

# A New Low Complex Reference Free Video Quality Predictor

Andreas Rossholm<sup>†‡1</sup>, Benny Lövestrom<sup>†2</sup>

<sup>†</sup>, *Department of Signal Processing  
School of Engineering  
Blekinge Institute of Technology  
Ronneby, Sweden*

<sup>2</sup>, *benny.lovstrom@bth.se*

<sup>†‡</sup>, *Video Technology  
Ericsson Mobile Platforms AB  
Lund, Sweden*

<sup>1</sup>, *andreas.rossholm@ericsson.com*

**Abstract**—In many applications and environments for mobile communication there is a need for reference free perceptual quality measurements. In this paper a method for prediction of a number of quality metrics is proposed, where the input to the prediction is readily available parameters at the receiver side of a communications channel. Since the parameters are extracted from the coded video bit stream the model can be used in user scenarios where it is normally difficult to estimate the quality due to the reference not being available, as in streaming video and mobile TV applications. The predictor turns out to give good results for both the PSNR and the PEVQ metrics.

## I. INTRODUCTION

There is a growing demand for objective quality measurement techniques estimating perceived video quality in mobile devices. This is of interest to, among others, the mobile phone industry, mobile network operators and software developers. A review of the area of perceptual video quality metrics can be found in [1]. The quality of a video encoder and decoder can be measured with different metrics. Two of the most commonly used metrics are the peak signal-to-noise ratio (PSNR) and the mean-squared-error (MSE) which can be calculated for each decoded frame and then averaged for the complete sequence.

Since both MSE and PSNR are based on a pixel-by-pixel comparison the metrics have some issues regarding the relation to the perceptual quality. This has resulted in the development of several new metrics as SSIM [2], a video adapted version of SSIM denoted VSSIM [3], NTIAs VQM [5], and Opticom's PEVQ [6]. However, all these metrics use the original frame as reference and calculate some ratio between this and the decoded frame, full-reference (FR). In many situations where perceptual quality is of interest, e.g. streaming video, video telephony, MBMS, and DVB-H the original frames are not available. Thus there is a need for reference free quality metrics, which can be implemented entirely at the decoder side of a transmission line.

There are a few no-reference (NR) metrics that have been developed but they often focus on one parameter such as

blur [7], blockiness [8], motion [9], or a combination of motion and distortion as [10]. These metrics all requires processing of the received frame and are thereby often less useful in real applications as in mobile equipment where the computational power frequently is limited.

## II. THE PROPOSED IDEA

When a video sequence is encoded to fulfil the required properties such as bit rate, frame rate and resolution, the encoder sets and adjusts a number of parameters. Some of these are set for the whole sequence while some are adjusted for each frame or within frames. The coding results in a bit stream consisting of motion vector parameters, coded residual coefficients and header information e.g. frame rate and quantization parameters (QP) value. From the bit stream it is also possible to calculate the number of intra blocks, number of inter blocks, number of skipped blocks, etc. The idea proposed in this paper is to predict the video quality using these parameters. The predictor is built by setting up a model and adapt its coefficients using a number of training sequences. The parameters used are available at the decoder and therefore the quality predictor is reference free.

Throughout this paper the video sequences are coded using the H.264 standard, since this is one of the most used video encoder for mobile equipment. The parameters chosen for the predictor are

- 1) Average QP value (Avg QP)
- 2) Bitrate /Frame rate (Bits/Frame)
- 3) Number of intra blocks (Intra [%])
- 4) Number of inter blocks (Inter [%])
- 5) Number of skipped blocks (Skip [%])
- 6) Frame rate
- 7) Number of inter blocks of size 16x16 (P16x16[%])
- 8) Number of inter blocks of size 8x8, 16x8, and 8x16 (P8x8 [%])
- 9) Number of inter blocks of size 4x4, 8x4, and 4x8 (P4x4 [%])

10) Average motion vector length (Avg MV [%])

Also, other parameters could be extracted and evaluated but these were chosen based on their expected potential contribution to the perceptual quality.

### III. THE METRICS PREDICTED

The proposed model will in this paper be evaluated in predicting the following quality metrics; PSNR, SSIM, VSSIM, NTIA VQM, and PEVQ.

PSNR, the peak signal-to-noise ratio, is defined as  $PSNR = 10 \cdot \log MAX_I^2 / MSE$  [dB], where  $MAX_I$  is the maximum value a pixel can take (e.g. 255 for 8-bit images) and the MSE is the average of the squared differences between the luminance values of corresponding pixels in two frames. To get a measure for a video sequence a simple averaging over frames is made.

