

Non-Referenced Objective Streaming Video Quality Evaluation in Cellular Networks of 3rd and 4th Generation

Ivica Kostanic, Hamad Almohamedh and Fahad Alqurashi

Abstract — This paper presents a novel methodology for quality evaluation of streaming video data services. The methodology requires no reference and it predicts subjective experience of the video quality. The predictions are based on a nonlinear mapping between objective technical metrics collected by the user equipment and subjective scores given by human evaluators. The objective metrics may be taken from various levels of the protocol stack. In the current implementation, the nonlinear mapping is accomplished through a neural network.

The performance of the methodology is tested using data UDP streaming video services over LTE (4G) and HSPA (3G). The agreement between the predictions and subjective quality evaluation scores is excellent.

Index Terms—Video Quality, Streaming Video, Quality of Experience, Wireless Data, QoE in LTE and HSPA.

I. INTRODUCTION

Over past several years, data traffic over cellular networks has experienced a tremendous growth. This growth has been fueled primarily by streaming video. According to some studies [1], data collected from various mobile providers suggests that mobile video is generating 50 percent of all mobile data. One may expect video streaming to account for over 60 percent of mobile data traffic in 2018, as consumers increase the number of videos they watch and upload. Fast, uninterrupted, video experiences encourage people to increase their video usage.

As cellular networks become used primary for streaming video, the quality of this service becomes a determining factor of the user experience. Therefore, having ability to provide a high quality video streaming becomes major differentiator between cellular service providers. The first step in provisioning a quality service is the assessment methodology. In common engineering practice, the Quality of Service (QoS) of streaming video is estimated through objective metrics of the radio link quality. Metrics like Signal to Noise Ratio (SNR), Carrier to Interference Ration (CINR), Bit Error Rate (BER) or Packet Erasure Rate (PER) are used. Although, these type of metrics are related to the overall quality of the network, they are not directly translatable into user's experience. The streaming video is ultimately judged by human subjects. The relationship between objective metrics and subjective perception of the quality is not a linear one. For that reason, better methods for video quality evaluation that take into account subjective perception of the quality need to be sought.

All authors are affiliated with Florida Institute of Technology, Melbourne, FL, USA. The corresponding author is Ivica Kostanic (kostanic@ft.edu)

The research described in this document is an attempt to develop a methodology that translates objective measures of the link quality into prediction of the subjective quality of experience. Such a methodology is of a great practical interests. It provides means for a cellular network provider to assess the quality of its video streaming service in an automated and economical manner.

The outline of the paper is as follows. Section II provides description of the video quality evaluation methodologies. Section III describes proposed MVQP algorithm and discusses its components, methodology behind the algorithm and presents the results of the experimental evaluation. Finally, Section IV discusses future directions for this work.

II. METHODS FOR VIDEO QUALITY EVALUATION

There are two principle methods for evaluation of video quality: subjective and objective. Subjective methods are regarded as the most relevant and accurate. These methods involve testing of human subjects. The testing is conducted under strictly controlled experimental environment. There are several approaches that may be used for subjective evaluation and some of them are thoroughly described in [2] and related documents. Although relevant and accurate, subjective evaluating methods are not very practical. Tests that involve humans and highly controlled experimental environment are inherently difficult to administer. As a result, they are expensive and do not lend itself to easy implementation in “day to day engineering” environment.

On the other hand, objective methods for evaluation of video quality are much more practical. In general, they are based on a set of fundamental metrics that are extracted from the quality of communication link. In their approach, objective methods for assessment of video quality may or may not user reference video. In objective methods with the full reference, quality evaluation algorithms have access to both the original, undistorted video, and the video that is distorted by the communication network. By comparing the distorted video with its undistorted version, the algorithm may determine the type and the extent of the distortion. Various types of distortion may be then mapped into some estimate of video quality. Such methods are quite successful [3-5]. However, the requirement for a full reference (i.e., the original video) represents a significant limitation. This is especially the case in applications that require real time evaluation of unknown streaming content – for example, streaming video of a broadcast.

On the other end of the objective evaluation spectrum are methods that require no reference [6-10]. These methods are

much more desirable from the practical standpoint. They may be utilized in variety of scenarios where the evaluated video is not available beforehand. However, the no-reference methods are usually highly challenging from the algorithmic development and computational standpoints.

In between the two ends of the spectrum, one finds many different approaches that requires some form of partial reference. Partial references are usually in the form of some video features that are extracted from the original undistorted videos and which are available to the evaluation algorithm.

