

Assessment of H.264 Coded Panorama Sequences

Michal Ries, Olivia Nemethova, Biljana Badic, Markus Rupp
Institute for Communications and Radio Frequency Engineering
Vienna University of Technology
Gusshausstr.25, A-1040 Vienna, Austria
(mries, onemeth, bbadic, mrupp)@nt.tuwien.ac.at

Abstract—The newest video coding standard H.264 allows providing video streaming for low bit and frame rates in acceptable quality in 3G wireless networks due to significant video compression gain while preserving the perceptual quality. This is specially suitable for video applications in 3G wireless networks. One of the most popular content types in 3G streaming is panorama (weather cams, traffic jams, city guides). The results for subjective perceptual quality evaluation of panorama sequences differ from other sequence types significantly. In this paper the difference of the panorama type is investigated and a prediction of a perceptual metric for panorama sequences based on different objective parameters is introduced.

I. INTRODUCTION

In the last decade many metrics have been proposed for video quality measurements. At the moment, four basic types of the video quality metrics are available: human perception video metrics [1], [2], objective degradation metrics as peak-to-signal to noise ratio (PSNR) [3], metrics based on watermarking [4], and combined metrics which take into account more than one objective video parameter [5], [6]. PSNR metrics are used mainly for historical reasons and are suitable for quality measurements of still but not moving pictures [7]. Human perception metrics are based on human perception models as well as the recognition of lossy video coding artifacts like blockiness, blurriness and jerkiness. Watermarking metrics are estimated by watermarking the original video sequences and the quality is measured according to the watermark degradation. This method depends strongly on the codec type. Video quality metrics which consider objective video parameters like spatial information (SI), temporal information (TI), frame rate (FR), bit rate (BR) and PSNR are suitable for different types of video codecs like MPEG2, MPEG4, H.263 [5].

The H.264/AVC codec (more details about it can be found in [8]) is a recent video coding standard of the ITU-T Video Coding Experts Group and the ISO/IEC Moving Picture Experts Group. This codec provides significant video coding improvements and new technical features that improve its compression performance keeping the same quality, what makes it in particular suitable for wireless networks. Blockiness [9] is generally considered to be one of the most annoying artifacts when using block-based coding. For this reason H.264/AVC defines an adaptive in-loop deblocking filter, where the strength of filtering is controlled by the values of several syntax elements. The blockiness is reduced without much affecting the sharpness of the content. Consequently,

the subjective quality is significantly improved. Furthermore, H.264 uses a small block-size transformation that allows the encoder to represent signals in a more locally-adaptive fashion. H.264 exhibits outstanding performance for low bit rates making this codec very suitable for narrow-band wireless applications.

Consequently to significant improvement of visual perceptual quality in H.264/AVC, most of the usually used metrics are not applicable [1], [5], [7] for video quality measurements. Additionally, the results for panorama type of sequences differs significantly from the other sequence types. According to subjective video quality assessments [9], the sequences with lower frame rates are more accepted although it is to expect that quality will decrease more with decreasing frame rate like in other types of sequences.

In this paper we analyze the coding performance of the H.264 codec using low bit and frame rate panorama sequences. Our analysis is based on the subjective quality evaluation results. We propose a new mean opinion score (MOS) prediction metric considering objective video quality parameters.

The paper is organized as follows: In Section 2 the panorama sequences selected for evaluation are described as well as the setup of the survey which we performed to obtain MOS values. Section 3 explains the used test setup. The results are further interpreted in Section 4. Focus is given on the methodology of metric design. Section 5 contains conclusions and final remarks.

II. PANORAMA SEQUENCES CHARACTERISTICS

According to [5] video sequences are characterized by spatial information (SI) and temporal information (TI). As subjective quality evaluation is a psycho visual experiment, the results strongly depend on the type and character of the sequence itself. The intention of this paper is to demonstrate the dependency of the MOS on the sequence character of panorama sequences by means of a survey. Panorama sequences have rather different character compared to typical soccer or TV news (talking head) sequences [7], [9], as is shown in Table 2.

