sciendo

Transport and Telecommunication, 2020, volume 21, no. 3, 181–190
Transport and Telecommunication Institute, Lomonosova 1, Riga, LV-1019, Latvia
DOI 10.2478/ttj-2020-0014

QOS TO QOE MAPPING FUNCTION FOR IPTV QUALITY ASSESSEMENT BASED ON KOHONEN MAP: A PILOT STUDY

Jaroslav Frnda[1], Marek Durica[1], Mihail Savrasovs[2], Philippe Fournier-Viger[3], Jerry Chun-Wei Lin[4]

[1]Department of Quantitative Methods and Economic Informatics,
Faculty of Operation and Economics of Transport and Communications, University of Zilina
010 26 Zilina, Slovakia
jaroslav.frnda@fpedas.uniza.sk
marek.durica@fpedas.uniza.sk
[2]Transport and Telecommunication Institute
Lomonosova 1, Riga, LV-1019, Latvia
savrasovs.M@tsi.lv
[3]School of Natural Sciences and Humanities, Harbin Institute of Technology
Shenzhen, 518055, China
philfv8@yahoo.com
[4]Department of Computing, Mathematics, and Physics, Western Norway University of Applied Sciences
Bergen 5063, Norway
jerrylin@ieee.org

This paper deals with an analysis of Kohonen map usage possibility for real-time evaluation of end-user video quality perception. The Quality of Service framework (QoS) describes how the network impairments (network utilization or packet loss) influence the picture quality, but it does not reflect precisely on customer subjective perceived quality of received video stream. There are several objective video assessment metrics based on mathematical models trying to simulate human visual system but each of them has its own evaluation scale. This causes a serious problem for service providers to identify a critical point when intervention into the network behaviour is needed. On the other hand, subjective tests (Quality of Experience concept) are time-consuming and costly and of course, cannot be performed in real-time. Therefore, we proposed a mapping function able to predict subjective end-user quality perception based on the situation in a network, video stream features and results obtained from the objective video assessment method.

Keywords: IPTV; Kohonen map; Mapping function; QoE; QoS

1. Introduction and Motivation

Video content has become a dominant part of all data traffic sent via packet networks based on IP protocol. Network convergence and digitalization happening over the last two decades enabled encoding and transmission of video via network infrastructure originally intended for data transfer (e-mails, file transfers, web service, etc.). Just as voice service, video is also time-sensitive service and because of that the transmission protocol TCP had to be replaced by a protocol without acknowledgment procedure of delivered packets. The UDP protocol appeared to be a solution (popular video-on-demand services such as YouTube or Netflix use TCP; but that is actually not a real-time broadcasting); however, missing data at the received side were not retransmitted, which could generate artifacts during video reconstruction and playback.

If IPTV (Internet Protocol Television) providers want to be competitive with classic terrestrial video broadcasting service, they must monitor and evaluate the service quality they offer. Given the fact that subjective tests require the involvement of a number of viewers, the picture quality is evaluated by objective video quality metrics. For this purpose, common metrics are used such as PSNR (Peak Signal to Noise Ratio), SSIM (Structural Similarity Index) or VQM (Video Quality Model). The first one is older but easy to compute metric, the second and third ones have a better correlation with human perception (Sevcik *et al.*, 2014; Frnda *et al.*, 2019).

Not only IPTV service is based on the technology of real-time video streaming. Gradually, improvements in the IT sector have extended the capacity of storage systems or network bandwidth. The concept of a smart city includes intelligent transport systems. These systems require the IP based video surveillance capable of monitoring and recording video over an IP network. The high-profile installations involving high image quality and frame rate for vehicle counting and classification or licence plate recognition. These automated real-time (or rather almost real-time) systems identify a number of traffic rule violations and incidents and have to be in operation 24/7. Due to this fact, these systems also rely on UDP protocol and may be affected by losses during the transmission process. These disruptions causing the video picture to become jerky, jittery or freeze for a specific amount of time while the codec (compression algorithm) tries to estimate missing picture information (Loktev *et al.*, 2019).

Typically, the user-perceived quality is best expressed by Mean Opinion Score (MOS), ranging from 1 to 5 (5 is the best). Each of the mentioned objective metrics uses various scales.

The key objective and motivation of this work is to introduce a method capable of tying the findings of the objective and subjective method. There are several studies that suggest mapping functions for translation of objective results into the subjective point of view.

The first attempt to utilize a neural network for this task is called PSQA (Pseudo Subjective Quality Assessment) (Mohamed and Rubino, 2002), and the authors use objective qualitative parameters like packet loss or bitrate as an input to neural network training with the objective score to calculate a subjective rating. Although only old codec MPEG 2 and low resolution (352 x 288) along with small-sized test sample were used, the paper served as a conceptual framework for this article.