SSIM, the Structural Similarity index, considers image degradations as perceived changes in the variation of structural information by combining measures of the distortion in luminance, contrast and structure between two frames [2].

VSSIM, the Video Structural Similarity index, is an adaptation of the SSIM metric to quality evaluation for video. This is performed by a weighted average, where the weights are determined by calculated motion vectors. VSSIM was developed using the VQEG (Video Quality Experts Group) Phase I test data set for FR-TV video quality assessment [4]. Since these data consist of larger frame sizes (SD-resolutions, 525-line and 625-line) than the QCIF used in this paper a modified VSSIM has also been used in the proposed solution to adapt it to smaller resolution. This is accomplished by down scaling the weighting coefficient  $K_M$  and its connected thresholds with a factor of 8, from 16 to 2 [3].

NTIA VQM, the National Telecommunications and Information Administrations general purpose Video Quality Model general model, is a reduced reference method containing linear combination of seven objective parameters for measuring the perceptual effects of a wide range of impairments such as blurring, block distortion, jerky/unnatural motion, and noise [5]. VQM general was included in the Video Quality Experts Group (VQEG) Phase II Full Reference Television (FR-TV) tests [4].

PEVQ, the Perceptual Evaluation of Video Quality, calculates measures from the differences in the luminance and chrominance domains between corresponding frames. Also motion information is used in forming the final measure [6]. PEVQ has been developed for low bit rates and resolutions as CIF ( $352 \times 288$ ) and QCIF ( $176 \times 144$ ). PEVQ is a proposed candidate for standardization of an FR video model within VQEG [4] which is in the process of starting verification tests for future standardization.

### IV. THE MATHEMATICAL MODEL

The problem can be presented as an observation matrix,  $X = [x_1 x_2 \dots x_N]$ , where  $x_1, x_2, \dots, x_N$  are a number of feature vectors that have been generated with different video content and codec setups. Each feature vector  $x_n$  consists

of extracted codec parameters denoted  $x_1, x_2, \dots, x_K$ . The corresponding quality measures for the different video content, PSNR, PEVQ, SSIM, VSSIM, and NTIM then correspond to the desired  $Y = [y_1 y_2 \dots y_N]$ .  $X$  and  $Y$  can be viewed as training data for a classification, mapping or regression problem. It is desired to find a function  $Z = f(x)$  that maps the given values in  $x$  to a specific value  $Z$ , e.g. an estimation of PSNR.

There are several different models solving the problem, that are more or less computationally complex. Because a low complex solution is required in order to have the possibility for an implementation in a mobile device, multi-linear regression is selected.

The multi-linear model is formulated as:

$$Y = \beta x + \epsilon \quad (1)$$

where  $\epsilon$  represents the unpredicted variation. The multi-linear regression estimates the values for  $\beta$  denoted  $\hat{\beta}$  that can be used to predict  $Z$  as

$$\hat{Z} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_K x_K \quad (2)$$

#### A. Predicted Metric Evaluation

To be able to evaluate the accuracy of the predicted metric *Pearson linear correlation coefficient* can be used. It is defined as follows:

$$r_P = \frac{\sum(\hat{Z}_i - \hat{Z}_{mean})(Z_i - Z_{mean})}{\sqrt{\sum(\hat{Z}_i - \hat{Z}_{mean})^2} \sqrt{\sum(Z_i - Z_{mean})^2}} \quad (3)$$

where  $\hat{Z}_{mean}$  and  $Z_{mean}$  are the mean value of estimated respective true data set. This assumes a linear relation between the data sets.

### V. VIDEO SOURCE SEQUENCES

To generate training and verification data different sequences with different characteristic (amount of motion, color, heads, animations) were used. The source sequences had QCIF ( $176 \times 144$ ) resolution and were generated with different frame rates, 30, 15, 10, and 7.5 frames per second (fps), and bitrates, approximately: 30, 40, 50, 100, 150, and 200 bits per second (bps). The video sequences were approximately 3 seconds long (90, 45, 30, and 23 frames) and they were encoded with the H.264/MPEG-4 AVC reference software, version 12.2 generated by JVT [11] using the baseline profile.

