Intelligent QoE Analysis Using Machine Learning

Alexander V. Ivchenko
Department of Radio Engineering and Cybernetics
Moscow Institute of Physics and Technology
Dolgoprudny, Russia
ivchenko.a.v@phystech.edu

Abstract— **The work is devoted to the analysis of the quality of the transmitted and reproduced multimedia content. End-user quality assessment is based on subjective perception, but it determines whether a person will use a particular product or service. At the same time, most users of multimedia services are not in their own networks of providers with the ability to provide and monitor the required quality of QoS (for example, IPTV) service, but in networks of other operators where information about the deterioration of the quality of the provided service comes post factum in the form of user complaints OTT services). Until now, there are no uniform methods for assessing the quality of service (QoS) and quality of perception (QoE). The proposed methodology is based on the methods of machine learning and neural networks and allows in an automatic mode to assess the quality of perception based on objective parameters. Methods of objectifying and parametrizing the evaluation are also considered.**

Keywords— machine learning; neural networks; QoE; QoS; metrology; distortion; automation

I. INTRODUCTION

Analysis of the quality of transmitted and reproduced multimedia content is a necessary condition for the effective functioning of digital information transmission systems. Quality assessment, based on subjective perception of a person, is called quality of perception - QoE (Quality of Experience). In turn, perceptions are influenced by objective indicators that determine the evaluation, called quality of service - QoS (Quality of Service) [1].

When using digital networks, data loss consists of information losses when the original signal is compressed and from the cases of discontinuity in the flow of data packets. Degradation is characterized by a gradual decrease in the sharpness of the picture, an increase in the likelihood of the appearance of mosaic artifacts, and, with large losses in the transmission path, "freezing" the video into a freeze frame. In this case, the dependence of the degree of degradation on the magnitude of the losses is essentially nonlinear, accelerating from an almost imperceptible increase with small losses to failure and complete cessation of reproduction in the worst cases.

As a result, there are two problems specific to digital multimedia broadcasting. First, because of the non-linearity of the correspondences, there is a large discrepancy between the measured quality of the QoS packet transmission parameters by the objective quality of any particular functional characteristic of the network and the actual "perceived" subscribers of the quality of the QoE service consumed. Secondly, there are many different parameters of degradation, in connection with which the problem arises of deducing a scalar estimate of the quality of services from the vector quantity of deviations.

A. State of the art

There are a number of works devoted to the analysis of subjective perception. In 2004, the International Telecommunication Union conducted a study [3] of methods and approaches for assessing quality. As a result, it was proved that the most progressive objective methods yield statistically equivalent results on large samples. The main reason for the equivalence was the dependence of expert judgment directly on the content.

S. Ke in 2012 [4] described a method of notifying the inconsistency of the state of the bandwidth on the user's device and the current bandwidth requirement. Notifications occur in semi-automatic mode via SMS, MMS, e-mail and the portal of the service provider. The QoE is evaluated by user questionnaires. A significant drawback of this solution is the multiple interaction with end users for a single evaluation of the QoE parameter, which virtually eliminates the possibility of building an effective automatic QoE monitoring system on a large scale.

Close to our study is the work of Z. Kahn [7], which proposes a method of measuring QoE, based on a neural network using the particle-particle optimization (PSO) method. The technique is designed for use in mobile applications. During the operation, the application polls the user about the correctness of the evaluation and determines the amount of error, in the case of exceeding the allowable limit, the algorithm for adaptation of weights is applied. In the course of the work, the frame structure is compared with the reference frames (without distortion), which is a significant obstacle for mass application, for example, to assess the quality of streaming video. The technique also operates with a fixed set of analyzed parameters.

Close work is the technique of U-vMOS [8]. Based on the "Content" group of parameters (resolution, frame rate and color gamut), "Upload" (content search time, start time of loading and browsing, channel switching time) and "Playback" (block image distortion, frequency and duration of playback interruptions) derived a formula by which the value of U-

978-1-5386-5824-6/18/$31.00 ©2018 IEEE

vMOS is calculated - separately for broadcast and for video on demand (1).

$$U - vMOS = 1 + (sQuality - 1) \cdot \cdot (\frac{\alpha(sUnteraction - 1) + \beta(sView - 1)}{4(\alpha + \beta)}) \quad (1)$$

The technique groups content into classes such as "active sport", "comedy", etc., defines a number of scenarios of user interaction with services and thresholds of characteristic values. The coefficients α and β depend on the content class and the action.

