QUALITY-OF-EXPERIENCE ASSESSMENT AND ITS APPLICATION TO VIDEO SERVICES IN LTE NETWORKS

KAN ZHENG, XIAOLI ZHANG, QIANG ZHENG, WEI XIANG, AND LAJOS HANZO

ABSTRACT

Reliable and repeatable video quality assessment is essential for performance analysis of wireless multimedia applications in the third generation LTE network. In this article we report on a database containing subjective assessment scores and the corresponding Quality-of-Service parameters of 70 video test sequences encoded with H.264, which are corrupted when transmitted over a wireless 3G LTE network simulator. Then, a new assessment method based on NN is proposed, whose weights are determined through training. The resulting pseudo-subjective assessment scores are then compared to the true MOS results in our database. Naturally, the accuracy of the NN-based prediction tool should be tested 'outside' of the set used for NN weight training. However, there are persistent residual errors between the predicted and subjectively evaluated MOS, which can be further reduced by particle swarm optimization, applied as post-processing of the NN weights to improve its accuracy. The proposed assessment method has potential applications, including Quality-of-Experience-aware network optimization for LTE network operators.

INTRODUCTION

With escalating demands for digital entertainment, digital video becomes an essential part of everyday life. Meanwhile, broadband wireless communication networks such as the Third Generation Partnership Project (3GPP) Long Term Evolution (LTE) networks have become more pervasive than their wire-line based counterparts. LTE is capable of supporting innovative new services, providing high quality of experience. Video traffic covers a wide range of services, such as video telephony, entertainment, advertising, as well as video-enabled call centers and services. The video services provided by LTE networks are expected to become increasingly more popular. Therefore, the provision of reliable video quality assessment is essential for the performance analysis of wireless multimedia applications in LTE networks.

The quality of experience (QoE), which is defined as the overall consecutive quality of an application or service, is commonly used for representing true user perception [1]. The end-to-end QoE is influenced by a wealth of factors in wireless networks. Therefore, accurate evaluation of the QoE of video services is a challenging task. In general, the factors affecting the QoE may be classified into subjective and objective categories [2]. The subjective factors include user emotion, experience, expectation, and so on, while the objective factors relate to both technical and non-technical aspects of the services. The network's end-to-end QoS parameters, the network/service coverage, and the terminal functionality are among typical technical factors. The convenience of service setup, service content, pricing, and customer support are some examples of the associated non-technical factors.

Since video services are ultimately watched by human observers, the subjective opinion of the observer is the best QoE indicator of video services. In a typical subjective assessment study, the original reference video sequence and the corresponding sequence received after channel-induced impairments are presented to a panel of observers, who give their subjective ratings. The mean opinion score (MOS) is the most commonly used metric representing these opinions [3]. There already exist standards and studies defining the procedures of subjective video quality evaluation [4, 5]. These methods differ in terms of their evaluation metrics with regard to the specific application considered, such as the reference video, the legitimate range of results, the length of the video sequence, the number of observers used, the type of display, and so on. Hence, subjective tests are the most pertinent QoE measurements, because the results are obtained directly by humans. However, subjective quality evaluation requires considerable human resources and time. It can neither be easily repeated nor used in real-time. Due to the limitations of regular subjective measures, researchers tend to rely on simple objective measures for video services. These methods are

K. Zheng, X. Zhang, and Q. Zheng are with Beijing University of Posts & Telecommunications.

W. Xiang is with University of Southern Queensland.

L. Hanzo is with University of Southampton

mainly based on certain objective metrics, such as the peak signal-to-noise ratio (PSNR) [6], structural similarity index metric (SSIM) [7], video quality metric (VQM) [8], and so on. Moreover, the weights of a given QoS/QoE mapping function can be generated by using statistical analysis [9]. However, the results provided by conventional objective techniques do not correlate well with those obtained by subjective approaches. Then, another approach, termed pseudo-subjective quality assessment (PSQA), was proposed due to its combined advantages of both subjective and objective methods [10]. The quality assessment model is first trained by the results gleaned from subjective approaches, which reflect the effects of subjective factors. Upon convergence, this model can be used for predicting the quality of video services without any human involvement [11]. However, these approaches have not been widely investigated at the time of writing.

In order to fill this gap, our work focuses on the critical assessment of video streaming over a LTE network, taking into account both the objective network parameters as well as the users' subjective factors. A reliable QoE assessment method is essential for wireless operators in designing efficient radio resource management schemes for the sake of meeting the anticipated future demand of video streaming services at satisfactory QoE levels. Our main contributions are summarized as follows.

• First, we report on the results of a large-scale subjective assessment campaign with the objective of building up a database for evaluating the QoE of video services in wireless networks with the aid of an LTE wireless network simulator, which generates impaired video sequences for different values of the controllable QoS parameters. Each impaired test sequence is associated with specific QoS parameters. Correspondingly, the opinion scores of human observers are collected and averaged for generating the MOS as the QoE metric for the given impaired test sequence. We collect the QoS parameters and their corresponding QoE metric in a database.

