, similar to the work in this paper. But they model the relationship between the video quality and the bitrate as linear, which is not the case for encoded videos.

In this paper, we propose a novel bitstream-based no-reference PSNR estimation method for H.264/AVC encoded videos, which requires the extraction of only two bitstream parameters: the average bitrate and the mean QP of the I-frames of the video. Using only two bitstream parameters makes this NR metric highly relevant for real-time applications. Additionally, we investigate the effect of various encoding

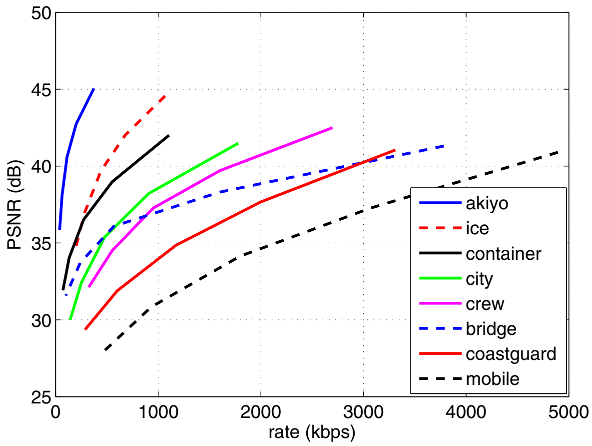


Fig. 1: PSNR versus bitrate for different videos.

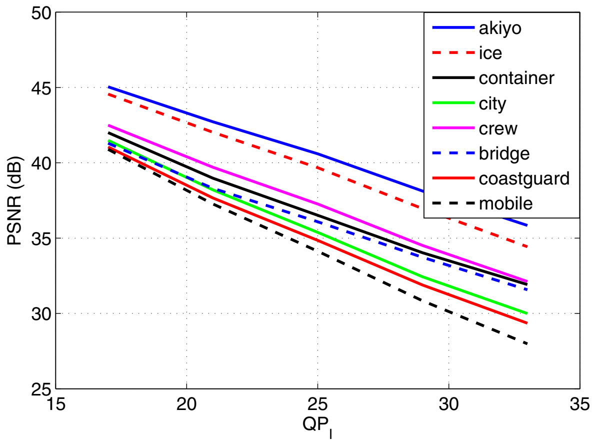


Fig. 2: PSNR as a function of the QP_I for different videos.

configurations on the PSNR and assess the robustness of our proposed NR metric towards these configurations. We further present an example application of mobile network resource allocation for HTTP video streaming users where our proposed method can be applied.

The rest of this paper is organized as follows. The proposed NR metric is presented in Section II. Its robustness towards videos not used for training and various H.264/AVC encoding configurations is discussed in Section III. Section IV discusses potential applications and Section V concludes this work.

II. Proposed No-Reference Metric

Our proposed NR metric estimates the average PSNR of the luminance component of the video signal at a given resolution. The estimation relies on only 2 input parameters: 1) the average bitrate of the video stream and 2) the mean QP of the I-frames of this video.

The bitrate alone is not sufficient for the PSNR estimation as the PSNR at a certain bitrate is strongly content dependent [14], as can be seen in Figure 1, which shows the PSNR as a function of the bitrate for 8 different videos encoded with H.264/AVC baseline in CIF format (352x288) and at 30 frames per second. The videos such as *akiyo* and *ice* that achieve a high PSNR at a low bitrate are considered low-demanding in opposition to high-demanding videos such as *coastguard* and *mobile* that require a high bitrate to obtain a high PSNR.

During the video encoding process, the quantization of the transform coefficients is the main mechanism for introducing distortion in the video. The coarseness of the quantization is determined by the QP. Therefore, the QP is directly related to the coding distortion. In H.264/AVC, each macroblock can be encoded with its own QP. Without having to decode the video frames, the bitstream can be parsed to extract the QP of each macroblock. In our PSNR estimation, we use the mean QP over all macroblocks of the I-frames as input parameter and denote it by QP_I in the rest of the paper. For video streaming, I-frames are inserted in regular short intervals in order to provide independently decodable entry points to the viewer. Using only the I-frames reduces the parsing time significantly. Figure 2 represents the PSNR as a function of the QP_I for 8 different videos. Again, QP_I alone is not sufficient for PSNR estimation as the PSNR-QP_I relationship is content dependent. We can see that low-demanding videos (e.g. *akiyo* and *ice*) tend to have a high PSNR for a given QP_I, compared to high-demanding videos (e.g. *coastguard* and *mobile*) which have a low PSNR for a given QP_I.