The sequences for training were: Foreman, Cart, Mobile, Shine, Fish, Soccer goal, and Car Phone resulting in 168 sequences for training. For verification five different parts from a cropped version of the 3G-sequence was used, where the five parts have different characteristics. The cropping was made to QCIF without the original letter box aspect ratio. Varying the bitrate and the frame rate in the same way as for the training data results in 120 verification sequences.

## VI. RESULTS

Applying the linear regression described above to the 168 training sequences for each of the parameters separately gives a measure of the correlation between this parameter and the quality metric. The outcome is shown in Figure 1, where it can be seen that the parameters have considerably varying correlation, but also that it differs between the metrics.

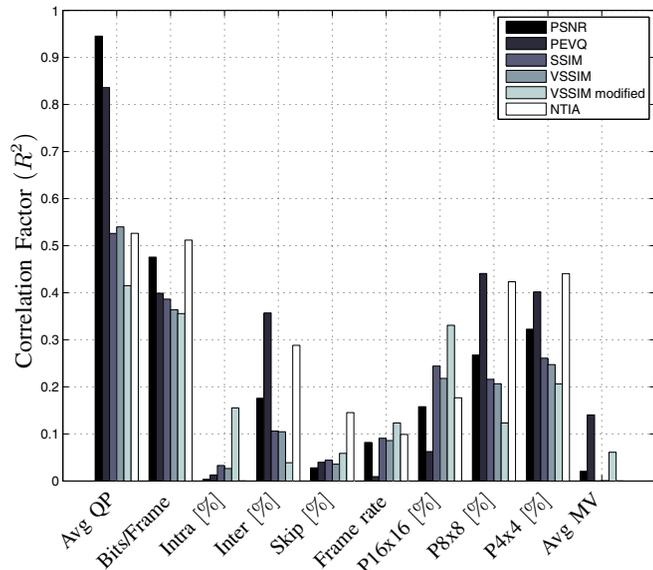


Fig. 1. The correlation factor  $R^2$  between each of the parameters and the metrics used

Using the full parameter set from Figure 1 for the multi-linear regression would give some redundancy. For this reason two parameters are removed before the regression and prediction are made, namely "Intra [%]" and "P16x16 [%]". The parameters "Intra [%]", "Inter [%]" and "Skip [%]" will together always add to a total of 100%, and since "Intra [%]" has the lowest correlation factor  $R^2$  this is removed. In the same way, "P16x16 [%]", "P8x8 [%]" and "P4x4 [%]" sums up to a total of 100% and therefore "P16x16 [%]" is removed. Performing the regression with the remaining eight parameters using the training sequences gives a prediction function  $\hat{Z}$  for each metric. The resulting coefficients in this function  $\hat{Z}$  (see Eq. (2)) is shown in Table I. These functions are applied to the verification sequences to predict the quality metric for these. Further, the quality metrics are calculated according to their definitions, and the *Pearson correlation coefficient*,  $r_P$ , from Eq. (3) is calculated and shown in Table II. In the Figures 2 – 5 the true metrics are plotted versus the predicted metrics. Note that the scale differs between the figures since the metrics have different ranges.

It can be seen from the table and the figures that the best prediction is obtained for the PSNR metric. This is expected since the JM encoder uses rate distortion optimization where the distortion measure is correlating with the PSNR. Also the PEVQ metric is well predicted, with a correlation coefficient of 0.94. This gives the possibility to implement the proposed, no-reference, metric in environments where full or reduced

Metric	$r_P$
PSNR	0.99
SSIM	0.62
VSSIM	0.62
VSSIM modified	0.72
NTIA	0.73
PEVQ	0.94

TABLE II

The Pearson correlation coefficient,  $r_P$ , for the prediction of the different quality metrics

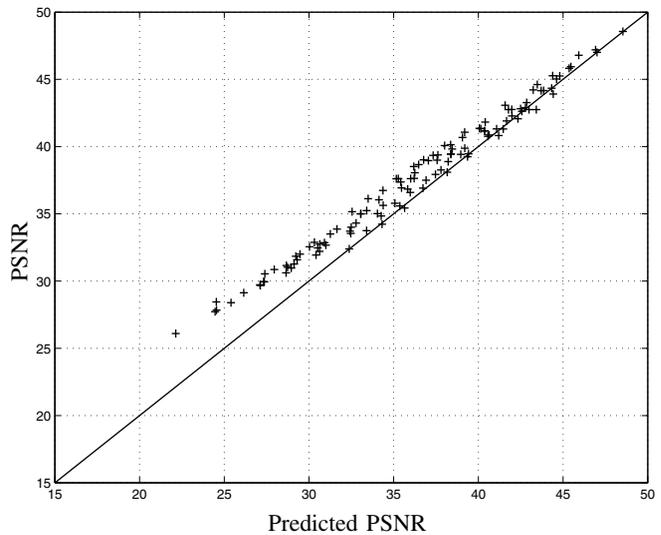


Fig. 2. PSNR in dB vs. predicted PSNR for each verification sequence,  $r_P = 0.99$

reference metrics are not possible to implement. The metric is also of low complexity, since the prediction is a simple calculation of the function in Eq. (2).

