A No-reference Hybrid Objective QoE Evaluation for MPEG-4 Encoded Video

Fan Liu, Yang Geng, Jichun Liu, Wenjing Li, and Xuesong Qiu
State Key Laboratory of Networking and Switching Technology,
Beijing University of Posts and Telecommunications,
Beijing, 100876, P. R. China
Email: lfyiyi1@gmail.com

***Abstract*—In order to monitor and ensure adequate Quality of Experience towards end users, this paper proposes a no-reference hybrid objective video quality metrics, based on five high-level parameters extracted from application layer, network layer and bitstream layer. We first cluster various videos based on video content characteristics and then design Back Propagation Neural Network for each content type. The performance evaluation demonstrates a high accuracy in real-time monitoring.**

I. INTRODUCTION

With the fast-emerging of various multimedia services, service providers must ensure and maintain adequate Quality of Experience to be recognized by end users. Thus it is urgent to propose efficient metrics to predict real-time QoE of end users with high accuracy. The existing QoE evaluation methods fall into subjective video quality metrics and objective video quality metrics. No-Reference (NR) metrics, objective assessment, only evaluates video quality based on received sequence and is always adopted for real-time QoE assessment.

The author proposes a no-reference hybrid objective video quality metrics for MPEG-4 encoded video to achieve real-time monitor with high accuracy. Firstly, we present video classification with a novel description of content characteristics. Then five high level parameters are extracted from three layers to evaluate video quality, including network layer—packet loss rate, application layer—content type and GoP (Group of Picture) size, and loss percent of I-frame and P-frame from bitstream layer. We establish Back Propagation Neural Net (BPNN) to predict MOS from the selected parameters.

The remainder of the paper is structured as follows. Section II describes related works and highlights the method adopted in this article. Based on video content characteristics, we cluster different videos into four content types in Section III. Our entire experiment process is described in Section IV: original video selection, impairment generation and subjective experiments. Section V designs and implements BPNN for each content type and analysis their performance. Finally, conclusions are drawn in Section VI.

II. RELATED WORKS

Greengrass et al. [1] propose impact of distorted I-, P- and B-frame is diverse in accordance with the reference relationship in GoP. Since impaired B-frame generally cannot be perceived by users, we only extract high level parameter, loss of I- and P-frame, and ignore the influence of B-frame.

The content type of video is another important factor affecting impairment visibility. Ostaszewska et al. [2] present the incompatibility of Spatial-perceptual Information and Temporal-perceptual Information [3] with human perception and give a proposal of modification in algorithm. We also cluster various video sequences based on the modified algorithm which has been proved to be of high performance.

Staelens et al. [4] present a novel no-reference bitstream-based objective video quality metric. However, it just predicts the impairment to be visible or not and does not provide quantitative QoE result, such as MOS value.

Until now, there exists no QoE assessment based on extracting parameters from three layers, including application layer, network layer and bitstream layer, and predicts perceived quality with MOS value accurately in real-time.

III. VIDEO CLASSIFICATION

A. Q3.SI and Q3.TI measurement

ITU-T Recommendation P.910 [3] defines Spatial-perceptual Information (SI) and Temporal-perceptual Information (TI) to quantity the spatial information and amount of motion, respectively. We analyze some video sequences and draw the waveform diagram of $std_{space}[M_n(i,j)]$ changing over time. More details of $std_{space}[M_n(i,j)]$ is shown in [3]. For example, Fig. 1 shows the waveform diagram of news.

As it is shown, since TI is defined as the maximum value over all the video sequence, only three peaks result in a high TI, while the majority of $std_{space}[M_n(i,j)]$ keep relatively small.

In order to best describe the spatial and temporal information, we compute upper quartile of defined SI and TI in statistics to better represent a video sequence, that is Q3.SI and Q3.TI:

$$Q3.SI = UpperQuartile_{time}\{std_{space}[Sobel(F_n)]\} \quad (1)$$

$$Q3.TI = UpperQuartile_{time}\{std_{space}[M_n(i,j)]\} \quad (2)$$

B. Cluster Results

In this paper, we cluster 20 MPEG-4 digital video sequences from the Video Trace Library and Consumer Digital Video

library. All the selected videos are displayed in a CIF resolution (352*288pixels), a frame rate of 30 frames per second and durations between 10 to 12 seconds. The selected videos span a wide range of spatial and temporal information.