As there is a correlation between low-demanding (i.e. lower bitrate for a particular PSNR than other videos) and higher PSNR for a given QP_I than other videos, using jointly the bitrate and the QP_I of a video enables to estimate its PSNR. The relation between the bitrate and PSNR can be approximated with a logarithmic function (Figure 1) and the relation between the QP_I and the PSNR is approximately linear (Figure 2). Based on these two relations, we estimate the PSNR as a function of the bitrate and the QP_I as follows:

$$PSNR = b_1 + b_2 \cdot \log(rate) + b_3 \cdot QP_I + b_4 \cdot rate \cdot QP_I \quad (1)$$

where we have a constant term, a term depending on the bitrate (expressed in kbps), a term depending on the QP_I and an interaction term.

The model parameters b_1 to b_4 are trained using representative training sequences by means of linear regression. Our training data consists of 8 video sequences [15] (*akiyo, bridge, city, coastguard, container, crew, ice, mobile*) encoded with the x264 software [16] in H.264/AVC baseline profile at 30 fps and CIF resolution. The videos are encoded in constant QP mode. Five representations are encoded with QP taking the values 17, 21, 25, 28 and 33, respectively.

Figure 3 shows the 40 training points as blue dots and the resulting PSNR model. The values for the model parameters b_1 to b_4 are presented in Table I.

TABLE I: Trained model parameters

Model Parameter	b_1	b_2	b_3	b_4
Value	74.791	-2.215	-0.975	$1.708 \cdot 10^{-5}$

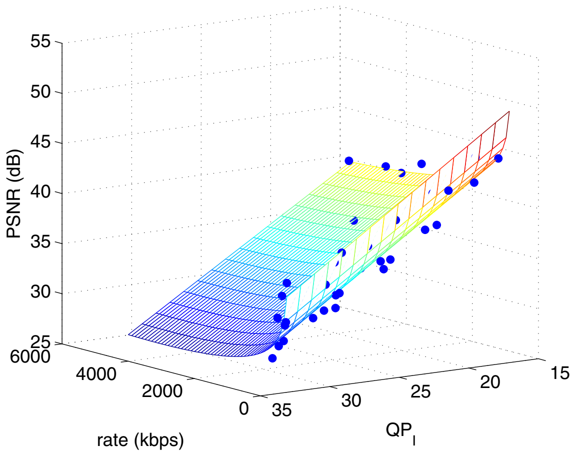


Fig. 3: Training data and fitted PSNR function.

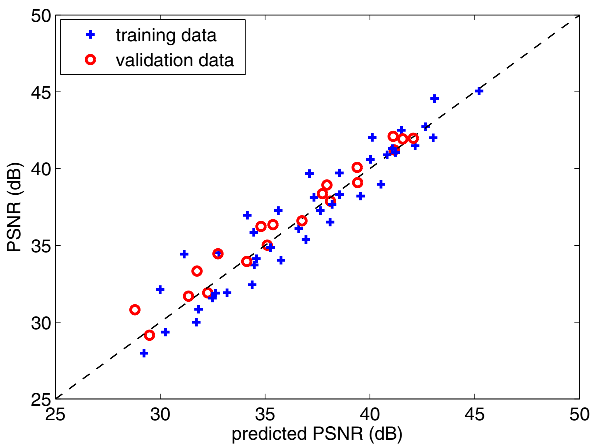


Fig. 4: Predicted vs. true PSNR.