• Second, a new QoE assessment method relying on a two-step structure is proposed for achieving good accuracy with a feasible implementation complexity. Thanks to its self-learning property and acceptable computational complexity, a feed-forward back propagation (BP) neural network (NN) is adopted for MOS prediction. Moreover, compared to the existing PSQA scheme, in order to avoid any local minima and to improve the achievable accuracy, particle swarm optimization (PSO) [12] is invoked for post-processing the NN's weights when necessary. To the best of our knowledge, there exists no prior work on integrating the NNs with PSO for successively predicting the video quality in the literature. The samples including the QoS parameters and QoE metric in the database can be used for training as well as for characterizing the performance of the proposed method "outside" the training set.

• Finally, the accuracy improvements achieved by the proposed QoE assessment method are characterized by our numerical results for
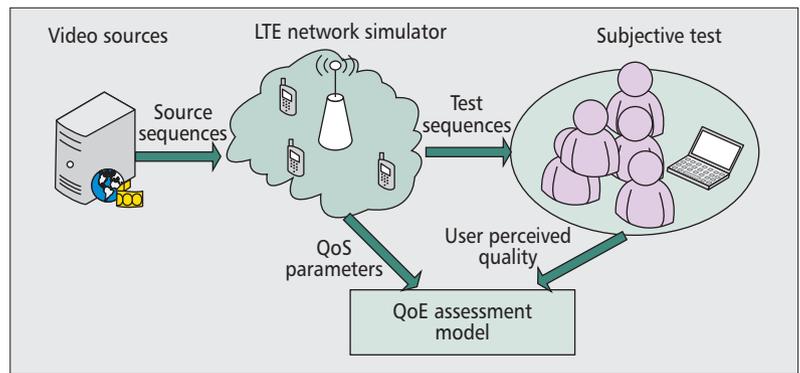
Figure 1. Framework of the video assessment method.

demonstrating the benefits of the proposed PSO-aided NN weight post-processing. As a result, our proposed method is capable of achieving reliable QoE prediction.

The remainder of this article is organized as follows. The next section commences with a brief overview of subjective quality assessment, while our new quality assessment method is proposed after that. Then our numerical results are presented. The final section concludes this article.

SUBJECTIVE QUALITY ASSESSMENT

Based on our subjective quality assessment, we can construct a database containing a large set of subjective assessment scores recorded for diverse video streams encoded with the H.264 codec, which is corrupted by a simulated LTE network. This database can then be utilized for studying the reliability of our proposed quality assessment method. The framework of the video assessment method is shown in Fig. 1, whose details are provided below.

SOURCE SEQUENCES

As shown in Fig. 2, six test video sequences in the YUV format are used in our experiments. These sequences are of the Common Intermediate Format (CIF) (352 × 288 pixel) image resolution, and the frame rate is 30 frames per second. The representative CIF was specifically chosen, since it is the recommended screen size of mobile phones and small handheld terminals as well as of tablet PCs. Each of the six video clips contains 300 frames with the duration of 10 seconds. The sequences have diverse content in terms of motion activity, objects, people, and so on.