III. MOBILE VIDEO QUALITY PREDICTION (MVQP)

The video quality evaluation method proposed in this research falls within the category of objective methods with no reference. However, the principle goal of the method is to produce predictions of the subjective quality assessment. In a way, the method attempts to combine advantages of both the objective and subjective methods with an additional requirement for a low complexity. The target computing hardware for the method implementation is 3G/4G smart phone. The method is named Mobile Video Quality Prediction (MVQP).

Fig. 1 shows the hardware required for the MVQP method. Besides the cellular (or some other communication) network under test, the solution requires a Server and Mobile phone (i.e. Use Equipment – UE). The server is “off-the-shelf” PC that runs a video streaming service. The service is implemented over UDP protocol. The PC is connected to the network under test through high speed dedicated links. The links need to have sufficient speed so that they do not interfere with the measurements. In other words, the algorithm assumes that there is no distortion introduced by the connectivity between the server and the communication network under test.

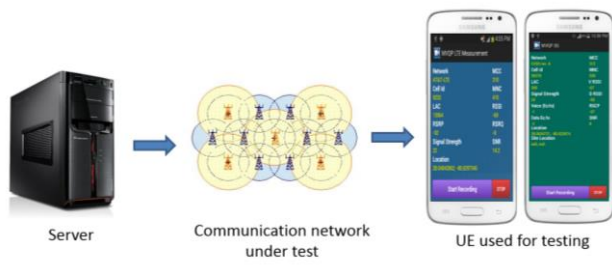


Figure 1. Hardware infrastructure required for MVQP

The computing platform implementing the quality evaluation algorithm is a 3G/4G smart phone. Besides custom application that hosts all the computations, the phone needs to have a GPS receiver. The GPS receiver provides coordinates for geo-referencing of the measurements. This is a fundamental requirement in cellular networks since the network performance is very much dependent on the geographical location.

Development of the MVQP method consists of five different stages. The stages are listed as:

1. Video database creation
2. Field measurements
3. Subjective evaluation
4. E-MOS app development
5. Experimental validation and testing

Different stages of the development are described as follows.

A. Video Database

In support of the MVQP development, a database of video recordings has been created. The videos are made publically available at: <http://research.fit.edu/wice/mvqp.php>. The database contains a wide variety of video recordings. The model of the camera used for the video collection is Sony PMW-F5 CineAlta Digital Cinema Camera 4K. This is a professional video camera and the videos are of a very high quality. More details on the database creation may be found in [11].

B. Field Measurements

The video database provides undistorted videos. To develop the video quality evaluation algorithm one needs a database of distorted videos. In principle, if the video is streamed over different communication networks, the networks may distort the video in different ways. For that reason, in this research, the distorted videos were created by either 3G or 4G cellular networks.

Generation of distorted videos is illustrated in Fig. 2.



Figure 2. Generation of distorted video database

The videos were streamed over commercial networks located in Melbourne, FL, USA. The networks are utilizing typical 3G and 4G equipment. At each measurement location, the distorted copy of the video (i.e. the video that traversed cellular network), was recorded. In parallel with the recording of the video, the measurement software records various parameters that are associated with the objective performance

of the communication link. Recorded parameters in case of 3G (HSPA) and 4G (LTE) interfaces are listed in Table 1.

Table 1. Communication link parameters measured during video streaming

	3G Performance parameter	4G Performance parameter
1	RSSI in dBm	RSSI in dBm
2	RSCP in dBm	RSRP in dBm
3	Ec/Io dB	RSRQ in dB
4	Packet Loss in %	Packet Loss in %

For both technologies, the parameters in Table 1 fall into one of three groups. The first group are parameters 1 and 2. They represent fundamental indicators of cellular network coverage. The parameter number 3 represents a fundamental indicator of interference on the radio link. Parameters 1-3 are physical layer parameters. Finally 4th parameter captures packet loss, which is an upper layer (Layer 4) parameter. More precise definition and explanation of the parameters in Table 1 may be found in [12].

One should note that, for both technologies, parameter sets specified in Table 1 represent just one of possible selections. Other possibilities may be adopted. For example, the set may be expanded to include other performance parameters that are available at the user equipment. Possible candidates include Bit Error Rate (BER), various delay parameters and radio link configuration parameters (Modulation and Coding Scheme, MIMO configuration, or rank of the channel). This particular research found that parameter set in Table 1 works well for quality evaluation of video streaming over UDP.