Three typical non-professional panorama test sequences were chosen, namely Winter Nature, Danube River and Traffic. The first and second sequences contain uniform but smooth

and relatively slow constant movement of the scene. The third sequence is static without camera movement. Sequence 1 is a video clip, a winter countryside with a few trees and bushes containing only a few edges. Sequence 2 contains big consistent glass surface of the buildings and the river Danube. Sequence 3 is obtained by a traffic camera. This is not a typical panorama sequence because the camera is static and slowly moving cars are observed. Snapshots of all these sequences are presented in Figures 1 - 3. In Tab. 1 the bit rates and



Fig. 1. Sequence 1: Winter Nature

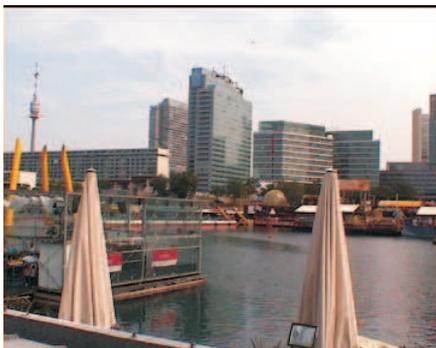


Fig. 2. Sequence 2: Danube River



Fig. 3. Sequence 3: Traffic

frame rates chosen for subjective perceptual quality test. We chose the values for FR and BR to cover typical scenarios and possibilities for 3G video streaming. The frame rates 7 and 10 at bit rate 24 kbps were not used because the codec H.264 does not allow such high spatial compression.

Frame Rate [fps]	Bit Rate [kbps]
3	24
3	44
3	80
5	24
5	44
5	80
7	44
7	80
10	44
10	80

TABLE I
H.264/AVC CODEC SETTINGS

III. TEST SETUP FOR VIDEO QUALITY EVALUATION

To evaluate the subjective visual perceptual quality, 15 unpaid persons (in the age between 20 and 35) were tested. These tests were performed according to [10], using an LCD screen with a QCIF resolution picture located in the middle of the gray background. Absolute category rating (ACR) tests were performed. Test subject were asked to evaluate the total quality with five grade scale (1-excellent, 2-good, 3-poor, 4-fair, 5-bad). Reverse MOS scale was used because such rating corresponds to the Austrian school scale system and tests subjects are familiar with this scale. Panorama sequences were presented in an arbitrary order, with additional condition that the same sequence (even differently degraded) did not appear in succession. Each H.264 coded sequence was played twice in arbitrary order (test persons did not know about it), to estimate the individual variance of the test persons. Sequences which test subject evaluated with individual variance higher than one were not used in further data processing. After the test, subjects were asked to fill out a small questionnaire in order to obtain information about their age, sex, education and experience with imaging.

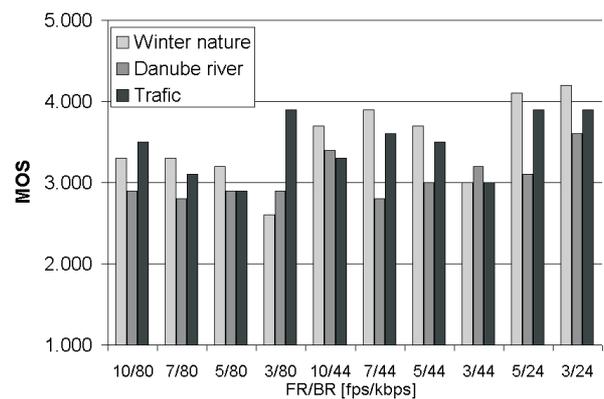


Fig. 4. MOS results for all sequences and frame/bit rate codec settings combinations.

Figure 4. shows averaged MOS values over all users. As already mentioned, ACR tests were performed. Therefore after averaging, the highest and lowest MOS values were not reached. The MOS varied between 2.5 and 4.5. For the first and second sequence with uniform camera movement test subjects favour lower frame rates and higher spatial resolution

(Figure 4). The only exception was observed for the lowest bit rates (24 kbps). The reason is that the spatial compression in bit rates less than 24 kbps is very high and too many spatial details are lost. In the third sequence (Traffic) the panorama effect was not observed because this sequence does not contain uniform camera movement.

IV. OBJECTIVE VIDEO QUALITY PARAMETERS AND METRIC DESIGN

The objective parameters like BR, FR, TI, and SI [9] describe properties of video codec and sequences. SI is computed from the image gradient. It is an indicator of the amount of edges in the image:

$$SI = \max_{\text{time}} \{ \text{std}_{\text{space}} [\text{Sobel}(F_n)] \} \quad (1)$$

TI is computed from the pixel-wise difference between successive frames. It is an indicator of the amount of motion in the video:

$$TI = \max_{\text{time}} \{ \text{std}_{\text{space}} [M_n(i, j)] \} \quad (2)$$

Sequence	TI	SI
Winter Nature	104.892	18.799
Danube River	152.030	28.862
Traffic	109.782	22.529
TV News	79.4	5.2
Soccer	85.8	22.9