III. Assessment of the Proposed Metric

A. Validation

To verify the robustness of the proposed model towards videos that don't belong to the training set, we use 4 additional validation video sequences *(bus, football, foreman, highway)*, encoded with the same encoding parameters as the training data. Figure 4 shows the relation between the predicted PSNR and the true PSNR for the training data (blue crosses) and the validation data (red circles). The points of the validation data lie within the cloud of the points of the training data, which indicates a good robustness.

The estimated PSNR points as a function of the rate are compared to the true PSNR points for the four validation videos in Figure 5. We also compute the Pearson correlation and the root mean square error (RMSE) between the true PSNR values and the predicted PSNR for the four validation sequences individually. Results are summarized in Table II. We see that the Pearson correlation is almost 1 for each individual video which means that the proposed PSNR estimation is performing well at estimating the shape of the rate-PSNR curve.

We also calculate the Pearson correlation and the RMSE for the four validation videos together and for the entire data set (8 training sequences plus 4 validation sequences). Results in Table III show that the overall Pearson correlation is still very high, while the overall RMSE is approximately 1 dB when all 12 videos are taken into account.

TABLE II: Metric performance for validation videos

	bus	football	foreman	highway
RMSE [dB]	0.246	1.447	0.244	0.981
Pearson correlation	0.999	0.999	0.999	0.998

TABLE III: Metric performance for video sets

	4 validation videos	all 12 videos
RMSE [dB]	0.891	1.043
Pearson correlation	0.983	0.968

B. Influence of the encoder configuration

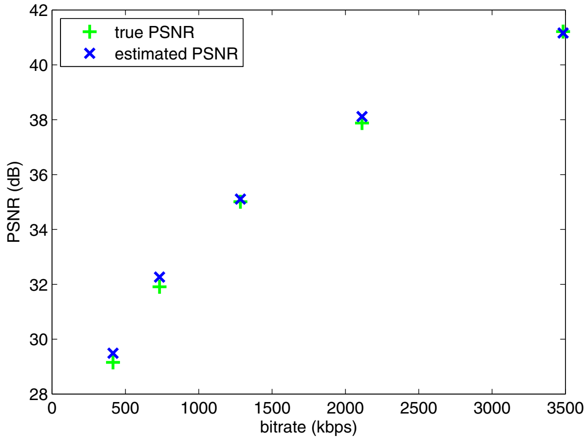
So far, we have only considered a specific H.264/AVC encoding format. We want to assess how sensitive our estimation method is towards different encoding methods. x264 implements three different encoding modes:

- **constant QP**: the QP is the same for each macroblock of a given frame type.
- **rate control**: the video is encoded at a target bitrate.
- **crf**: crf is an x264 specific encoding mode. The crf parameter can take on any value between 0 and 51, similar to the QP. In crf mode, the QP is adapted depending on motion characteristics.

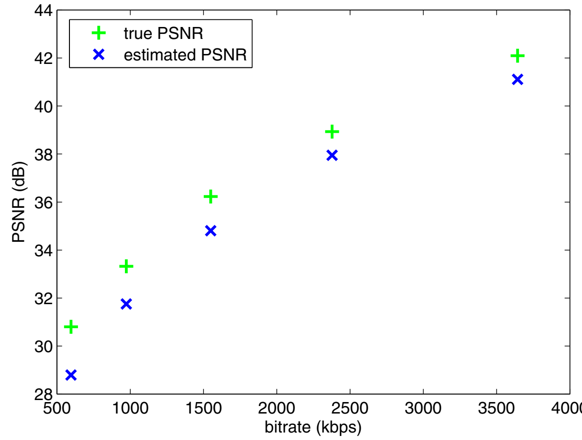
We encode 8 different CIF videos (*akiyo, bus, city, coastguard, container, crew, foreman, highway*) at 30 fps in baseline profile. The videos are encoded with each of the three encoding modes in five different representations. In constant QP mode, the QP takes the values 17, 21, 25, 28 and 33. In rate control mode, the videos are encoded at 500 kbps, 1 Mbps, 1.5 Mbps, 2 Mbps and 2.5 Mbps. In crf mode, the crf parameter takes the value 19, 21, 23, 25 and 27. As a result, we have $8 \cdot 3 \cdot 5 = 120$ encoded videos from which we extract the mean QP of the I-frames, the bitrate and the overall PSNR.