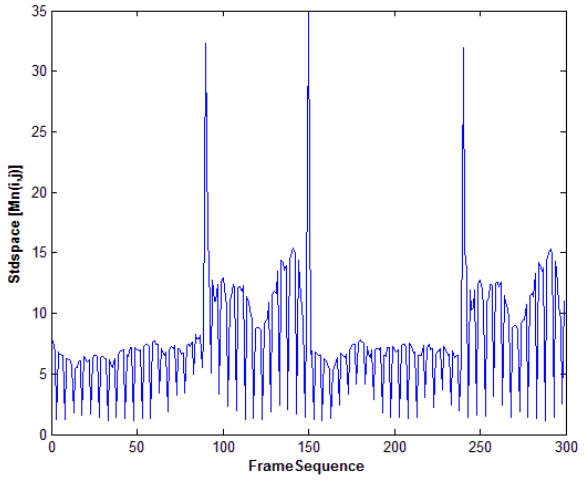

Fig.1. The waveform diagram of $std_{space}[M_n(i,j)]$ changing over time in news

Q3.SI and Q3.TI of original video sequences are first calculated, and then we choose K-means cluster analysis based on Euclidean distance to classify various videos into four content types. The video classification result is shown in Fig.2.

From the cluster result, we can conclude the chief features of each type: Type 1 includes sequences with small moving region of interest on static background. Videos with low or media movement in complicated environment are classified into type 2. Type 3 has high motion, but the video always contain movement in vast monotonous background, such as soccer. Video in type 4 is always highly moving in background with high spatial details, such as mobile.

IV. Hybrid QoE Evaluation

A. Source Video Sequences Selection

Based on section III, we select eight video sequences from the four content types: news, mother and daughter, container, hall, football, soccer, mobile and flower. The screenshots of each video type are shown in Fig. 3.

B. Impairment Generation

We employ MPEG-4 encoding to the original YUV video sequences with GoP size of 9, 12 and 15 which are typically used in an IPTV environment. Then we use Network Simulator 2 (NS2) to simulate real network. We have set packet loss rate from 0.01 to 0.99, and collect statistics of I-frame and P-frame loss percent synchronously. Besides, we mainly set packet loss rate from 0.01 to 0.3 and only 1% of loss rate larger than 0.9.

Fig. 3. Screenshots of eight selected video sequences from four content type

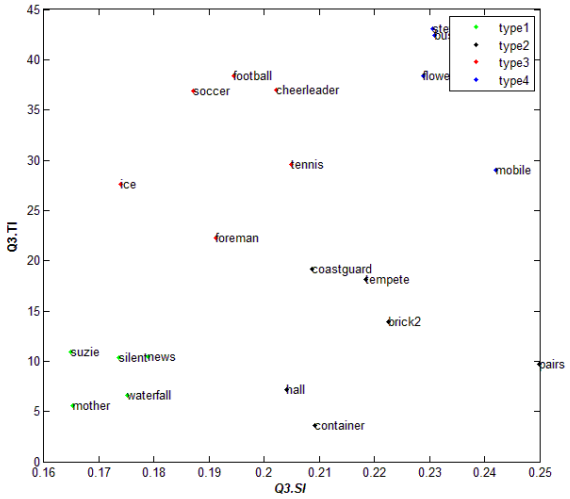

Fig. 2. Video classification Based on content characteristics, Q3.SI and Q3.TI

C. Subjective Quality Assessment

Based on the Single Stimulus (SS) method [5], we select 50 subjects from various professions with age ranges from 18 to 35. After watching each sequence, subjects select a score between 0 and 5 to inform the video quality level [5].

After subjective experiment, we analyze the sample data by means of the mean score and confidence interval.

V. Experiment and Results Analysis

A. BPNN Design

We design a particular BPNN for each content type. To illustrate the structure of our designed BPNN, BPNN with five hidden layers is depicted in Fig. 4. Four neural nodes in input layer receive selected parameters: GoP Size, loss rate of packet, I-frame and P-frame, respectively. The output layer contains a single neural node representing the estimated MOS value.

The number of hidden layer nodes has significant influence on performance of BPNN. The author refers to previous experience to get estimate of nodes number:

$$\sum_{i=0}^{n} C_M^i > k \qquad (3)$$

Where k is sample size, M is number of hidden layer nodes, n is number of input layer nodes. If $i > M$, we define $C_M^i = 0$. Then we change the node number around the computed result and obtain the most appropriate number for each BPNN.

While adjusting weights and thresholds of NN, We apply Motentum Backpropagation BP Algorithm (MOBP) which is based on Least Mean Square Algorithm [6] and introduces momentum factor α in weight update stage, thus Correction value of weights keeps a certain inertia:

$$\Delta w(n) = -\eta(1-\alpha)\nabla e(n) + \alpha w(n-1) \qquad (4)$$

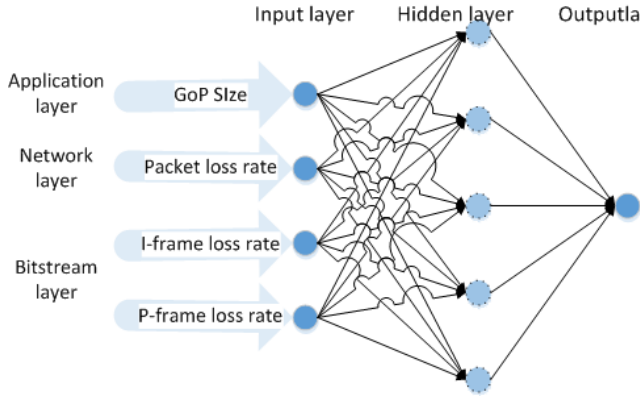

Fig. 4. The structure of BPNN for a particular content type, for example five neural nodes in hidden layers

B. Experiment and Results Analysis

After collecting samples and designing BPNN, the samples of each content type are normalized and then trained with corresponding BPNN. If the BNPP converges successfully and achieves required performance, it is what we want. The process is depicted in Fig. 5.

