



Copyright © IEEE.
Citation for the published paper:

This material is posted here with permission of the IEEE. Such permission of the IEEE does not in any way imply IEEE endorsement of any of BTH's products or services. Internal or personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution must be obtained from the IEEE by sending a blank email message to pubs-permissions@ieee.org.

By choosing to view this document, you agree to all provisions of the copyright laws protecting it.

A NO-REFERENCE MACHINE LEARNING BASED VIDEO QUALITY PREDICTOR

Muhammad Shahid, Andreas Rossholm, and Benny Lövström

Department of Electrical Engineering, Blekinge Institute of Technology
SE-37179 Karlskrona, Sweden

Corresponding author email: muhammad.shahid@ieee.org

ABSTRACT

The growing need of quick and online estimation of video quality necessitates the study of new frontiers in the area of no-reference visual quality assessment. Bitstream-layer model based video quality predictors use certain visual quality relevant features from the encoded video bitstream to estimate the quality. Contemporary techniques vary in the number and nature of features employed and the use of prediction model. This paper proposes a prediction model with a concise set of bitstream based features and a machine learning based quality predictor. Several full reference quality metrics are predicted using the proposed model with reasonably good levels of accuracy, monotonicity and consistency.

Index Terms— Video Quality, H.264/AVC, Bitstream Features, No-Reference, Support Vector Machine

1. INTRODUCTION

The recent advancement in digital imaging technology and availability of efficient transmission systems have resulted in a proliferation of videos more than ever before. Videos transmitted to and from mobile devices will account for 66% of the global mobile data traffic by 2017 as per forecasts [1]. Video services that have gained wide interest are many and television broadcast, DVD, Blu-Ray, Mobile TV, Web TV etc. are some to name. One of the key characteristics of video services is perceptual aspect of the quality of experience (QoE) as observed by the end user. Quality of visual media can get degraded during capture, storage, transmission, reproduction and display due to distortions which might occur at any of these stages. The true judges of video quality are humans as end users of the video services. The scientific process of evaluation of video quality by humans is called subjective quality assessment and it is usually given as mean opinion score (MOS) value. However, subjective evaluation is often too inconvenient, time-consuming, expensive and it has to be done by following special recommendations in order to yield reproducible standard results. These reasons give rise to the need of some intelligent ways of automatically estimating the perceived quality that can be performed swiftly and economically. This non-subjective way of quality evaluation is known

as objective quality assessment. Due to the overwhelming number of approaches on which objective methods of video quality assessment (VQA) are based, there are many ways we can classify them. For example, *data metrics* such as PSNR and MSE consider only the fidelity of the signal and remain agnostic to the visual content. On the other hand, *picture metrics* do take into account the visual information present in the signal. Moreover, objective metrics can also be divided on the basis of the reference information used for the quality estimation. Such metrics fall in one of the three categories namely full-reference (FR), reduced-reference (RR) and no-reference (NR) metrics [2].

In the area of video quality, Video Quality Expert Group (VQEG) has the paramount role of validation of objective metrics for International Telecommunication Union (ITU) [3]. ITU has been the cardinal body for issuing recommendations and performing standardization activities for objective methods of video quality. Based on the normative reports from ITU, objective metrics have been classified into the following five groups in terms of the type of data used.

1. Media-layer models: Models found in this category estimate quality value using the video signal. Applications such as codec comparison and optimization are common uses of these models.
2. Parametric packet-layer models: These models do not access the media signal but employ the packet-header information to estimate visual quality.
3. Parametric planning models: In these models, the quality planning parameters for network and terminals are employed for the assessment of quality.
4. Bitstream-layer models: These models use the encoded bitstream and packet-layer information for the quality assessment.
5. Hybrid models: A composite of two or more models mentioned earlier belongs to this category.

PSNR and MSE have been the traditional pixel by pixel comparison objective metrics which, however, don't correlate well with perceptual assessment of humans [4]. It is desirable to have perceptually relevant objective metrics, hence Structural Similarity Index (SSIM) and Perceptual Evaluation of Video Quality (PEVQ) included in ITU-T Rec. J.247 are preferred.