We perform a three-way analysis of variance (ANOVA) for testing the effect of the QP_I, the bitrate and the encoding mode on the PSNR. The ANOVA results are reported in Table IV, which indicate that both QP_I and bitrate have significant impact ($p < 0.05$) on the PSNR. The interaction between QP_I and bitrate is also found to be significant ($p < 0.05$). On the other hand, the encoding mode doesn't have a significant impact on the PSNR ($p > 0.05$). This suggests that the encoding mode doesn't need to be additionally taken into account during PSNR estimation and thus that our PSNR estimation method is robust towards different encoding modes.

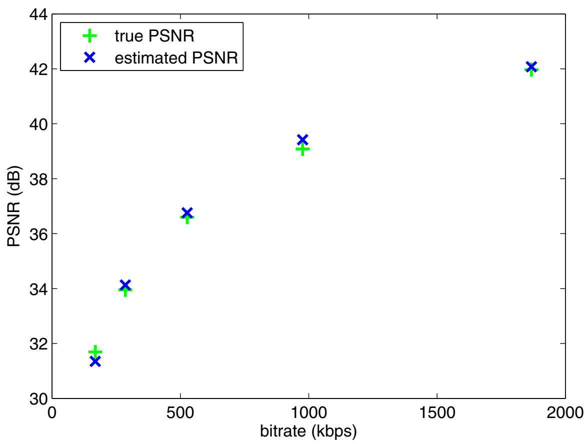
To confirm this, we use the model trained with constant QP encoded videos in Section II (Table I) in order to estimate the PSNR of the 8 videos encoded either with rate control or



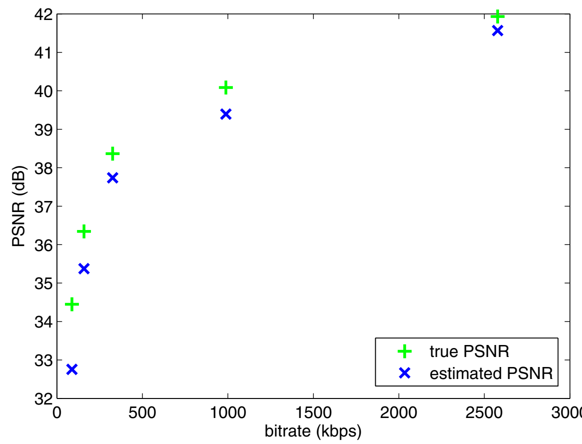
(a) bus



(b) football



(c) foreman



(d) highway

Fig. 5: True and estimated PSNR for 4 validation videos.

TABLE IV: Three-way ANOVA with encoding mode

Source	Mean Sq.	df	F-value	p-value
QP_I	378.4	1	321.51	<0.0001
bitrate	8.06	1	6.85	0.0101
mode	0.69	2	0.59	0.5569
QP_I : bitrate	18.43	1	15.65	0.0001
Error	1.177			

with crf mode. The resulting RMSE and Pearson correlation are presented in Table V. The Pearson correlation is still very high ($\geq$ 0.95) and the RMSE is still close to 1 dB, which confirms that our NR PSNR estimation method is robust against different encoding modes.

C. Influence of the encoding profile

In the same way, we investigate the influence of the H.264/AVC profile on our PSNR estimation method. Therefore, we compare three different encoding profiles: **baseline, main** and **high**. We encode 12 different CIF videos (*akiyo, bridge, bus, city, coastguard, container, crew, football, foreman, highway, ice, mobile*) at 30 fps in constant QP mode. The videos are encoded with each profile in five different representations (QP takes the values 17, 21, 25, 28 and 33). We have thus $12 \cdot 3 \cdot 5 = 180$ encoded videos, from which we extract the QP_I, the bitrate and the PSNR.