Using the approach illustrated in Fig 2, a substantial number of distorted videos was collected for each of the two technologies. Total number of videos was 80 in the case of 3G and 100 in the case of 4G.

C. Subjective Evaluation

In this step, each of the distorted video is associated with a Mean Opinion Score (MOS). The MOS values are obtained from controlled subjective evaluation tests. The ITU specifies several methods for subjective testing [2]. The research described in this document uses Absolute Category Rating (ACR) method. According to the ACR, at the beginning of testing, the subjects are educated briefly on the assessment process. After a brief demonstration, they are instructed to evaluate videos using standard opinion scoring in accordance with descriptions given in Table 2. It is important to note that videos are played using the same end devices that are used in the real network. In this particular case, the *screen* used for video presentation is the screen of the smart phone.

The ACR assessment method requires that the videos are presented to the testing subject in accordance to the algorithm shown in Fig. 3. To accommodate the ACR testing, a separate smart phone application was developed. The application provides the automation of the testing process, score collection and data analysis

Through the subjective testing, each distorted video is evaluated exactly 10 times. Based on the evaluation, the

MOS value associated with each distorted video is computed as the average value of individual opinions.

Table 2. Standard opinion scoring table

MOS	Quality	Impairment
5	Excellent	Imperceptible
4	Good	Perceptible but not annoying
3	Fair	Slightly annoying
2	Poor	Annoying
1	Bad	Very Annoying

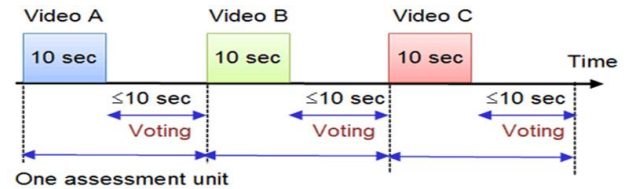


Figure 3. One assessment unit of the ACR subjective testing method

D. Development of EMOS Application

After the subjective tests, two sets of data are available for each distorted video. On one hand, there is a set of objective metrics listed in Table 1. These metrics are calculated by the UE since they are needed by the link adaptation and mobility management associated with regular call processing. On the other hand there are MOS values obtained from the subjective ACR testing. What is needed is the mapping between the two.

This research assumes that the mapping between objective metrics and the MOS values is a nonlinear one. To approximate this non-linear mapping a neural network is used. Two neural network architectures are considered: Multi-Layer Perceptron (MLP) and Radial Basis Function (RBF) neural network. Out of the two, the RBF NN is selected due to its simplicity and ease of training. The architecture of an RBF NN is presented in Fig. 4.

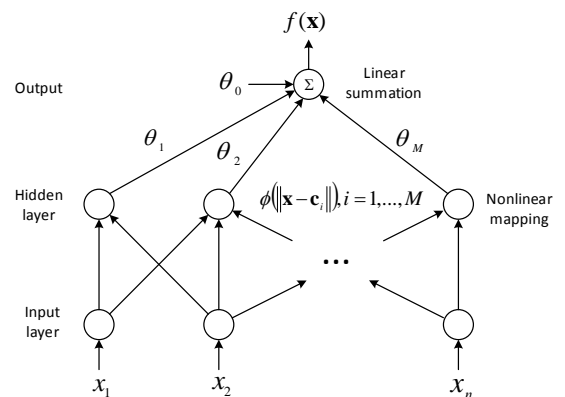


Figure 4. Architecture of the RBF NN

As seen, the network consists of three layers. The input layer accepts the objective measures of quality that are associated with distorted video. The hidden layer performs a

fixed nonlinear transformation, with adjustable parameters, so that the input space is mapped into another vector space. The output layer combines the outputs of hidden layer in a linear fashion. The output of the neural network is given as

$$f(\mathbf{x}) = \theta_0 + \sum_{i=1}^M \theta_i \phi(\|\mathbf{x} - \mathbf{c}_i\|) \quad (1)$$

where $\mathbf{x} \in \mathfrak{R}^{n \times 1}$ is the vector of input parameters, $\phi(\bullet): \mathfrak{R}^+ \rightarrow \mathfrak{R}$ is the kernel function, $\|\bullet\|$ denotes Euclidian norm, θ_i for $0 \leq i \leq M$ are network weights, $\mathbf{c}_i \in \mathfrak{R}^{n \times 1}$ for $0 \leq i \leq M$ are the centers and M is the number of centers.