Our contribution in this paper builds on our previous works [5] and [6]. The use of neural network for image quality prediction has given results better than the use of linear regression technique [6]. However, neural network techniques often face the problem of overfitting and the computational complexity of neural network based methods can grow with increase in dimension of input space. In order to circumvent such issues, support vector machine technique is used [7]. Mainly, this paper presents: carefully performed subjective tests of quality assessment on a varying content of video sequences, a rational selection of the bitstream features based on their contribution towards quality estimation, improved results by using a better regression technique and a comprehensive comparison of results using commonly used FR quality metrics such as PSNR, SSIM, MSSIM and PEVQ. The remaining part the document is organized as follows. A description of the test media generation and its subjective quality assessment, used in this work, is provided in Section 3. It is followed by the details of the proposed model of quality prediction in Section 4. A comparison of the proposed method with similar techniques is presented in Section 5. Finally, some conclusive remarks about this work are drawn in Section 6.

2. RELATED WORK

With regards to the use of quality metrics in practice, such as in-service monitoring, only NR metrics are suitable as availability of a reference video signal (FR) or some features of the reference video (RR) is not possible mostly. Bitstream-layer models provide a low complexity solution for NR quality assessment. The parameters used in these methods can be obtained without fully decoding a video. A low complexity solution of video quality prediction based on bitstream-layer parameters is found in [5]. Low complexity has been achieved by using a simple multi-linear regression system for estimating a relationship between the parameters and quality values. An improvement of this approach was presented in [6] where the required number of parameters have been reduced for computational efficiency and the prediction accuracy has been improved by the virtue of the usage of artificial neural network (ANN). The results presented in both of these methods were not, however, validated by subjective scores. A set of 48 bitstream parameters related to slice coding type, coding modes, various statistics of motion vectors and QP value was used in [8] to predict quality of HDTV videos. Partial least squares method was used as a tool for regression between the feature set and subjective MOS. However, the computational complexity for processing such a huge set of parameters on slice level can be a point of concern. Two methods tested on MPEG-4 encoded videos and based on several macroblock (MB) level statistics, derived from bitstream and reconstructed videos, were reported in [9] for PSNR and in [10] for SSIM. In the metric targeted for for PSNR esti-

mation, Shanableh [9] has employed spectral regression and reduced model polynomial network. A multi-pass prediction system based on stepwise regression has been used in [10]. The video features, in both of these metrics, constitute mainly coding information of an MB, some relative measures of motion vectors of neighboring MBs and a couple of numerical values related to texture of an MB. Similarly, a PSNR estimator for H.264/AVC encoded video is presented in [11] where bitrate, QP value and coding mode are used as the features for quality prediction. Prediction of perceptual quality (subjective MOS) can be an applicable improvement of these methods.

3. SUBJECTIVE VIDEO QUALITY ASSESSMENT

Subjective evaluation of video quality is the fundamental ground truth to test and compare the performance of objective methods of quality assessment. It involves setting up an experimental environment according to specific recommendations where a prescribed number of people watch the test set of video sequences and rate the videos in terms of the visual quality. International Telecommunication Union (ITU) has two sub organizations namely Radiocommunication Unit (ITU-R) and Telecommunication Unit (ITU-T) which are responsible for standardizing such recommendations.