TABLE V: Metric performance for various encoding modes

	rate control	crf
RMSE [dB]	1.078	1.147
Pearson correlation	0.979	0.950

We test the effect of the QP_I, the bitrate and the profile on the PSNR by performing a three-way ANOVA on our data, as presented in Table VI. The QP_I, the bitrate and the interaction of QP_I with bitrate all have significant impact on the PSNR ($p < 0.05$). On the other hand, the impact of the profile on the PSNR is not significant ($p > 0.05$). Again, this suggests that the encoding profile doesn't need to be additionally taken into account during PSNR estimation and thus that our PSNR estimation method is robust towards different encoding profiles.

TABLE VI: Three-way ANOVA with encoding profile

Source	Mean Sq.	df	F-value	p-value
QP_I	1240.19	1	542.2	<0.0001
bitrate	61.87	1	26.65	<0.0001
profile	0.97	2	0.49	0.8116
QP_I : bitrate	118.04	1	50.85	<0.0001
Error	2.32			

We use the model with the parameters from Table I to estimate the PSNR of the 12 videos encoded with the main and the high profile. Table VII contains the resulting RMSE, which is approximately 1.16 dB in both cases and the Pearson correlation, which is above 0.96 in both cases. This confirms that the proposed estimation method is robust against various encoding profiles.

TABLE VII: Metric performance for various H.264 profiles

	main	high
RMSE [dB]	1.163	1.168
Pearson correlation	0.964	0.963

D. Influence of the encoder

We also examine the effect of using a different encoder software. Therefore, we additionally encode 12 CIF videos (*akiyo, bridge, bus, city, coastguard, container, crew, football, foreman, highway, ice, mobile*) at 30 fps in baseline profile with the JM reference software [17]. The videos are encoded in constant QP mode with QP values 17, 21, 25, 28 and 33. We use this data together with the videos encoded with x264 and same parameters (baseline profile, constant QP).

We perform a three-way ANOVA for testing the effect of the QP_I, the bitrate and the encoder software on the PSNR. The results are reported in Table VIII and confirm that the QP_I, the bitrate and the interaction of QP_I and bitrate have a significant impact on the PSNR ($p < 0.05$). On the contrary, the encoder doesn't have a significant impact on the PSNR ($p > 0.05$).

TABLE VIII: Three-way ANOVA (encoder)

Source	Mean Sq.	df	F-value	p-value
QP_I	910.16	1	390.62	<0.0001
bitrate	43.33	1	18.6	<0.0001
encoder	3.18	1	1.36	0.2455
QP_I : bitrate	74.44	1	31.95	<0.0001
Error	2.33			

The model with the parameters from Table I is used to predict the PSNR of the videos encoded with the JM reference software. It results in a Pearson correlation of 0.972 and a RMSE of 1.203 dB.

In this section, we showed that our proposed NR PSNR estimation method is not only valid for a particular H.264/AVC encoding structure but on the contrary robust against various encoding softwares, modes and profiles. Although the model parameters have been trained for a single configuration (x264 encoder, baseline profile, constant QP), the metric always provides a Pearson correlation above 0.95 and a RMSE between 1 dB and 1.2 dB on average. For best performance, careful selection of the training data should be considered.

IV. Applications

The average bitrate of an encoded video is easy to extract. The QP values of the macroblocks can be extracted from the bitstream after entropy decoding. Then, as the model parameters b_1 to b_4 have been trained beforehand, the calculation of the estimated PSNR with Equation (1) is straightforward. That is, the QP extraction is the only non-trivial step in our proposed PSNR estimation. Limiting the parsing to the I-frames of the video reduces the overall complexity. This low complexity makes the proposed NR metric applicable in real-time and scalable to multiple concurrent video streams, e.g. somewhere in the network to monitor the PSNR of several video streams.

The emerging adaptive HTTP streaming technologies, e.g. recently standardized as Dynamic Adaptive Streaming over HTTP [18], are based on the concept of making the video content available at different bitrates, that is, at different qualities. Estimating the PSNR of a video at different bitrates enables to then reconstruct its rate-PSNR curve, either by having a set of bitrate-PSNR points (e.g. Figure 5) or by using an interpolation model, e.g. [19]. Our proposed PSNR estimation method provides Pearson correlation above 0.99 when considering individual videos (cf. Table II). This means that the shape of the estimated curve will be almost the same as the true curve.