This technique has several disadvantages. U-vMOS operates with a fixed set of parameters for analysis. If one of them is missing or measured with unacceptable accuracy, then the technique does not work. The breakdown of content into classes and user states on the script is rough and introduces, strictly speaking, an error of an arbitrary magnitude - for example, in a football match there are breaks in which multimedia content qualitatively changes. The technique can evaluate only those distortions that arose due to degradation after the passage of the communication channel. If distortions arose before this point - they will not be taken into account. This applies to spot glare, the effect of "Newton Circles", post-processing distortions, etc.

B. *Statement of the problem*

The task we have set is to create an automatic method for analyzing subjective perception based on objective parameters. The solution should be:

- modular (adding or removing parameters from the analysis);
- universal (application in systems with different architectures and equipment);
- applicable on stationary and mobile devices;
- able to use the scale of the national IP-broadcasting network.

Also, the solution should not require a standard – undistorted content for analysis.

II. Methodology

A. *Finding the QoE function*

The general form of the function QoE from some objective indicator V is shown in Fig. 1 and is described by the formula (2), where b1, b2, b3 are scaling factors.

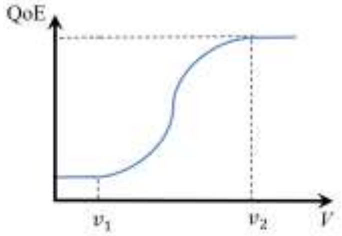

Fig. 1. 1. Dependence of QoE on the integral exponent V

$$QoE = b_1 / \left[1 + exp\left(b_2 \left(V - b_3\right)\right)\right] \quad (2)$$

At values less than v1, there is no deterioration in the perception quality, similarly, at a level of more than v2, the user does not notice any improvement in quality. The exponent V can be considered as, $V = \sum_{i=1}^{n} w_i \left(W, P\right) \psi_i$, W = {w_i}, P = {p_j}, w_i – calculated weighting coefficients, ψ_i – значения объективных показателей, p_j – values of objective indicators, pj - characteristic features of content. The QoE function and the weighting coefficients are nonlinear, so we have the problem of multidimensional nonlinear regression.

To solve this problem, a multi-layer neural network was written based on the open software library for machine learning Google TensorFlow [19]. To collect data on the operation of the network, the Wireshark traffic analyzer [5] was used. A number of experiments with shells for the core led to the selection of the Keras library [20]. Preliminary experiments on test data show up to 82% accuracy

B. *Objectification of parameters*

Analyzed indicators are divided into groups: network and equipment; characteristic features of content; expert evaluation. Among the first group, it is proposed to monitor [11–15]: packet loss in the network; traffic speed; delay; jitter; the speed of erroneous sound packets; screen resolution; fractality of audio and video.

Fig. 2. Shots of the series "Castle" and "Kamenskaya."

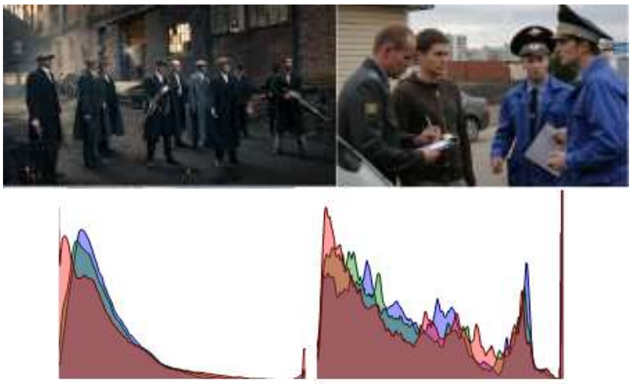

Fig. 3. Personnel and corresponding histograms of the films "Peaky blinders" and "Glukhar"

Among the characteristic features of video imagery it is suggested to take into account [16–18]:

- activity of the movement;
- the frequency of changing the camera angle and changing the scene;

- distortions of various kinds;
- depth of field;
- saturation.

Motion activity is estimated based on the number and value of the length norm of the motion vectors available from the decoder. The analysis of the change of angles is caused by the decrease of QoE by long static scenes. The frequency of changing the angle depends on the amount of motion in the frame. To track this parameter, we use the algorithm for plot change analysis, proposed by A.V. Dvorkovich and V.P. Dvorkovich, who consists in analyzing the change in the histogram of the distribution of brightness in neighboring frames [10].