The nonlinear mapping performed by the network is determined by the values of weights θ_i and centers \mathbf{c}_i . The centers are usually selected so that they sample adequately the space spanned by input parameters, while the value of weights are determined through the RBF NN training process. More on the network training process may be found in [13, 14].

After the network is trained, the output prediction from a given input vector is performed in accordance with (1).

Therefore, after training, all that one needs to code is the expression in (1) with a determined set of weights. This makes practical implementation of the algorithm very effective from the complexity standpoint and therefore, quite suitable for the computing platform that consist of only 3G/4G smart phone.

E. Experimental Verification

The set of distorted videos with the associated objective metrics and MOS scores is divided into two subsets. One subset is used for training of the neural network and the other one is used for testing of the network prediction accuracy. The results obtained for the two subsets in case of 3G data are presented in Figs 5 and 6. For the 4G data the results are given in Figs 7 and 8. As one may see, the figures show a good agreement between predictions of the MVQP algorithm and MOS values from subjective testing. The average error between prediction and measurements is essentially zero, with standard deviation of 0.24 and 0.3 in 3G, and 0.25 and 0.4 in 4G for the training and testing data sets.

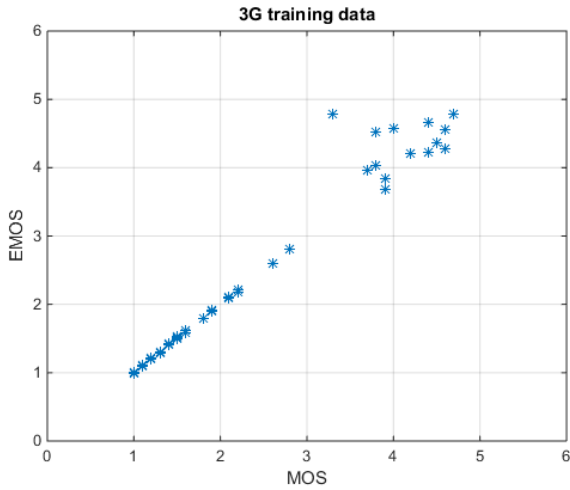


Figure 5. Performance of the MVQP algorithm for 3G training data set

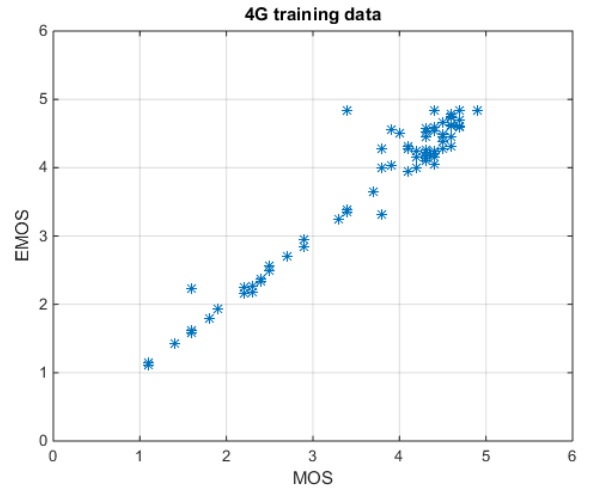


Figure 7. Performance of the MVQP algorithm for 4G training data set

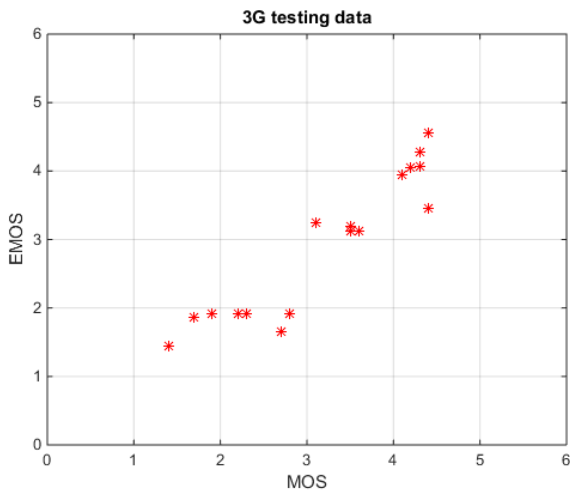


Figure 6. Performance of the MVQP algorithm for 3G testing data set

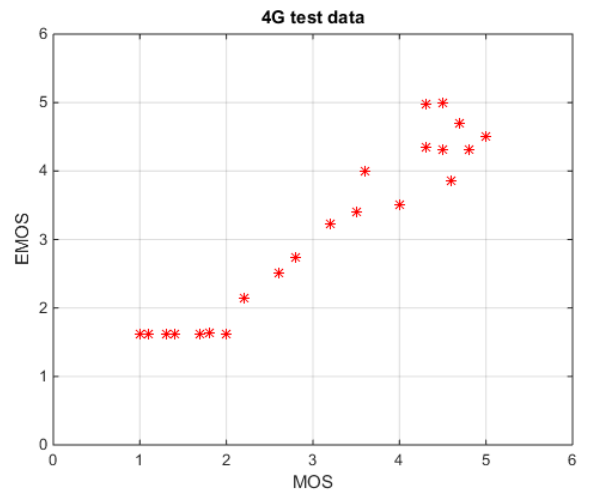


Figure 8. Performance of the MVQP algorithm for 4G testing data set

IV. SUMMARY AND FUTURE RESEARCH

The research presented in this paper demonstrated feasibility for subjective video quality prediction from objective measurements and through a nonlinear mapping that may be performed using RBF NN. The research resulted in a mobile phone application that performs video quality assessment in real time. Such an application provides a tangible benefit in a day to day management of the network performance.

The research may be extended in two directions. First, other types of data streaming over 3G and 4G cellular networks could be considered. In particular, video streaming over TCP/IP and VoIP would be logical next steps. Additionally, even though the approach was developed and tested using 3G and 4G cellular, it is not specific to any particular air interface. Therefore, it should be relatively easy to extend the research to WiFi, WiMAX, or any other wireless interface that supports streaming data services.