## VII. CONCLUSION

A low complex, no-reference method to predict perceptual quality metrics, based on extracted video bit stream parameters, has been suggested. For the PSNR and PEVQ metrics a very good precision is achieved, while for the other the correlation is weaker. The result for PSNR is expected since rate distortion optimization is used, while the good result also for PEVQ gives great potential to use the method in an no-reference quality evaluation application. The good result for PEVQ is also promising since PEVQ is developed to measure the perceptual quality for low resolution and bitrates, and is proposed for standardization.

Looking further into the behavior of each parameter in the prediction function shows that there are dependencies of non-linear type for some of these, which suggests a modification of the prediction function to take this into account. This will be addressed in further investigations, and is expected to further improve the accuracy. To get a more robust prediction the method will also be tested with other video sequences and encoders.

Metrics	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$	$\hat{\beta}_7$	$\hat{\beta}_8$	Scale
PSNR	69.77	-0.94	-0.12	-0.04	-0.02	-0.03	-0.07	0.06	-0.03	1.0 exp -0
SSIM	112.96	-0.49	0.11	-0.41	-0.03	0.03	0.64	-0.06	-0.02	1.0 exp -2
VSSIM	117.01	-0.55	0.04	-0.37	-0.04	0.01	0.53	-0.00	-0.08	1.0 exp -2
VSSIM modified	120.83	-0.55	-0.13	-0.42	-0.08	0.00	0.50	0.16	0.24	1.0 exp -2
NTIA	49.96	-0.54	1.47	-0.66	0.14	0.39	2.16	-1.05	0.65	1.0 exp -2
PEVQ	58.84	-0.95	0.10	-0.10	-0.02	0.21	0.31	-0.14	-0.30	1.0 exp -1

TABLE I  
The values of  $\beta_i$  in Eq. (2) for the different metrics, resulting from the regression

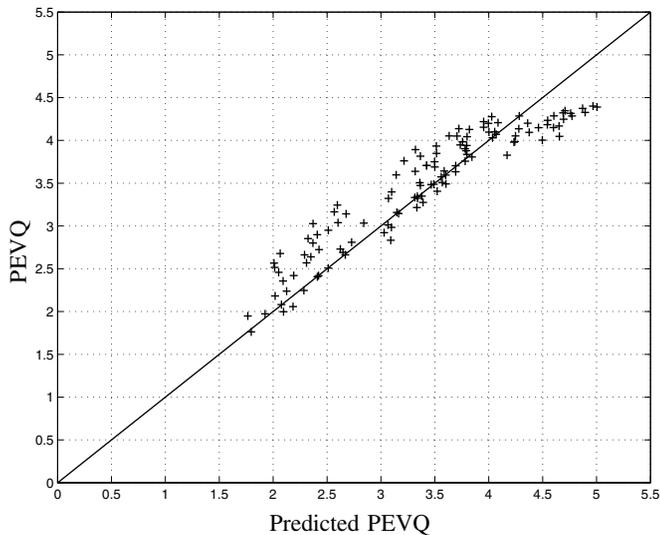


Fig. 3. PEVQ vs. predicted PEVQ for each verification sequence,  $r_P = 0.94$

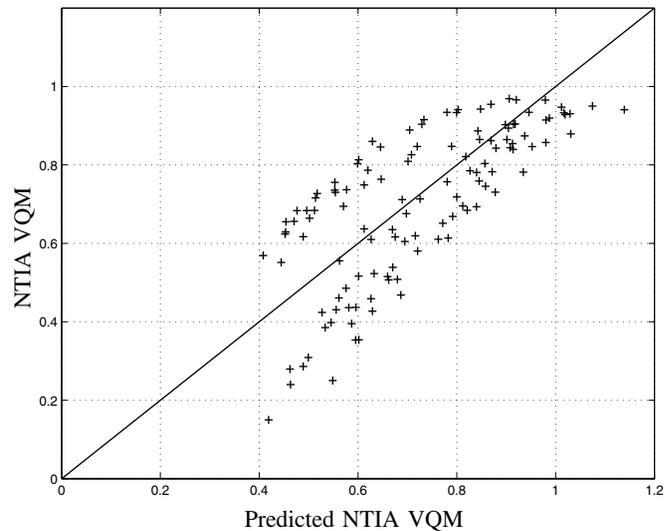


Fig. 5. NTIA VQM vs. predicted NTIA VQM for each verification sequence,  $r_P = 0.73$