3.1. Test Video Sequences

Six different video sequences of CIF and QCIF spatial resolutions were selected in raw progressive format with different amount of motion contents and texture complexity. A method put forward by ITU-R P.910 [12] recommends an approach to characterize videos on the basis of their spatial perceptual information (SI) and temporal perceptual information (TI). SI and TI are calculated in the luminance plane of a video. The test videos used in this work have been selected to represent both of these two parameters from low to high range and Fig. 2 shows a plot for the chosen videos, displayed in Fig. 1. The selected videos were encoded at two frame rates and five different bitrates, resulting into a test set of 120 videos, Table 1 shows these values in detail. The encoding was done using JM reference software (<http://iphome.hhi.de/suehring/tml>) based on H.264/AVC standard in baseline profile mode. This encoding approach is specifically useful in applications that require low delay (such as video conferencing) as bi-predictive coding (B frames) are not used in it and for the hand-held devices with low computational power. Each of 120 video sequences were of 20 seconds in length. The first frame is coded as intra (I) frame and has comparatively high quality which is supposed to reduce for upcoming successive frames due to predictive coding. In order to have relatively uniform quality of the whole test sequence, last 10 seconds of each video were selected for the subjective quality assessment to remove the I frame effect and to avoid forgiveness effect [13].



Fig. 1. A snapshot of the test video sequences

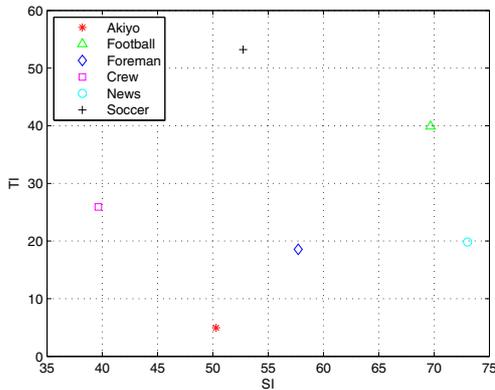


Fig. 2. SI and TI plot computed for luminance component of selected CIF and QCIF videos

3.2. Subjective Quality Assessment Setup

We considered the recommendations given by ITU-R BT 500-12 [14] for the lab setup of our experiments. Particularly, the method followed was single stimulus quality evaluation where a test video sequence is shown once without the presence of any explicit reference, corresponding to the reality where users see only the processed version of videos [15]. A detailed account of these subjective tests is presented in [16]. In order to obtain reliable results out of raw subjective scores, screening of the observers was employed to discard observers that are considered as outliers. The algorithmic details of these steps are reported in Annex 2 of [14]. After performing the refining process, mean opinion score (MOS) was calculated and used in this work.

4. THE PROPOSED METHOD

The proposed method consists of extraction of visual quality relevant bitstream parameters/features and building of a machine learning based model for quality prediction. These parameters were selected carefully to keep the complexity in control and to get reasonable coding information which can represent the degradations caused on visual quality by compression. Here, we describe the extracted parameters and the rationale of including a feature in the proposed model. Note that the terms *parameter* and *feature* are used alternatively to

Table 1. The Test Video Sequences

Resolution	Video Sequence	Bitrate(kbps)
CIF @ 30,15 fps	Akiyo	200,400,600,800,1000
	News	200,400,600,800,1000
	Foreman	200,400,600,800,1000
	Crew	200,400,600,800,1000
	Soccer	200,400,600,800,1000
	Football	200,400,600,800,1000
QCIF @ 30,15 fps	Akiyo	100,200,300,400,500
	News	100,200,300,400,500
	Foreman	100,200,300,400,500
	Crew	100,200,300,400,500
	Soccer	100,200,300,400,500
	Football	100,200,300,400,500

mean the same in this context. We broadly divide the chosen bitstream parameters into two types namely *Coding Distortion* and *Content Structural Information* as given below:

- **Coding Distortion:** Quantization is the main source of distortion introduced during the video encoding and the Quantization Parameters (QP) value is used to steer the scale of quantization. The QP value has its impact on the bitrate and it is observed that a video sequence coded at higher bitrate has higher visual quality and the quality is degraded by lowering the coding bitrate when the encoding settings is fixed.
- **Content Structural Information:** In H.264/AVC, the 16x16 macro block (MB) can be subpartitioned into blocks of sizes 4x4, 4x8, 8x4 or 8x8, 8x16, 16x8 depending on the coding mode being chosen for the sake of minimal error in prediction [17]. The relative selection of these alternatives during encoding provides an estimate of the structural information of the content present in a video. The ratio of intra blocks out of the total blocks in an inter frame would give an estimate of temporal complexity present in a video sequence. Motion Vector (mv) based statistics can be used to characterize the motion contents of a video.