Researchers (Valderrama and Gómez, 2016) managed to select different attributes for the neural network training process, such as various lengths of the group of pictures (GOP), prioritization policies (BestEffort and DiffServ), or bottlenecks in the test network topology. Although Pearson´s coefficient achieved a level of more than 0.9 in their paper, only low-resolution pictures (740 x 480) and the limited number of packet loss scenarios (1%, 5%, and 10%) were used. D. Mocanu (Mocanu *et al.*, 2015) has summarized the applicability of several tools for machine learning, and his test has shown a higher correlation coefficient in the results achieved by neural networks. The research team of J. Søgaard (Søgaard *et al.*, 2015) proposed a regression function for video quality calculation with a correlation coefficient varying from 0.7 to 0.9 based on the video content of dynamic and static scenes.

The improvement of the SSIM method can be considered as the main benefit of the next work (Loh and Bong, 2018). The research team led by Woei Tan Loh incorporated the concept of spatial and temporal video quality into the SSIM index. Although they achieved improvement in accuracy, the computational time of the proposed method was 50% higher in comparison to the "classic" SSIM metric.

Newer video codec H.265, along with Pearson's coefficient slightly over 0.92 but without UHD (Ultra High Definition) resolution, was used as an effort to synthesize a regression function for subjective score prediction in papers (Cheng *et al*., 2017) and (Anegekuh *et al*., 2015).

Authors (Mustafa and Hameed, 2019) created several testing scenarios (different packet loss, 12 types of scene, various bitrates) but they used only CIF and 4CIF resolution with H.264 codec. They decided to apply new metric NoDFI on several machine learning tools (neural network, naïve bayes and decision tree) and obtained prediction accuracy ranging from 0.86 to 0.88. Article (Akhtar *et al*., 2019) provides a summary of the latest trends in this research field and points out several approaches, e.g. difference between linear and non-linear relationship among the QoE and QoS parameters. Authors in (Bampis *et al*., 2018) included spatial and temporal information as an additional input into the prediction process, and they received accuracy rate approximately 0.9 in the term of Pearson correlation coefficient. Paper produced by (Gu *et al*., 2019) described no-reference prediction model features such as contrast, sharpness or brightness. The team tested their metric on 6 video databases and gained accuracy ranging from 0.73 to 0.9 based on the selected test database. Advantages and disadvantages of Back propagation neural network usage for video quality estimation were introduced in (Yuan and Wang, 2019). Several scenarios were tested and improvement schemes of neural network were suggested by the authors. They provided a selection of important video sequence characteristics and summary of their impact on quality prediction. The created model worked with accuracy level about 0.91.

Collective of authors of this proposed article also contributed to this research. We introduced a hybrid method for IPTV quality assessment based on a backpropagation neural network (Frnda *et al.*, 2019). Our proposed application takes scene dynamic range, packet loss, bitrate and resolution into consideration as input features followed by two codecs (H.264 and H.265/HEVC). While writing this article, our model can estimate the subjective perception of video quality for two codecs and three resolutions, which is something that none of the previously mentioned models offer.

All the above-mentioned works attempted to design a computational model able to extrapolate the subjective perception of video quality from the set of potential feature candidates.

We decided to use Kohonen map for its unsupervised learning and clustering. It is a type of neural network that comes with generalization ability to predict the data which it has not trained on. In fact, network administrators do not need to know the exact MOS value as other models do. They need to know if the delivered video quality is good or if it is getting worse. Our developed system has kept all the benefits of our above-mentioned hybrid method, namely fast service quality prediction for both the most commonly used video codecs nowadays.

We realize that some of the published papers gained a better accuracy level of prediction but none of them used Kohonen map for quality estimation, therefore this paper serves as a pilot study and investigates the suitability of unsupervised learning and clustering techniques for future research.

2. Methodology

In our previous paper, we created a distorted video database based on subjective and objective picture quality evaluation. We chose the SSIM method as an objective assessment method representative due to its good correlation with subjective perception, as well as due to its ability to use a suitable scale range for SOM modelling. It is a full reference metric, i.e. the image quality prediction is based on an initial distortion-free image as a reference. The analysis results from the measurements of contrast and luminance altogether with the structural similarities of the referenced video sequence are shown in Figure 1. Throughout the video quality investigation process, the scores range from 0 – no similarities to 1 – an identical sequence as the reference. The formula for the final combination represents the similarity measure of a test signal *y* with a reference signal *x*, and it is defined as follows (Sevcik *et al*., 2014; Frnda *et al*., 2019):

$$SSIM(x,y) = [l(x,y)]^{\alpha}[c(x,y)]^{\beta}[s(x,y)]^{\gamma}, \quad (1)$$

where $l(x,y)$ represents luminance component, $c(x,y)$ is contrast component and $s(x,y)$ describes structure component of SSIM formula. Parameters $\alpha > 0$, $\beta > 0$, $\gamma > 0$ point out to weighted combination of above-mentioned components.