Rate-distortion curves or utility curves can be used for quality-based resource allocation, as e.g. demonstrated in [20]. In a resource constrained network, the resources are distributed to the different users depending on how much utility they gain or they lose, respectively. To identify this gain or loss in utility correctly, the accurate steepness of the curves is of particular importance. This is provided by our NR metric as it presents a very high Pearson correlation.

A. Example application

To illustrate the effectiveness of the proposed NR metric, we present an example application based on our prior work in [20]. An LTE mobile network with multiple adaptive HTTP video streaming users in one cell is considered. The radio resources are allocated to K users in a way to maximize the overall satisfaction:

$$\underset{(\alpha_1,\ldots,\alpha_K)}{\arg\max} \quad \sum_{k=1}^{K} U_k(\alpha_k) \tag{2}$$

$$subject\ to \quad \sum_{k=1}^{K} \alpha_k = 1 \tag{3}$$

where (2) determines the resource share α_k of each user k that maximizes the sum of utilities and (3) constrains on the available resources. The utility on a Mean Opinion Scale

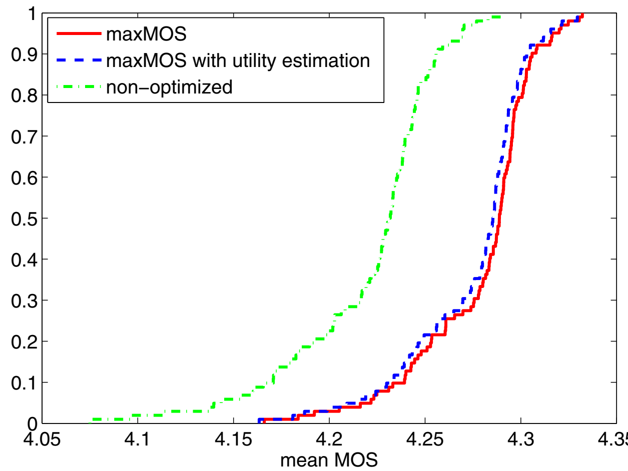


Fig. 6: cdf of mean MOS.

(MOS) [21] scale is calculated based on the PSNR with the VQM_P metric from [22]. The values on a scale from 0 to 100 are converted to the MOS scale from 1 to 4.5 using Annex B from [23].

An LTE cell with 5 MHz bandwidth and 10 mobile streaming users is simulated. Figure 6 shows the distribution of the mean MOS over all users in the cell for 100 simulations, where the users are streaming different videos and the channel conditions are changing from one simulation to another. The green curve on the left represents the mean MOS over the users when no resource optimization is carried out, that is, the resources are shared equally between the users. The red curve on the right represents the optimal mean MOS over all users, when the optimization is done with the true PSNR values. Finally, the blue curve shows the mean MOS over all users when the optimization is done based on the estimated PSNR curves. The PSNR values are estimated with the proposed NR metric. We see that the optimization result with the estimated values is very close to the optimization result with the true values. In fact, there is an average loss of less than 1% on the MOS scale compared to the optimization with the true PSNR values.

V. Conclusion

In this paper, we propose a novel NR metric which estimates the PSNR of a H.264/AVC video based on only the average bitrate of the video and the mean QP of the I-frames. The low-complexity of the method makes it relevant to real-time applications. The NR metric provides a Pearson correlation of 0.99 for individual videos and a RMSE of approximately 1 dB on average. We show that our PSNR estimation is robust against various H.264/AVC encoding configurations. We also provide an example application where our proposed NR metric is used to estimate rate-PSNR curves as a basis of a quality based resource allocation. In future work, we plan to integrate different video resolutions and framerates in the NR metric as well as investigate the pertinence of the proposed PSNR estimation method for HEVC encoded video.

References

[1] Sandvine, "Global internet phenomena report," 2013.

[2] B. Girod, "Digital images and human vision," A. B. Watson, Ed. Cambridge, MA, USA: MIT Press, 1993, ch. What's wrong with mean-squared error?, pp. 207–220.