As the rate control mechanism in the used JM reference software is driven to find the optimum coding modes based on a requested bitrate, the above mentioned two types have an aggregate impact on the visual quality and it differs depending on the content of a sequence.

4.1. Feature Extraction of H.264/AVC Bitstream Data

In the light of the discussion given before, following features were extracted from the H.264/AVC coded bitstream data of the test video sequences. **1.** Average bitrate. **2.** Percentage of Intra coded macroblocks. **3.** Percentage of Inter coded macroblocks. **4.** Percentage of skipped macroblocks. **5.** Percentage of inter macroblocks of size 16x16. **6.** Percentage of inter macroblocks of size 8x8, 16x8, and 8x16. **7.** Percentage of inter macroblocks of size 4x4, 8x4, and 4x8. **8.** Average of motion vector difference lengths in X direction. **9.** Average of motion vector difference lengths in Y direction. **10.** Average of absolute motion vector lengths in X direction. **11.** Average of absolute motion vector lengths in Y direction. **12.** Percentage of macroblocks with absolute motion vector of length zero. **13.** Percentage of macroblocks with motion vector difference of length zero. **14.** Motion Intensity (I) defined by:

$$\sum_{i=1}^N \sqrt{MVX_i^2 + MVY_i^2} \quad (1)$$

where N is the total number of macroblocks in each frame and MVX_i and MVY_i are the absolute motion vector values of the i-th macroblock in X and Y directions respectively. **15.** Motion Intensity (II) defined by:

$$\sqrt{MVX^2 + MVY^2} \quad (2)$$

where MVX and MVY are the average of absolute motion vectors in each frame in X and Y directions respectively. **16.** Percentage of Intra coded macroblocks of size 4x4 in I frames. **17.** Percentage of Intra coded macroblocks in P frames. **18.** Average quantization parameter.

The features were computed on frame level and an average value of each feature was obtained for a video. In this work, the coded sequences had only one I frame (1st one), so the feature number 16 was not included in our current model.

4.2. Proposed Prediction Model

For VQA, in order to predict the quality value based on an objective model, first we need to train such a model on a set of video features with known quality value. Given a matrix X of n features of m videos with the corresponding n values of the quality in vector y , where m and n are real positive numbers, as expressed by the following:

$$X = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_n\} \quad X \in R^{m \times n} \quad (3)$$

$$\mathbf{y}^T = \{y_1, y_2, y_3, \dots, y_n\} \quad \mathbf{y} \in R^{n \times 1} \quad (4)$$

A given prediction model is trained using this n-dimensional feature set for target values of quality measure given in y . This training process can be performed by building a functional model between the video features set and the corresponding quality values set. The trained model is used to predict quality measure of another set of n-dimensional feature

set of any k test videos, for which it was not trained with. The accuracy of the prediction can be evaluated by using the actual quality measure of the k videos.

One of the regression analysis techniques relies on kernel based learning methods. These methods solve any problem by mapping the input data set into high dimensional feature space via linear or nonlinear mapping. Support vector machine (SVM) is a supervised kernel based learning algorithm and it is commonly used for classification and regression analysis [18]. Regression computations in SVM are done based on structural risk minimization (SRM) principle which employs capacity control to avoid over-fitting issue. In order to simplify the implementation of SVM, in which a solution of inequality constraints is sought, Suykens et al. [19] developed a variant of SVM called least square support vector machines (LS-SVM). It reformulates the standard SVM to solve linear Karush-Kuhn-Tucker equation systems and it helps decrease computational complexity. In essence, LS-SVM transforms quadratic programming into a set of linear equations which are easier to solve. In our work, we adopted LS-SVM algorithm for regression analysis due to its known ability of handling non-linear data and the simplicity of computation.