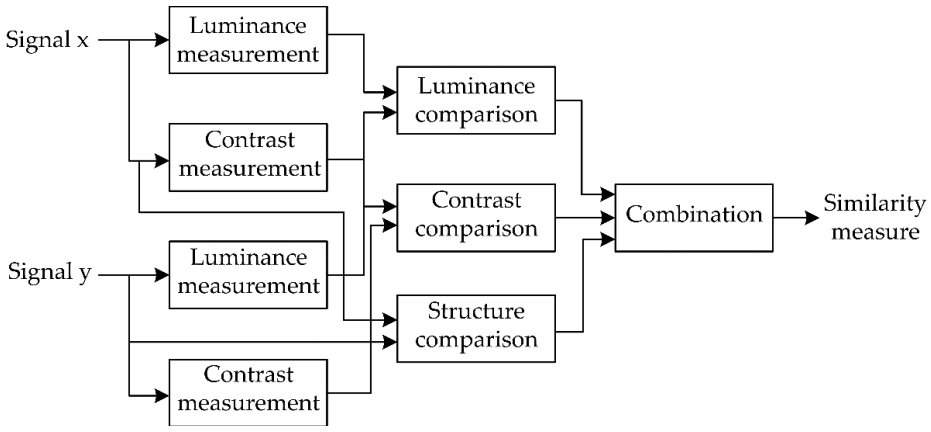
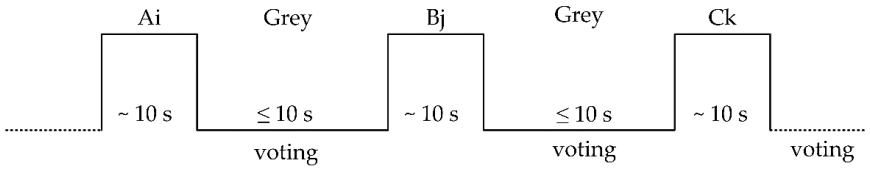


Figure 1. SSIM metric architecture

The absolute category rating (ACR) is a category judgement method standardized by ITU-T. The video sequences are presented individually, i.e. one at a time. Each presentation continues with the video quality evaluation done by the invited observers, and the standard MOS scale (as described in Table 1) is used for the assessment. As depicted in Figure 2, voting limit is set to ten seconds. ACR method reflects on the real situation when customers also cannot compare delivered video stream with reference one provided by content owner (typically TV channel). The testing room (lighting conditions, viewing distance) with a TV screen (24" Dell P2415Q UHD) was prepared to meet the recommendations stated in (ITU-T P.913, 2016). 60 observers participated at the age ranging from 18 to 35 years and with a male domination of 38:22. Observers had a little break every half hour, and total session duration took two hours (ITU-T P.913, 2016).

The scene character, from sport broadcastings through action movies to TV news, can be described by the Time and Spatial information (TI/SI). The recommendation (ITU-T P.910, 2008) defines several categories of video content based on these two parameters. Various scenes in terms of the SI and TI parameters were released by the Shanghai Jiao Tong University's research team (Song *et al*., 2013).

These video sequences last only 10 seconds and contain 300 frames, i.e. the frame rate is 30. The descriptive characteristics for the scenes mentioned above are displayed in Figure 3. For more information about preparation of the testing videosequences please see our paper (Frnda *et al*., 2019).



Ai – Sequence A under test condition i
Bj – Sequence B under test condition j
Ck – Sequence C under test condition k

Figure 2. ACR testing procedure

a) b)

c) d)

*Figure 3.*Types of the scene (clockwise) – a) a very dynamic scene (fast camera moving); b) night scene (static people sitting close to a bonfire, flames are changing their shape very dynamically); c) static scene (static shooting, low motion of vehicles); d) sport scene (marathon)

Table 1. Five-degree scale of MOS

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

TV broadcast bitrate typically ranges from 10 to 15 Mbps, which is based on technical limitations of broadcast transmission of digital terrestrial television. The TV signal is divided into 7 or 8 MHz wide channels. By using 64-QAM (Quadrature amplitude modulation) digital modulation, maximal transmission signal bandwidth is a little above 31 Mbps (8 MHz channel wide). According to this restriction, one multiplex (several digital signals are joint into one signal via shared transmission channel) consists of three 10 Mbps or two 15 Mbps (premium quality) TV stations.

Our previously mentioned research pointed on the significant video sequence features suitable to be used as an input data vector for training. At first, the scene type is a significant parameter that describes how codec masking techniques can approximate the absent picture information. In static shooting with monochromatic background, an algorithm can effectively forecast lost part, but in a

dynamic scene where the frames are redrawn very often, impairments caused by packet loss are more noticeable. Both codecs (H.264 and H. 265 respectively) have a dissimilar compression ratio, therefore data loss impacts on codec with higher compression more (H.265 has about double compression in comparison to its predecessor H.264). Bitrate characterises the number of bits that are processed in a unit of time. Higher resolution involves higher bitrate to secure appreciable picture quality, but high bitrate (e.g. 15 Mbps) could be useless if small resolution is set – which does not increase the perceptual experience. In this case, packet loss effect is minimised due to loss of redundant information.