[3] Y. Ou, Z. Ma, T. Liu, and Y. Wang, "Perceptual Quality Assessment of Video Considering Both Frame Rate and Quantization Artifacts," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 3, pp. 286–298, 2011.

[4] Y. Peng and E. Steinbach, "A novel full-reference video quality metric and its application to wireless video transmission," in *IEEE International Conference on Image Processing (ICIP)*, Brussels, Belgium, Sep. 2011.

[5] S. Winkler and P. Mohandas, "The Evolution of Video Quality Measurement: From PSNR to Hybrid Metrics," *IEEE Transactions on Broadcasting*, vol. 54, no. 3, 2008.

[6] D. S. Turaga, Y. Chen, and J. Caviedes, "No reference PSNR estimation for compressed pictures," *Signal Processing: Image Communication*, vol. 19, no. 2, pp. 173–184, Feb. 2004.

[7] A. Ichigaya, M. Kurozumi, N. Hara, Y. Nishida, and E. Nakasu, "A Method of Estimating Coding PSNR Using Quantized DCT Coefficients," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 2, pp. 251–259, Feb. 2006.

[8] A. Eden, "No-Reference Estimation of the Coding PSNR for H.264-Coded Sequences," *IEEE Transactions on Consumer Electronics*, vol. 53, no. 2, pp. 667–674, May 2007.

[9] T. Brandao and M. Queluz, "No-reference quality assessment of H.264/AVC encoded video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, no. 11, Nov. 2010.

[10] T. Na and M. Kim, "A Novel No-Reference PSNR Estimation Method with regard to De-blocking Filtering Effect in H.264/AVC Bitstreams," *IEEE Transactions on Circuits and Systems for Video Technology*, 2013.

[11] ITU-T, "Rec. P.1202 Parametric non-intrusive bitstream assessment of video media streaming quality," Oct. 2012.

[12] C. Keimel, M. Klimpke, J. Habigt, and K. Diepold, "No-reference video quality metric for HDTV based on H.264/AVC bitstream features," in *IEEE International Conference on Image Processing (ICIP)*, Brussels, Belgium, Sep. 2011.

[13] S.-O. Lee, K.-S. Jung, and D.-G. Sim, "Real-time Objective Quality Assessment based on Coding Parameters Extracted from H.264/AVC Bitstream," *IEEE Transactions on Consumer Electronics*, vol. 56, no. 2, pp. 1071–1078, May 2010.

[14] F. Yang and S. Wan, "Bistream-based quality assessment for networked video: A review," *IEEE Commun. Mag.*, Nov. 2012.

[15] "YUV Video Sequences." [Online]. Available: http://trace.eas.asu.edu/yuv/

[16] VideoLAN, 2013. [Online]. Available: http://www.videolan.org/developers/x264.html

[17] Fraunhofer HHI, "JM reference software 18.2," 2012. [Online]. Available: http://iphome.hhi.de/suehring/tml/

[18] ISO/IEC 23009-1, "Information technology - Dynamic adaptive streaming over HTTP (DASH) - Part 1: Media Presentation description and segment formats," Tech. Rep., Apr. 2012.

[19] L. Choi, M. Ivrlac, E. Steinbach, and J. Nossek, "Sequence-level methods for distortion-rate behavior of compressed video," *Proc. IEEE ICIP '05*, Sep. 2005.

[20] A. El Essaili, D. Schroeder, D. Staehle, M. Shehada, W. Kellerer, and E. Steinbach, "Quality-of-experience driven adaptive HTTP media delivery," in *IEEE International Conference on Communications (ICC 2013)*, Budapest, Hungary, Jun. 2013.

[21] ITU-T, "Rec. P.800 Methods for subjective determination of transmission quality," Aug. 1996.

[22] S. Wolf and M. Pinson, "Video quality measurement techniques," NTIA, Tech. Rep. TR-02-392, Jun. 2002.

[23] ITU-T, "Rec. G.107 The E-model: a computational model for use in transmission planning," Dec. 2011.