TABLE II
SI AND TI PARAMETERS

The PSNR metric cannot be used because there are very small variances between PSNR values for encoded and original sequences. This is caused by the fact that the temporal prediction can be made very exact as the camera movement is linear and uniform. Therefore, also for the higher compression gains, the quality given by PSNR remains similar for encoded sequences. The universal perceptual metric Q_m proposed in [7]:

$$Q_m = -0.45\text{PSNR} + 17.9 - 0.1(\text{FR} - 5) \quad (3)$$

where FR is frame rate of encoded video sequence and constant coefficients were obtained by evaluating of the data set in the survey. Please note, that prediction considers a six grade scale, the best note being one. Therefore, we have adapted this metric to our five grade scale:

$$Q'_m = 0.8Q_m + 0.2 \quad (4)$$

In the Q_m metric, the FR causes insignificant linear offset in the mapping of PSNR on the MOS. As can be seen in Figure 5 for lower PSNR, Q_m values are out of five-grade MOS scale. If we compare this quality metric with our data set, we can conclude (see Figure 5) that this metric is not suitable for perceptual quality prediction of panorama sequences. The correlation coefficient r of Q_m metric and to our data set is only 46.7 %. To evaluate the quality of the fit of different

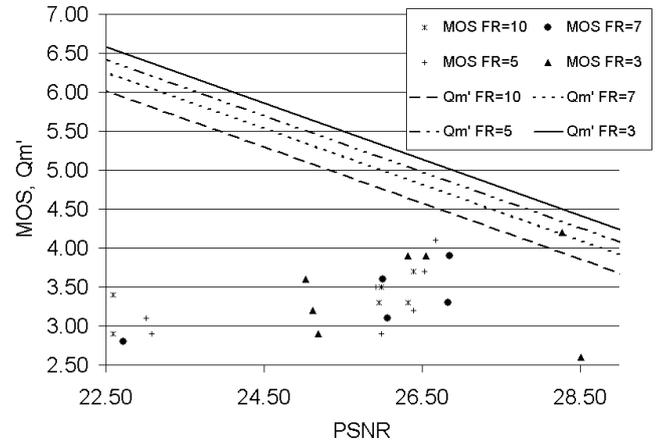


Fig. 5. Q_m prediction over PSNR and FR: metric (4) and our data

metrics for our data, we used a correlation factor defined as follows:

$$r = \frac{\mathbf{x}^T \mathbf{y}}{\sqrt{(\mathbf{x}^T \mathbf{x})(\mathbf{y}^T \mathbf{y})}} \quad (5)$$

In our case vector \mathbf{x} corresponds to MOS and vector \mathbf{y} to the metric prediction.

There are three basic reasons why this metric does not fit with our data set. First, this metric does not take into account the sequence character of panorama sequences. PSNR is not a suitable parameter for quality prediction of H.264/AVC encoded sequences (PSNR remains similar for different BR). Finally, a perceptual quality metric cannot be described by a linear function because even the simplest model of human vision [2] is not linear.

These experiences determine our future steps in metric design. We propose our metric only for the first two sequences with uniform camera movement as they have the same sequence character. We choose couples of objective parameters which reflect spatial and temporal sequence characteristics. Finally, we choose a three dimensional non-linear model. First, the efficiency of mapping the objective video parameters (FR, BR) on MOS is evaluated.

The first metric we investigated is based on basic codec parameters FR and BR according to [7]. These parameters were investigated because perceptual quality can be estimated without additional computation of the objective parameters and therefore, the original sequence is not required.

The second metric is based on SI and TI parameters. This metric requires a higher calculation effort but it is independent on the original video sequence.

The third metric is based on the relative difference of SI (SI_{rd}) and TI (TI_{rd}) (6) to the original sequence. The disadvantage of this method is that SI and TI have to be calculated for both the original and the encoded sequence:

$$SI_{rd} = \frac{(SI_{orig} - SI_{copy})}{SI_{orig}} \quad (6)$$

$$TI_{rd} = \frac{(TI_{orig} - TI_{copy})}{TI_{orig}}. \quad (7)$$

For metric fitting, a three-dimensional five-parameter polynomial model of second order was used:

$$y = a + bx_1 + cx_1^2 + dx_2 + ex_2^2 \quad (8)$$

The model has been chosen due to its simplicity and rather good fit with the measured data. On the other hand, this model reflects the non-linear human vision model [2] with two independent parameters. These parameters were obtained after the regression analysis:

X_1, X_2	a	b	c	d	e
BR, FR	4.48	-0.0447	2.82×10^{-4}	4.45×10^{-2}	6.40×10^{-5}
SI, TI	39.63	-1.39	4×10^{-3}	2.52	1.82×10^{-2}
SI _{rd} , TI _{rd}	8.92	-3.44	-32.91	5.12	2

TABLE III

POLYNOMIAL MODEL COEFFICIENTS OF (9) OBTAINED AFTER LINEAR REGRESSION ANALYSIS

X_1, X_2	Corr. coeff. R in %
BR, FR	55.03
SI, TI	79.13
SI _{rd} , TI _{rd}	90.97

TABLE IV

CORRELATION COEFFICIENTS FOR PROPOSED METRICS

The parameters BR and FR exhibit the worst fit. The reason is that the codec compression is non-linear. The video sequence is spatially and temporally compressed. The compression ratio depends on the amount of spatial and temporal information, leading to poor correlation of BR and FR parameters on MOS. These parameters can be used for metric design for uncompressed video. The second and third proposed metric have significantly better gain because SI and TI parameters are calculated from decompressed frames. The SI and TI parameters alone do not represent the sequence character. If we take the relative difference of SI and TI, we can recognize in Figure 6 that the sequence character change after compression is reflected. The best fit has the relative difference of SI and TI.