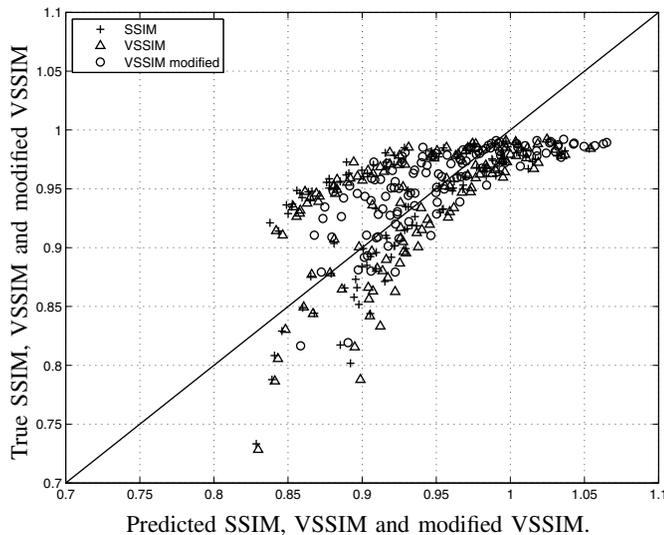


Fig. 4. SSIM, VSSIM and modified VSSIM versus the predicted values for each verification sequence,  $r_P = 0.62$ ,  $0.62$  and  $0.72$

## REFERENCES

[1] S. Winkler, "Perceptual Video Quality Metrics - A Review", Chapter 5, in H.R. Wu and K.R. Rao, Eds., *Digital Video Image Quality and Perceptual Coding*, pp.155-179, CRC Press, 2006.  
[2] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image

quality assessment: From error visibility to structural similarity" *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600-612, April. 2004.  
[3] Z. Wang, L. Lu and A.C. Bovik, "Video quality assessment based on structural distortion measurement", *Signal Processing: Image Communication*, Special issue on Objective video quality metrics, vol. 19, no. 2, February 2004.  
[4] "FR-TV Phase II Final Report", 2003-08-25, VQEG: The Video Quality Experts Group, <http://www.vqeg.org/projects/>, (verified 2008-01-18)  
[5] M. H. Pinson, S. Wolf, "A New Standardized Method for Objectively Measuring Video Quality". *IEEE Transactions on Broadcasting*, vol. 50, no. 3, September. 2004.  
[6] "PEVQ Advanced Perceptual Evaluation of Video Quality (PEVQ Whitepaper)", *Opticom*, <http://www.pevq.org>, (verified 2008-01-18)  
[7] P. Marziliano, F. Dufaux, S. Winkler, T. Ebrahimi, "A No-Reference Perceptual Blur Metric", *IEEE International Conference on Image Processing 2002*, volume 3, pp. III-57 - III-60, Rochester, USA, September. 2002.  
[8] W. Zhou, A. C. Bovik, B. L. Evans. "Blind measurement of blocking artifacts in images", *IEEE International Conference on Image Processing 2000*, volume 3, pp. 981 - 984, Vancouver, Canada, September. 2000.  
[9] M. Ries, O. Nemethova, M. Rupp, "Motion Based Reference-Free Quality Estimation for H.264/AVC Video Streaming", *2nd International Symposium on Wireless Pervasive Computing (ISWPC '07)*, pp. 355 - 359, San Juan, Puerto Rico, USA, February 2007.  
[10] F. Yang, S. Wan, Y. Chang, H. R. Wu, "A novel objective no-reference metric for digital video quality assessment", *Signal Processing Letters, IEEE*, vol. 12, issue 10, pp. 685- 688, October 2005.  
[11] "H.264/MPEG-4 AVC REFERENCE SOFTWARE", Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG (ISO/IEC JTC1/SC29/WG11 and ITU-T SG16 Q.6), 24th Meeting: Geneva, CH, 29 June - 5 July, 2007