Despite our expectation, after performing the tests we have no evidence to support the importance of screen resolution. Most packets belong to I-frames (they comprise an entire image), thus the most part of dropped packets affects the I-frame integrity. In case of stronger compression of the I-frame, the quality loss can be more noticeable at high resolution when compared to low resolution (assuming comparable bitrates). On the other hand, one UDP packet contains seven PES (Packetized Elementary Stream) packets with a maximum size of 188 Bytes in MPEG transport stream. Therefore, one PES packet contains less fundamental information in the higher resolution and its loss should not be so destructive for macroblock reconstruction with positive effect on the error propagation in GOP (GOPs are specific picture frames grouped together and played back so that the viewer registers the spatial motion of the video). We can assume that the final perceived video quality is mostly affected by aspects such as bitrate, codec type, time and spatial information of the video sequence (Yuan and Wang, 2019; Loh and Bong, 2018; Frnda *et al.*, 2019).

The whole list of the selected features and output parameters is summarized in Table 2. Bold text parameters served as an input vector for SOM modelling, ACR (italic) represents the subjective perception of end-users.

Table 2. Testing dataset parameters

Parameter	**Description**
Types of codec	H.264 AVC, H.265 (HEVC)
Bitrate [mbps]	**5, 10, 15**
Packet loss [%]	**0.1, 0.2, 0.3, 0.5, 0.75,1**
Resolution	HD, FullHD, UHD[a]
Objective assessment method	**SSIM**
Range of scenes	**Static, night, sport, very dynamic**
Subjective assessment method – ACR	*MOS scale (classification classes)*

[a] HD stands for High Definition (1280×720), FullHD (1920×1080) and UHD for Ultra High Definition (3840×2160).

As the results indicate, we made feature selection because the rest of the parameters did not have a significant impact on clusters making. We can state that unused feature candidates were approximated by the four selected parameters and by SSIM intervals presented in Table 3.

2.1. Self-organizing Map (Kohonen Map)

Self-organizing maps (SOM) learn to classify input vectors based on the way they are grouped in the input space. They vary from other artificial neural networks as they use competitive learning to modify their weights in comparison to error-correction learning (such as backpropagation supervised learning). SOM can solve many classification problems because competitive learning creates clusters of neurons while each of the created clusters can be interpreted as classification class.

A SOM consists of only two layers, the first one is the input layer and the second one is the output layer or a competitive layer. SOM can provide a graphical representation of a multidimensional dataset in a form of two-dimensional output layer (2D grid) with as many classes as the output layer has neurons (in the worst case).

In competitive learning, output neurons compete between themselves to be selected and only a single neuron is active at any time. The output neuron that wins the "battle" is called the *winner takes all neuron*.

This approach was introduced by Finnish professor Teuvo Kohonen, thus SOM is often called a Kohonen map (Kohonen, 1998). This type of neural network does not demand target outputs to be contained in the dataset, therefore no error (loss) function is presented either. One of the easiest ways to evaluate the accuracy of the trained model is a success rate which represents comparing the total number of testing attempts with the number of cases that have correct class assignation according to the output neuron location.

The output of SOM is an exact position of exciting neuron within the output grid layer. In case of good neurons adaptation on the vector of inputs, SOM can make a cluster – a group of neurons with a similar response to particular elements of the training sample. During the testing phase, each sample causes just one neuron excitation. Based on the position of that neuron, we can assess if it is a part of the cluster and what kind of class is assigned to the cluster.

2.1.1. SOM Learning Algorithm

The first phase of the training process is the weight initialization of each node (small random selected normalized value from 0 to 1). An input vector is chosen randomly from the training set and Best Matching Unit is calculated (BMU). BMU is the winning neuron and its weight vector is the most similar to the input vector according to the Euclidean distance as follows (Blazekova and Vojtekova, 2019).

$$d_j = \sqrt{\sum_{i=0}^{n}(x_i(t) - w_{ij}(t))^2}, \tag{2}$$

where $x_i(t)$ are individual elements of the input vector and w_{ij} is weight between i-th input and j-th output neuron. Then BMU is a neuron with the smallest distance:

$$d_{j*} = min(d_j). \tag{3}$$

Weight adaptation is given by this learning formula:

$$w_{ij}(t+1) = w_{ij}(t) + \eta(t)h(j^*, j)(x_i(t) - w_{ij}(t)), \tag{4}$$

where η is a learning rate and $h(j^*, j)$ is a function that defines how weights of neurons inside the radius are updated. The neighbourhood is an area where the weights are adjusted to make them more like the input vector. Each iteration makes the radius of the neighbourhood shrink over time (e.g. by exponential decay function), until only one neuron rests. The basic height of neighbourhood function h for SOM map is:

$$h(j^*, j) = \begin{cases} 1, & if\ d(i^*, i) \le r(t) \\ 0, & otherwise \end{cases}, \tag{5}$$

where $d(i^*, i)$ is the distance between the winning neuron i^* a specific neuron i, and r represents the size of a radius in time t which also decreases with time. However, the surroundings of the winning neuron may not be square as shown in Figure 4. Many applications use circular or hexagonal topology. Instead of traditional sequential training and weights updating, we chose a batch algorithm. The training data are sent to the SOM at once and update the weights after that. Both approaches are iterative, but the batch version is much faster (Carvalho *et al.*, 2016; Ramirez-Alonso and Chacon-Murguia, 2016).

After training, it is necessary to evaluate or visualize the achieved progress. One of the well-known tools for this is called U-matrix. The U-matrix (unified distance matrix) shows the location and size of potential clusters as it is depicted in Figure 5. Visualisation is based on the Euclidean distance between the input vectors and network response in a gray scale (MATLAB uses yellow to black scale) image. Light areas can be distinguished as clusters and dark areas as cluster borders (Kohonen, 1998). This can be a useful demonstration when one tries to find clusters in the input dataset without having any prior knowledge about the clusters.

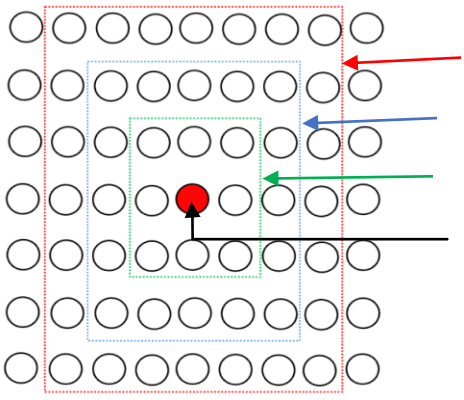


Figure 4. Neighbourhood reduction during training SOM

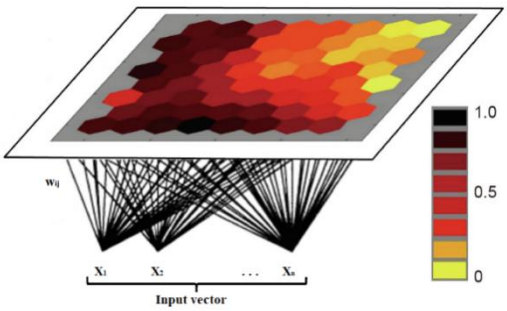


Figure 5. Visual representation of created clusters by the U-matrix

From the created testing video dataset, we also extracted information on the SSIM objective metric intervals related to at least point 3 dedicated to the MOS scale. When the picture quality is worse than 3 on the MOS scale, offered service is poor with annoying visual impairment, thus results showed in Table 3 can serve for very fast video quality estimation.

Table 3. Measured SSIM intervals related to the MOS scale

Video-sequence	ACR (MOS)	H.264 (SSIM)	H.265 (SSIM)
Night scene	≥ 4	0.99-0.95	N/A[a]
	≥ 3< 4	0.95-0.92	0.98-0.95
	≥ 2 < 3	0.92-0.77	0.95-0.89
Static scene	≥ 4	0.99-0.98	N/A
	≥ 3 < 4	0,98-0,94	0.98-0.96
	≥ 2 < 3	0.94-0.86	0.96-0.92
Sport scene	≥ 4	N/A	N/A
	≥ 3 < 4	0.98-0.94	N/A
	≥ 2 < 3	0.94-0.86	0.95-0.91
Very dynamic scene	≥ 4	N/A	N/A
	≥ 3 < 4	0.96-0.89	N/A
	≥ 2 < 3	0.89-0.60	0.93-0.72

[a] N/A stands for Not Applicable. MOS value was not reached.

According to the obtained results, the MOS value 4 and better is roughly related to the SSIM score 0.97; the MOS value between 3 and 4 lies within interval 0.94 - 0.969; and the rest is below 3 on the MOS scale.

3. Results

Several tools and applications exist for Kohonen map implementation. We prefer MATLAB with its Neural Network Toolbox. This toolbox allows using Kohonen map for clustering, as well as for supporting of a batch algorithm. The created dataset contains 432 video sequences affected by packet loss (Frnda *et al.*, 2019). We decided to divide it in ratio 95:5 (testing set vs. hold-out set). A relatively small portion dedicated to the hold-out set occurred due to the unsupervised learning mechanism; therefore, we needed to have much more data for correct SOM modelling.

Selected video features mentioned in Table 2 have an impact on the visualization quality of degraded video sequences. To keep the minimum satisfactory level of end-users, the service providers have to secure at least score 3 of the MOS scale. This requirement led us to define three (like traffic light) potential clusters for Kohonen map implementation:

- Green – suitable quality.
- Orange – sufficient quality. Potential risk of quality deterioration.
- Red – Inadequate quality. Network intervention is needed.