Depth of field determines the area of the frame on which the viewer's attention is concentrated (Figure 2). On the right frame there is a huge amount of fairly clear uninformative details, whereas on the frame to the left, attention is focused solely on the actor. The percentage of object contours is used as a numeric indicator. To analyze the saturation, histograms of the frames are examined (Fig. 3).

C. Analysis of distortions

To learn the distortion analysis module, the TID2013 database [21] was chosen, containing 24 types of distortion. For the initial analysis, the following were selected:

- Gaussian noise;
- spatially correlated noise;
- impulse noise;
- local block-wise distortions.

A convolutional neural network based on the TensorFlow core and the Keras shell was written. Processing of data arrays is implemented on the basis of the Numpy library [12]. The work directly with the images is done using the OpenCV library [23]. To improve the accuracy of the neural network and reduce the risk of retraining, the Dropout algorithm [6] is applied. With the resolution of 512x384 pixel images, 91% accuracy and 94% completeness in the classification were achieved. A full-color analysis was used, since the analysis attempt in gray gradations lowered the accuracy to 40%.

D. Methods of application

The work introduces a natural assumption that within a single segment of the network, geographically limited by the geography of distribution of subscribers, a single set of technologies and equipment is used. This allows you to apply the QoE estimation function, computed once, for the entire network segment.

The first stage is the tuning step. The operator selects a number of parameters that can be analyzed using the means available to him. In this network, a utility is launched that collects network statistics and analyzes the characteristic features of the content. At the same point, content analysis is performed using the MOS technique [9].

At the second stage in the laboratory, the analysis of the obtained data is carried out, including the analysis of the set of indicators used. As a result, the operator also receives the accuracy value of the QoE estimate, which can be obtained with this sets of parameters and its subsets. It is specified whether there are dependencies between the indicators, is it possible to exclude some of the data from the analysis.

Fig. 4. Example of distortion from TID2013 from left to right: Gaussian noise, spatially correlated noise, impulse noise, block-wise distortion.

At the second stage, the methodology can conclude that the data is insufficient for the desired accuracy of the estimate. In this case, it is required to return to the first stage and include additional data in the analysis, based on the recommendations of the work.

Thus, after the second stage there is a sufficient set of parameters and functions of weight coefficients. At the third stage - the stage of application, it is enough to collect statistics and characteristic features of the content (at the control point or on the end user device) in order to obtain estimation of QoE according to the formula with the weight coefficients obtained in the previous stage (Fig. 5).

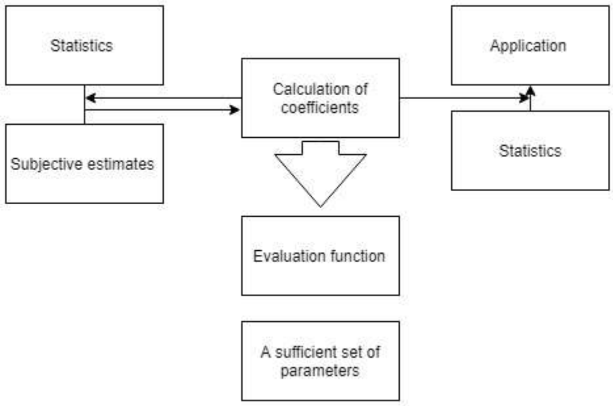

Fig. 5. Methods of application

III. Results

A QoE analysis methodology has been developed, its implementation and preliminary laboratory tests are carried out. The recovery of the QoE function shows 82% accuracy, but it requires a lot of expert evaluation to adjust the weights of the neural network and further work.

For analyzer's training, from 100 to 600 images were used, evenly distributed over 4 classes of distortion. As the training sample, 80% of the collected material was used and, accordingly, 20% was the test sample. In this case, the peculiarity of the database TID2013 is 5 degrees of distortion of the same frame. With such a small sample, when filing the same frames, the classification of distortions dropped to 40% of the accuracy. Also, work is under way to analyze the degree of distortion, as well as the case of several types of distortion in one frame.