In order to illustrate the performance of our designed BPNNs, Regression R values are shown in Fig. 6. Regression R values measure the correlation between output and target, that is correlation coefficient of estimated MOS and subjective MOS value.

After BPNN for a particular content type is trained with high performance, the author inputs the testing samples and compares the estimated MOS with values from actual subjective experiment. The test performance of our BPNNs for four content types is measured by Mean Squared Error (MSE) and Regression R Values. MSE and R value for each BPNN are depicted in Table I.

TABLE I
MSE AND REGRESSION R VALUE FOR TESTING PERFORMANCE

Content Type	MSE	R value
Type 1	0.97761	0.0216
Type 2	0.9702	0.0128
Type 3	0.971	0.0165
Type 4	0.9672	0.0323

VI. CONCLUSION

In this paper, we present a no-reference hybrid Objective QoE evaluation for MPEG-4 encoded video. Based on a novel description of video content characteristic, we propose a video classification mechanism by means of cluster analysis and systematize various video sequences into four content types. In QoE evaluation stage, five high level parameters are extracted from three layers: application layer, network layer and bitstream layer, and a hybrid QoE assessment metrics achieves real-time monitoring with high performance.

ACKNOWLEDGMENT

This research is supported by the Funds for Creative Research Groups of China (61121061), National Key Technology R&D Program (2012BAH06B02), and Chinese Universities Scientific Fund (BUPT2012RC0608).

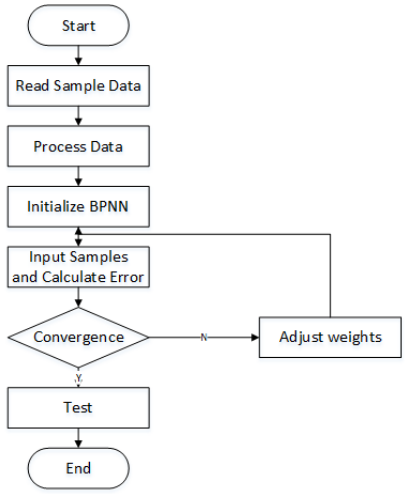

Fig. 5. Flow charts of objective experiment process based on BPNN

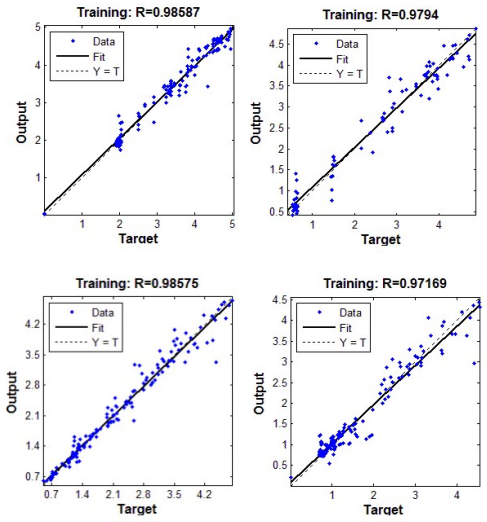

Fig. 6. Correlation coefficient of estimated MOS and subjective MOS value for content type 1, 2, 3, 4 respectively

REFERENCES

[1] J. Greengrass, J. Evans, and A. C. Begen, "Not all packets are equal, part 2: The impact of network packet loss on video quality," Internet Computing, IEEE, vol. 13, no. 2, pp. 74–82, 2009.

[2] A. Ostaszewska and R. Kłoda, "Quantifying the amount of spatial and temporal information in video test sequences," in Recent Advances in Mechatronics. Springer, 2007, pp. 11–15.

[3] P. ITU-T RECOMMENDATION, "Subjective video quality assessment methods for multimedia applications," 1999.

[4] N. Staelens, G. Van Wallendael, K. Crombecq, N. Vercammen, J. De Cock, B. Vermeulen, R. Van de Walle, T. Dhaene, and P. Demeester, "No-reference bitstream-based visual quality impairment detection for high definition h.264/avc encoded video sequences," Broadcasting, IEEE Transactions on, vol. 58, no. 2, pp. 187–199, 2012.

[5] I. R. Assembly, Methodology for the subjective assessment of the quality of television pictures. International Telecommunication Union, 2003.

[6] Z. Pin, "Matlab neural network application design," 2013.