The suitability of selected video features is proven by Figure 6. None of the input parameters can be seen cross-correlating.

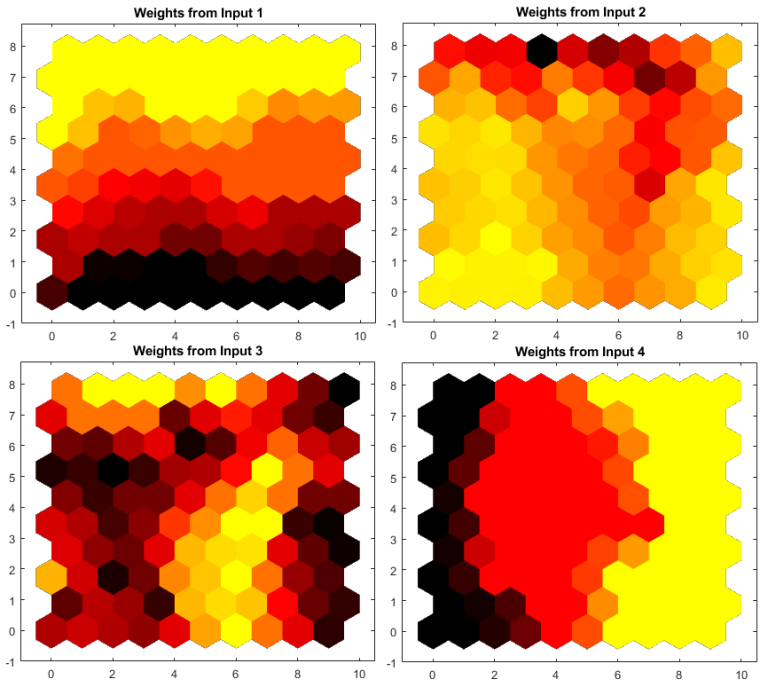


Figure 6. The U-matrix weight plane for each element of the input vector (type of scene, SSIM value, packet loss and bitrate). Input 1 has connections that are very different than those of input 2, 3 and 4

Successfully trained SOM can classify sample data into different clusters. In order for the samples to be predicted, the first step is to judge which cluster they belong to by using SOM. As depicted in Figure 7, after the training process had been done, we could identify several clusters and labelled them. Any pattern may belong to one class in this case. Figure 8 shows the hit diagram with each neuron showing the number of input vectors that it classifies. As shown by the figure, the trained network joints three inputs from the hold-out set with wrong neurons. We obtained a success rate slightly above 0.86% for 10x10 topology, as per results in Table 4.

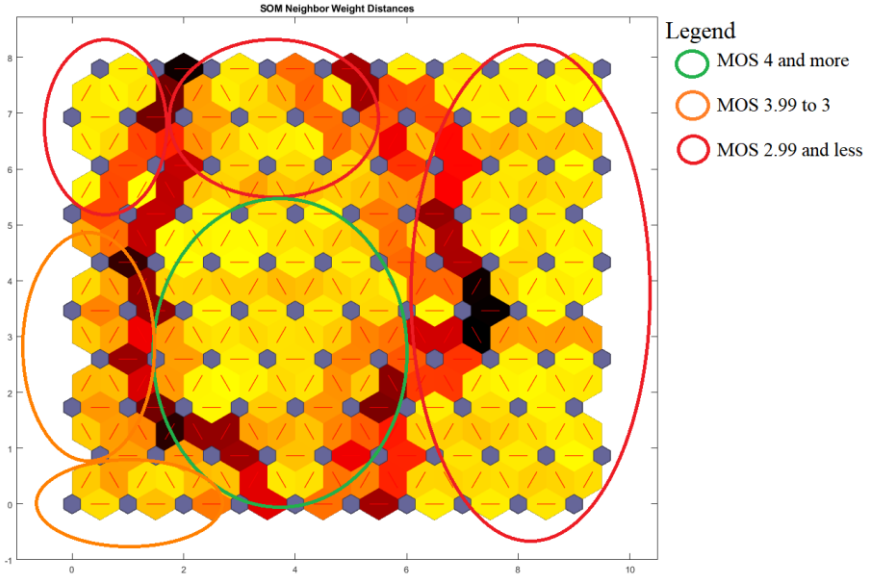


Figure 7. Kohonen map topology with designed clusters

One of the advantages of Kohonen map usage is the overfitting robustness. Since there is no target output vector that overemphasizes individual data points, SOM can avoid overfitting and improve the reliability of the results (Dybskaya and Sverchkov, 2017). Another advantage of using a SOM is that the data are easily interpreted and understood. The reduction of dimensionality and grid clustering makes it easy to observe similarities in the data. We take full advantage of this feature and we have prepared a concept of fast real-time IPTV quality classifier for network providers based on Kohonen map.