CONCLUSIONS

The obtained results demonstrate the possibility of applying the proposed approach of automatic estimation of QoE perception quality using convolutional neural networks and machine learning methods. Further work will be aimed at increasing the base of statistical and expert assessments, as well as a database of distortions, optimizing the network, taking into account the growth in the number of images.

Additional work will be conducted on processing and audio information.

ACKNOWLEDGMENT

Author thanks Dvorkovich Alexander Viktorovich for the consultations and the materials provided.

REFERENCES

[1] Recommendation ITU-T G.1080. Quality of experience requirements for IPTV services. ITU-T. 12/2008. 44 p.

[2] Ivchenko A.V. Intellectual method for assessing the quality of perception. Dvorkovich AV. Moscow. Proceedings of the Conference (reports); international (ISBN 978-5-905278-29-7). All-Russian (UDC 621.391: 004 (082). BBK: 32.511. (In Russian).

[3] ITU-T Tutorial. Objective perceptual assessment of video quality: Full reference television. ITU-T. 2004. 218 p.

[4] Pat. RF No. 2582573. Sh. Ke. Method of notifying the user's bandwidth. Publ. 21.12.12. (In Russian).

[5] WireShark.Free and Open Source packet analyzer. Version 2.6.0. URL: https://wireshark.org

[6] Srivastava N., Hinton G., Krizhevsky A., Sutskever I., Salakhutdinov R.. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. Journal of Machine Learning Research. 2014. No. 15. Pp. 1929-1958.

[7] Pat. CN 103179592A. Kan Z. Z., Yuan H., Wenbo W. Method for measuring QoE. 2015.

[8] Kunstler J., Dobardziev A. Wireline Video QoE Using U-vMOS. URL: https://ovum.informa.com/resources/product-content/2016/11/11/15/09/analyst-white-paper-wireline-video-qoe-using-u-vmos

[9] Recommendation ITU-T P.10/G.100. Vocabulary for performance and quality of service. ITU-T. 07/2006. 42 p.

[10] Dvorkovich V.P., Dvorkovich A.V. Digital video information systems (theory and practice). Moscow. Technosphere. 2012. 1009 p. (In Russian).

[11] Abry P., Flandrin P., Taqqu M.S., Veitch D.. Wavelets for the analysis, estimation and synthesis of scaling data. Self-Similar Traffic and Performance Evaluation, edited by K. Park and W. Willinger. John Wiley & Sons. 2000. Pp. 39-88.

[12] Shelukhin O.I., Osin A.V., Akhmetshin R.R. Evaluation of the self-similarity of speech traffic by the wavelet method with automatic scaling limits determination. Electrical and information complexes and systems. 2007. No. 1. Pp. 15-20. (In Russian).

[13] OpenCV. Open Source Computer Vision Library. Version 3.4.0. URL: https://opencv.org/ (In Russian).

[14] Recommendation ITU-R BT.601-7. Studio encoding parameters of digital television for standard 4:3 and wide-screen 16:9 aspect ratios. ITU-R. 03/2011. 20 p.

[15] Reiter U., Brunnstrom K., De Moor K., Larabi M.-Ch., Pereira M., Pinheiro A., You J., Zgank A. Factors Influencing Quality of Experience. Quality of Experience. Advanced Concepts, Applications and Methods, edited by S. Moller and A. Raake. Springer. T-Labs Series in Telecommunication Services. 2014. Pp. 55-72.

[16] Mariner xVu. URL: http://www.marinerxvu.com/

[17] Datamertie GX. URL: http://www.ip-label.co.uk/products/datametrie-global-experience/

[18] Hughes Sysyique. URL: https://hsc.com/campaigns/better-QoE-for-effective-OTT-streaming-services

[19] TensorFlow. An open source machine learning framework for everyone. Google. Version 1.6.0. URL: https://tensorflow.org/

[20] Chollet F. Google, Microsoft, others. Keras. GitHub repository. 2015. URL: https://keras.io

[21] Ponomarenko N., Jin L., Ieremeiev O., Lukin V., Egiazarian K., Astola J., Vozel B., Chehdi K., Carli M., Battisti F., Jay Kuo C.-C. Image database TID2013. Peculiarities, results and perspectives. Signal Processing: Image Communication. Vl. 30. Jan. 2015. Pp. 57-77.

[22] NumPy. Package for scientific computing with Python. Version 1.14.2. URL: http://numpy.org/