Radial Basis Function (RBF) was selected as the kernel function for realization of implicit mapping of given input data into higher dimensional feature kernel space which results in obtaining better training and testing errors. An optimization algorithm was employed for tuning the hyper parameters in order to achieve better generalization performance and prediction accuracy. There are three optimization algorithms: simplex which can be employed for all kernels, grid search which is limited to 2-dimensional tuning parameter optimization and line search which can only be employed to linear kernel. We need to tune the hyper parameters for building LS-SVM model, so we employed simplex optimization algorithm. The mechanism of the tuning process is operated by coupled simulated annealing (CSA) which is better than multi start gradient descent optimization [20]. We considered leave one out cross validation (LOOCV) as a cost function for estimating the performance of LS-SVM model. This algorithm is controlled by the performance metric mean square error (MSE) and offers good generalization ability [21]. The performance and accuracy of LS-SVM model depends on setting of (σ^2, γ) , where σ^2 is width of kernel and γ is regularization parameter. For each pair of hyper parameters (σ^2, γ) , LOOCV method is performed on training set to estimate the prediction error and thus a robust model is obtained by selecting those optimal pair of hyper parameters. Interested reader is referred to [19] for a detailed treat of these concepts.

4.3. Benchmark Measurements

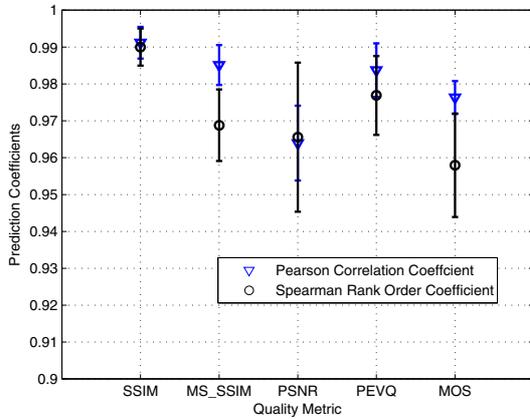
According to VQEG phase II [22], performance of an objective quality prediction model can be evaluated by three parameters which describe prediction accuracy using Pearson linear

Table 2. Comparison of LS-SVM with MLR [5] and ANN [6] based methods

VQM	SSIM			MOS			PEVQ			PSNR			MSSIM		
M/S	LSSVM	ANN	MLR												
PCC	0.9939	0.980	0.9482	0.9846	0.9748	0.9590	0.9922	0.9641	0.9133	0.9735	0.9158	0.9030	0.9850	0.9618	0.9635
SROC	0.9956	0.9794	0.9537	0.9641	0.9435	0.9129	0.9883	0.9564	0.9376	0.9985	0.8727	0.9115	0.9726	0.9396	0.8981
OR	0.050	0.133	0.200	0.1242	0.1653	0.200	0.075	0.075	0.2333	0.0550	0.1000	0.2333	0.3250	0.4667	0.2333
MSE	0.0002	0.001	0.0014	9.7877	14.30	30.21	0.0093	0.0210	0.1037	0.6340	1.6697	2.1423	0.000	0.000	0.000

Table 3. Comparison of LS-SVM with MLR [5] and ANN [6] based methods (Reduced set of features)

VQM	SSIM			MOS			PEVQ			PSNR			MSSIM		
M/S	LSSVM	ANN	MLR												
PCC	0.9888	0.9745	0.8012	0.9771	0.9420	0.6740	0.9801	0.9350	0.6117	0.9670	0.8944	0.7238	0.9702	0.9361	0.8530
SROC	0.9820	0.9644	0.8478	0.9528	0.8987	0.5843	0.9818	0.9356	0.5345	0.9409	0.8661	0.7182	0.9627	0.8846	0.8184
OR	0.100	0.100	0.6400	0.150	0.100	0.4600	0.075	0.100	0.3400	0.1500	0.150	0.800	0.2000	0.300	0.8600
MSE	0.0003	0.004	0.1914	10.50	14.288	715.62	0.0269	0.0842	1.1642	0.7328	1.5311	1732.2	0.000	0.0005	9.7636

**Fig. 3.** Prediction Accuracy and Monotonicity Coefficients Values of the Proposed Model, under 95% CI

correlation coefficient (PCC), monotonicity using Spearman rank order correlation coefficient (SROC) and consistency using outlier ratio (OR). We have used these metrics to present the results of this work.