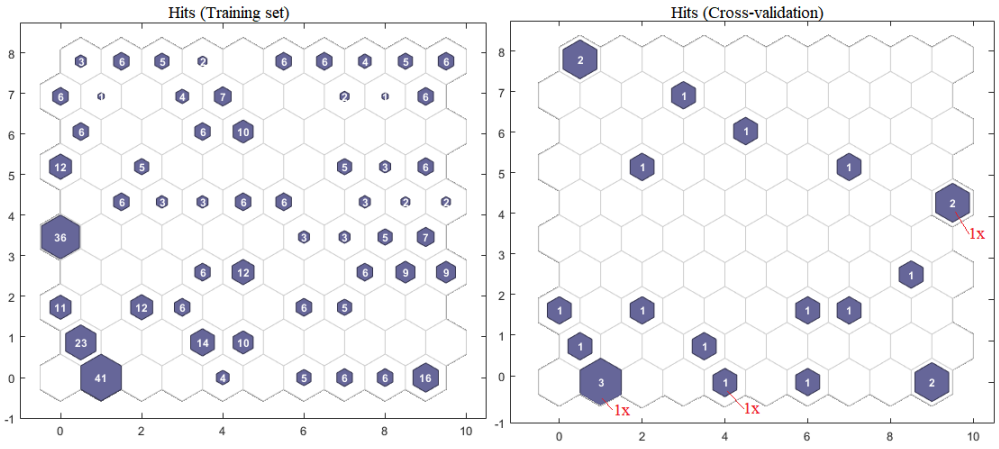


Figure 8. The number of inputs assigned to neurons – Training set (left) and Cross-validation (right). The red number in the Cross-validation set indicates the number of wrong input assignments related to the created clusters

Table 4. Cluster classification success rate for testing set. Bold text represents the selected topology

Topology	Number of correct assignments	Success rate in %
8x8	13/22	59.1
9x9	17/22	77.3
10x10	**19/22**	**86.4**
11x11	18/22	81.8

We can also measure the effectiveness of the model not only by the accuracy parameter. We considered MOS value 3 as a decision boundary and created a 2-class classification problem (good or bad video quality). Confusion Matrix is a performance measurement for machine learning classification. It combines four potential possibilities of classification namely: True positive (TP), True negative (TN), False positive (FP) and False negative (FN). TP and TN represent the right classification while FP and FN are Type I and Type II errors. It is extremely useful for measuring Recall, Precision and F1 score.

- **Recall** (or sensitivity) represents the number of correctly predicted positive cases as follows:

$$Recall = \frac{TP}{TP+FN} = \frac{12}{12+2} = 0.857. \quad (6)$$

  As it can be seen from formula 6, a model that produces no false negatives has a recall of 1.0.
- **Precision** represents the number of correctly predicted positive identifications against the total predicted positive cases:

$$Precision = \frac{TP}{TP+FP} = \frac{12}{12+1} = 0.923. \quad (7)$$

- **F1 score** is the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account. Accuracy works best if false positives and false negatives have a similar cost. If the cost of false positives and false negatives are very different, it's better to look at both Precision and Recall.

$$F_1\ score = \frac{2\times(Precision\times Recall)}{Precision+Recall} = 0.888. \quad (8)$$

4. Conclusion

In this paper, we proposed a novel IPTV quality estimator based on the Kohonen map. We assume that it is more important for service providers to know the quality of the offered video stream than being informed about the exact MOS value. Our model can deliver quality prediction in real-time for both nowadays most used video codecs concurrently.

This paper serves as a starting point, and our results have demonstrated that Kohonen map (or in other words self-organized map) can be useful and competitive with other machine learning techniques - e.g. backpropagation neural networks, decision trees or fuzzy systems - in this research field. In our future work, we intend to focus on data set extension and we plan to cover additional scenes for better dynamic range variability. More data samples also secure sufficient amount of data in order to evolve meaningful clusters. Data deficiency or unimportant data in the weight vectors will add randomness to the groupings. The weight vectors must be based on data that can successfully group and distinguish inputs.

Terrestrial TV broadcast bitrate oscillates between 5- 15 Mbps. H.265 as a part of DVB-T2 has been in trial operation by public service broadcasters since 2016 and it seems like the standard bitrate for H.265 is going to settle on 15 Mbps in near future (which will be fully adopted by IPTV operators because content owners will prepare video stream with uniform parameters for both broadcasting approaches); therefore, our model will meet the actual video standards for the upcoming years. We will also investigate the ways to combine other objective video quality assessment methods with our model.

Acknowledgement

This work was supported by the Institutional research of Faculty of Operation and Economics of Transport and Communications—University of Zilina no. 11/PEDAS/2019 and partially supported by the Travel Award sponsored by the open access journal Electronics published by MDPI.