The relative difference of SI and TI parameters represents the sequence character in the best way as can be seen from the correlation coefficient (90.97%).

V. CONCLUSION

In this paper we compared, investigated and proposed three different perceptual quality metrics for panorama sequences. We evaluated performance of H.264/AVC codec for low bit and frame rate panorama sequences by the psycho-visual experiment. We chose panorama sequences and codec settings typical for streaming services in 3G network. Our results have shown that panorama sequences with camera movement have different character than panorama sequences with a static camera. We proposed three different quality metrics and we show the corresponding fit of different objective video parameters.

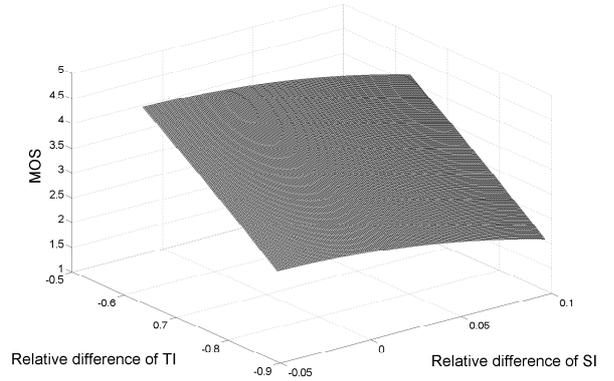


Fig. 6. Relation between relative difference of SI TI and MOS for sequences with uniform camera movement

The best fit we obtain for a metric based on relative objective parameters. The ratio of relative difference of SI and TI for panorama sequences with uniform camera movement is almost constant. These parameters are most suitable for metric design.

VI. ACKNOWLEDGMENT

The authors would like to thank mobilkom austria AG&Co KG for supporting their research. The views expressed in this paper are those of the authors and do not necessarily reflect the views within mobilkom austria AG&Co KG.

REFERENCES

- [1] S. Winkler, "A perceptual distortion metric for digital color video", in Proc. SPIE Human Vision and Electronic Imaging, vol. 3644, San Jose, California, pp. 175-184, Jan. 1999.
- [2] E.P. Ong, W. Lin, Z. Lu, S. Yao, X. Yang, F.Moschetti, "Low bit rate quality assessment based on perceptual characteristics", Image Processing, 2003, Proceedings, 2003 International Conference on, Vol. 3, pp. 182-192, Sept. 2003.
- [3] N.K. Ngan, C.W. Yap, K.T. Tan, *Video coding for wireless communication systems*, Marcel Dekker, NY, 2001
- [4] S. Winkler, E.D. Gelasca, T. Ebrahimi, "Perceptual quality assessment for video watermarking", in Proc. International Conference on Information Technology: Coding and Computing (ITCC), Las Vegas, NV, pp. 90-94, Apr. 2002.
- [5] ANSI T1.801.03 - 2003, "American National Standard for Telecommunications - Digital transport of one-way video signals Parameters for objective Performance assessment," American National Standards Institute.
- [6] M.H. Pinson, S. Wolf, "A new standardized method for objectively measuring video quality", IEEE Transactions on broadcasting, Vol. 50, Issue: 3, pp. 312-322, Sept. 2004.
- [7] G. Hauske, T. Stockhammer, R. Hofmaier, "Subjective Image Quality of Low-rate and Low-Resolution Video Sequences," Proc. International Workshop on Mobile Multimedia Communication, Munich, Germany, Oct. 2003.
- [8] T. Wiegand, G.J. Sullivan, G. Bjontegaard, G.A. Luthra, "Overview of the H.264/AVC Video Coding Standard", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 13, no. 7, pp. 560-576, Jul. 2003.
- [9] O. Nemethova, M. Ries, E. Siffel, M. Rupp, "Quality Assessment for H.264 Coded Low-Rate and low-Resolution Video Sequences", Proc. of Conference on Internet and Information Technologies (CIIT), St. Thomas, US Virgin Islands, pp. 136-140, Nov. 2004.
- [10] ITU-T Recommendation P.910, "Subjective video quality assessment methods for multimedia applications", Sept. 1999.