4.4. Simulation Results

The set of the extracted parameters of 120 video sequences was divided into a training and a test set. The training data was obtained from a set of 80 videos selected randomly from the 120 videos, that makes two third of the total. The proposed LS-SVM model was trained on the training set for a variety of VQA metrics including subjective MOS. The training was done separately for each quality metric and the training data was expected to be different for each due to random selection. The proposed method has been tested for its prediction ability of subjective MOS, PEVQ, PSNR, SSIM and MSSIM. As can be seen in the regression plot given in Fig-

ure 3, the proposed model is able to predict subjective MOS with reasonably high accuracy and a similar trend is observed for other quality metrics also. Magnitude of the error bars of PCC values are below 0.005 that shows a high accuracy level in the prediction.

5. COMPARISON WITH CONTEMPORARY TECHNIQUES

Table 2 reports the statistical results achieved after using the features listed in Section 4.1. Results are shown for up to four degrees of precision in most cases. Under each quality metric column, values of the considered performance measures are presented for the three methods. For all of the quality metrics, LS-SVM has outperformed the ANN and MLR methods in all the considered performance measures. Specifically, SSIM and MSSIM have been predicted with comparatively high accuracy, monotonicity and consistency. This result is in-line with findings related to structure based quality metrics for their ability of estimating the visual quality. That explains the widespread adaptation of these metrics for a variety of image processing applications such as image restoration, contrast enhancement, watermarking, and rate-distortion optimization in standard video compression [23]. Performance results on subjective MOS are also promising except relatively high outlier ratio for all three methods. That might occur due to relatively high variability in the subjective scores, which is inherent in subjective assessment. It was observed that certain features can be omitted from the prediction model without any significant loss in performance. A discussion on the rationale of deletion of features was provided in [6]. Consequently, feature number 2, 3, 4, 6, 16, 17 were deleted to form a reduced set of features. Table 3 provided a comparative performance of the proposed method when tested on the reduced set. The simplification of the model for a lesser number of parameters is valid as indicated by the performance measure values. The performance of LS-SVM based model

remains almost the same even with the reduced set of features, in contrast to the competitor models. It is ascertained from the results that the use of support vector machine is advantageous in learning the bitstream based features of an encoded video for its quality prediction.

6. CONCLUSION

We have presented a model of no-reference video quality estimation based on a well-known machine learning method known as LS-SVM. Selection of the quality-relevant features was done based on the rationale of encoding fidelity, structural information of contents, motion information, coding distortion and spatio-temporal complexity. The proposed method was tested for its prediction ability of quality metrics PSNR, PEVQ, SSIM, MSSIM and subjective MOS. Its performance was evaluated using three standard methods of performance measure and reasonably acceptable results were obtained. Moreover, a comparative analysis with two similar methods has been performed to prove its application. The contribution of this paper is two-folds; introduction of a visual quality estimation model with bitstream-based features and comparison of contemporary techniques used for regression. The LS-SVM based predictor performs marginally better than ANN based, and reasonably better than linear-regression based, predictors for accuracy, consistency, monotonicity and error in prediction. The performance of the proposed LS-SVM model remained substantially high while a reduced set of features was used for the quality prediction. This gain in performance can be attributed to the fact that SVM is less prone to overfitting and hence has better generalization capabilities. Future work in this area is to build a model that can account for the impact of network distortions on the visual quality.