V. ACKNOWLEDGMENT

The authors would like to acknowledge support of WiCE at Florida Tech and PCTEL.

VI. REFERENCES

- [1] – Ericsson, Ericsson Mobility Report, June 2014, free download from: <http://www.ericsson.com/res/docs/2014/ericsson-mobility-report-june-2014.pdf>.
- [2] ITU-R Recommendation BT.500-11, "Methodology for the subjective assessment of the quality of television pictures," tech. rep., International Telecommunications Union, 2000.
- [3] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P., "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing* 13(4), 600–612, 2004
- [4] Sheikh, H. R. and Bovik, A. C., "An evaluation of recent full reference image quality assessment algorithms," *IEEE Transactions on Image Processing* 15, 3440–3451, November 2006.
- [5] Seshadrinathan, K. and Bovik, A. C., "Unifying analysis of full reference image quality assessment," in *IEEE Intl. Conf. on Image Proc.*, 2008.
- [6] M. Montenovo, A. Perot, M. Carli, P. Cicchetti, and A. Neri, "Objective quality evaluation of video services," in *Second International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, Jan. 2006.
- [7] S. Pechard, D. Barba, and P. Le Callet, "Video quality model ' based on a spatio-temporal features extractions for H.264- coded HDTV sequences," in *Proc. Picture Coding Symposium*, Nov. 2007.
- [8] M. Ries, O. Nemethova, and M. Rupp, "Performance evaluation of video quality estimators," in *Proc. European Signal Processing Conference EUSIPCO*, Sep. 2007, pp. 159–163.
- [9] Y. Fu-Zheng, W. Xin-Dai, C. Yi-Lin, and W. Shuai, "A noreference video quality assessment method based on digital watermark," in *Proc. 14th IEEE Personal, Indoor and Mobile Radio Communications 2003*, vol. 3, Sep. 2003, pp. 2707– 2710.
- [10] A. Eden, "No-reference estimation of the coding PSNR for H.264-coded sequences," *IEEE Trans. Consum. Electron.*, vol. 53, no. 2, pp. 667–674, May 2007.
- [11] Fahad Al Qurashi, Hamad Almohamedh, Ivica Kostanic "RAW Video Database of Mobile Video Quality Prediction (MVQP)," *International Conference on Image Processing, Computer Vision, and Pattern Recognition*, July, 2014.
- [12] E. Dahlman, S. Parkvall, J. Skold and P. Berming, *3G Evolution, HSPA and LTE for Mobile Broadband*, 2nd ed., Academic Press, 2008.
- [13] F. M. Ham and I. Kostanic, *Principles of Neuro-computing for Science and Engineering*, McGraw Hill, 2001.
- [14] S. Chen, C. F. Cowan, and P. M. Grant, "Orthogonal least squares learning algorithm for radial basis function networks," *IEEE Trans. on Neural Networks*, vol. 2, pp. 302- 309, 1991