References

1. Akhtar, Z., Siddique, K., Rattani, A., Lutfi S.L., Falk, T.H. (2019) Why is Multimedia Quality of Experience Assessment a Challenging Problem? *IEEE Access*, 7, pp. 117897-117915.
2. Anegekuh, L., Sun, L., Jammeh, E., Mkwawa, I., Ifeachor, E. (2015) Content-Based Video Quality Prediction for HEVC Encoded Videos Streamed Over Packet Networks. *IEEE Transactions on Multimedia*, 17(8), pp. 1323-1334.
3. Bampis, C.G., Li Z., Bovik, A.C. (2019) Spatiotemporal Feature Integration and Model Fusion for Full Reference Video Quality Assessment. *IEEE Transactions on Circuits and Systems for Video Technology*, 29(8), pp. 2256-2270.

4. Blazekova, O., Vojtekova. M. (2019) Using of Parallel Coordinates in Finding Minimum Distance in Time-Space. *Communications - Scientific Letters of the University of Zilina*, 21(3), pp. 3-7.
5. Carvalho, F.D.T., Bertrand, P., Simoes, E.C. (2016) Batch SOM algorithms for interval-valued data with automatic weighting of the variables. *Neurocomputing*, vol. 182, pp. 66-81.
6. Cheng, Z., Ding, L., Huang, W., Yang, F., Qian, L. (2017) A unified QoE prediction framework for HEVC encoded video streaming over wireless networks. *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, pp. 1-6.
7. Dybskaya, V. V., Sverchkov, P. A. (2017) Designing a Rational Distribution Network for Trading Companies. *Transport and Telecommunication Journal*, 18(3), pp. 181-193.
8. Frnda, J., Nedoma, J., Vanus, J., Martinek, R. (2019) A Hybrid QoS-QoE Estimation System for IPTV Service. *Electronics*, 8(5), 585.
9. Gu, K., Tao, D., Qiao J., Lin, W. (2018) Learning a No-Reference Quality Assessment Model of Enhanced Images With Big Data. *IEEE Transactions on Neural Networks and Learning Systems*, 29(4), pp. 1301-1313.
10. International Telecommunications Union, ITU-T P.910. (2008) Subjective video quality assessment methods for multimedia applications.
11. International Telecommunications Union, ITU-T P.913. (2016) Methods for the subjective assessment of video quality, audio quality and audiovisual quality of Internet video and distribution quality television in any environment.
12. Kohonen, T. (1998) The self-organizing map. *Neurocomputing*, 21, pp. 1-6.
13. Loh, W., Bong, D.B.L. (2018) A Just Noticeable Difference-Based Video Quality Assessment Method with Low Computational Complexity. *Sensing and Imaging*, 19, Article number: 33.
14. Loktev, Daniil A., Loktev, Alexey A., Salnikova, Alexandra V., Shaforostova, Anna A. (2019) Determination of the Dynamic Vehicle Model Parameters by Means of Computer Vision. *Communications - Scientific letters of the University of Zilina*, 21(3), pp. 28-34.
15. Mocanu, D.C., Pokhrel, J., Pablo Garella, J., Seppänen, J., Liotou, E., Narwaria, M. (2015) No-reference video quality measurement: added value of machine learning. *Journal of Electronic Imaging*, 24(6).
16. Mohamed, S., Rubino, G. (2002) A Study of Real-Time Packet Video Quality Using Random Neural Networks. *IEEE Transactions on Circuits and Systems for Video Technology*, 12(12).
17. Mustafa, S., Hameed, A. (2019) Perceptual quality assessment of video using machine learning algorithm. *Signal, Image and Video Processing*, 13, pp. 1495–1502.
18. Ramirez-Alonso, G., Chacon-Murguia, M.I. (2016) Object detection in video sequences by a temporal modular self-adaptive SOM. *Neural Computing & Applications*, 27, pp. 411-430.
19. Sevcik, L., Voznak, M., Frnda, J. (2014) QoE Prediction Model for Multimedia Services in IP Network Applying Queuing Policy. In: *17th International Symposium on Performance Evaluation of Computer and Telecommunication Systems (SPECTS) part of SummerSim Multiconference*, pp. 593-598.
20. Søgaard, J., Forchhammer, S., Korhonen, J. (2015) Video quality assessment and machine learning: Performance and interpretability. In: *7th International Workshop on Quality of Multimedia Experience (QoMEX)*.
21. Song, L., Tang, X., Zhang, W., Yang, X., Xia, P. (2013) The SJTU 4K video sequence dataset. In: *5th International Workshop on Quality of Multimedia Experience (QoMEX)*.
22. Valderrama, D., Gómez, N. (2016) Nonintrusive Method Based on Neural Networks for Video Quality of Experience Assessment. *Advances in Multimedia*, volume 2016.
23. Yuana, Y., Wang, C. (2019) IPTV video quality assessment model based on neural network. *Journal of Visual Communication and Image Representation*, 64, 102629.