7. REFERENCES

- [1] Cisco Visual Networking Index, "Global mobile data traffic forecast update, 2012-2017," *Cisco white paper*, 2013.
- [2] Stefan Winkler, "Video quality and beyond," *Proc European Signal Processing Conference*, vol. 446, no. 3, pp. 150–153, 2007.
- [3] K. Brunnstrom, D. Hands, F. Speranza, and A. Webster, "VQEG validation and ITU standardization of objective perceptual video quality metrics," *Signal Processing Magazine, IEEE*, vol. 26, no. 3, pp. 96–101, May 2009.
- [4] S. Winkler and P. Mohandas, "The evolution of video quality measurement: From PSNR to hybrid metrics," *IEEE Transactions on Broadcasting*, vol. 54, no. 3, pp. 660–668, sept. 2008.
- [5] A. Rossholm and B. Lövström, "A new low complex reference free video quality predictor," in *IEEE 10th Workshop on Multimedia Signal Processing*, Oct. 2008, pp. 765–768.
- [6] Muhammad Shahid, Andreas Rossholm, and Benny Lövström, "A reduced complexity no-reference artificial neural network based video quality predictor," in *4th International Congress on Image and Signal Processing*, Oct. 2011, vol. 1, pp. 517–521.
- [7] J.A.K. Suykens, J. De Brabanter, L. Lukas, and J. Vandewalle, "Weighted least squares support vector machines: robustness and sparse approximation," *NEUROCOMPUTING*, vol. 48, pp. 85–105, 2002.
- [8] 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*, Sept. 2011, pp. 3325–3328.
- [9] T. Shanableh, "No-reference PSNR identification of MPEG video using spectral regression and reduced model polynomial networks," *IEEE Signal Processing Letters*, vol. 17, no. 8, pp. 735–738, Aug. 2010.
- [10] T. Shanableh, "Prediction of structural similarity index of compressed video at a macroblock level," *IEEE Signal Processing Letters*, vol. 18, no. 5, pp. 335–338, May 2011.
- [11] M. Slanina, V. Ricny, and R. Forchheimer, "A novel metric for H.264/AVC no-reference quality assessment," in *6th EURASIP Conference focused on Speech and Image Processing, Multimedia Communications and Services*, Jun. 2007, pp. 114–117.
- [12] "Subjective video quality assessment methods for multimedia applications," September 1999, ITU-T, Recommendation ITU-R P910.
- [13] V. Seferidis, M. Ghanbari, and D.E. Pearson, "Forgiveness effect in subjective assessment of packet video," *Electronics Letters*, vol. 28, no. 21, pp. 2013–2014, Oct. 1992.
- [14] "ITU-R Radio communication Sector of ITU, Recommendation ITU-R BT.500-12," 2009.
- [15] Stefan Winkler, *Digital Video Quality: Vision Models and Metrics*, Wiley, 2005.
- [16] Muhammad Shahid, Amitesh Kumar Singam, Andreas Rossholm, and Benny Lovstrom, "Subjective quality assessment of H.264/AVC encoded low resolution videos," in *5th International Congress on Image and Signal Processing*, Oct. 2012, pp. 63–67.
- [17] Iain E. Richardson, *The H.264 Advanced Video Compression Standard*, John Wiley & Sons, Ltd, 2010.
- [18] Vapnik.V, *Statistical learning theory*, Wiley, 1998.
- [19] Suykens J A K and Vandewalle.J, "Least squares support vector machine classifiers," June 1999.
- [20] S. Xavier-de Souza, J.A.K. Suykens, J. Vandewalle, and D. Bolle, "Coupled simulated annealing," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 40, no. 2, pp. 320–335, Apr. 2010.
- [21] I. H.Witten and E.Frank, *Data mining: Practical machine learning tools and techniquess*, Morgan Kaufmann, third edition, 2011.
- [22] "Full reference television phase II subjective test plans," 2002, Objective Quality Model Evaluation Criteria.
- [23] D. Brunet, E.R. Vrscay, and Zhou Wang, "On the mathematical properties of the structural similarity index," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1488–1499, Apr. 2012.