As defined in the ITU-T P.910 recommendation, both spatial information (SI) and temporal information (TI) can be used to characterize the quality of video sequences [4]. Spatial domain perceptual quality measurements are typically based on the Sobel filter, which is applied to each luminance frame at each time instance. Once the standard deviation of the (352 × 288)-pixel CIF image from the original sequence is computed, the maximum value within the entire 300-frame sequence is taken as the SI. By contrast, temporal-domain perceptual quality measurements are based on the pixel difference between the two neighboring frames. Then the TI is calculated as the maximum standard deviation of the pixel difference over the (352 × 288)-
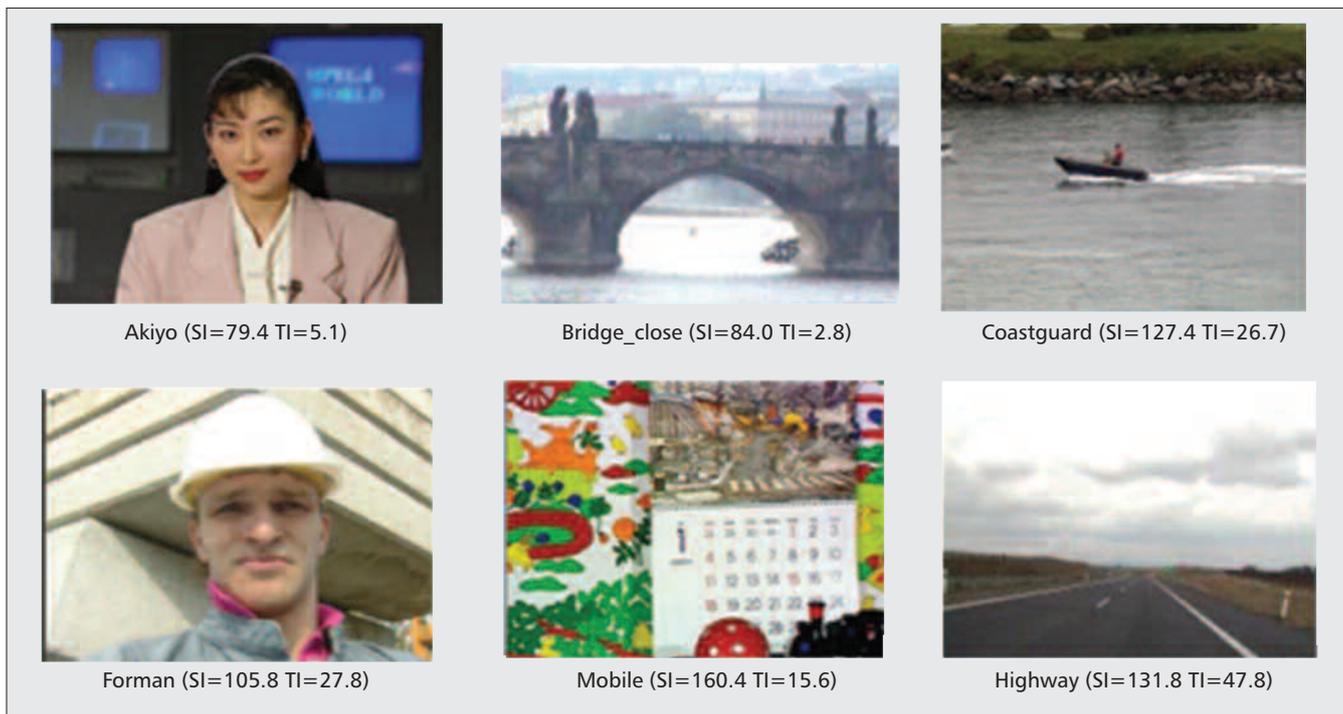
Figure 2. Illustration of the source video sequences.

pixel CIF space within the entire video clip. The SI and TI results recorded for the six test videos are provided in Fig. 2, where it can be seen that the test sequences exhibit a wide range of SI and TI values.

TEST SEQUENCES

In our questionnaire-based experiments, a set of test sequences were generated from the six source sequences. The test sequences suffer from different levels of distortions after being transmitted over the LTE wireless network simulator. The main processes of the video transmission are described in the following sections.

Video Compression — Owing to the bandwidth limitation of wireless networks, it is necessary to compress a video source before its transmission by means of compression algorithms. H.264 is an industrial video compression standard that converts digital video into a format that requires less bandwidth for transmission [13]. H.264/SVC (scalable video coding) has the ability to adaptively support different qualities of the transmission links. The H.264-encoded video contains the so-called base layer (BL) and several enhancement layers (ELs). The former provides the basic video quality for transmission over low-quality channels, while the latter is capable of increasing the video quality when the channel is capable of supporting an increased data rate. According to the available bandwidth, dropping one or more enhancement layers is an efficient method to avoid real-time video transcoding [13].

The H.264 family enhances the encoding loop consisting of a combination of intra-frame coding and inter-frame coding relying on motion compensation. Each video frame is either intra-coded (I), forward predictive coded (P), or bi-directionally predictive coded (B), according to its specific position in the Group of Picture (GoP) structure, such as IBBPBBPBB. The first I and P frames serve as a reference for the first two B frames. These reference frames must be transmitted first. Hence, the transmission order and display order may be different from each other.

As illustrated in Fig. 3, the encoding process is implemented by the JSVM encoder [14]. After encoding, an H.264 video file is generated. In order to obtain the trace file, the video file is fed into the Bit Stream Extractor to produce the original Network Abstraction Layer Unit (NALU) trace file. However, the NALU trace file does not contain the frame index information. Therefore, this trace file is processed by F-N Stamp to generate NALU trace files containing the information of the frame index [15]. The trace file information is then stored for transmission over LTE wireless networks.

Transmission over LTE Wireless Networks — A system-level 3G LTE wireless network simulator based on the OPNET software is used to qualify the impact of wireless transmissions on the source sequences. In this simulator, the video sequences are delivered over the downlink of the 3G LTE wireless network. A video server provides live/on-demand video content, where the video files are saved in their H.264-encoded versions [13]. At the evolved Node B (eNB), the video is sent to the user equipment (UE) using the available radio resources, for studying the effects of the wireless channel on video transmission. The details of our simulation procedures are shown as follows.

Step 1: Initialization: The parameters are initialized before simulations. All the UEs are dispersed uniformly across the cell. First, the trace file of the test video sequence is obtained, which contains all the parameters of each frame of the
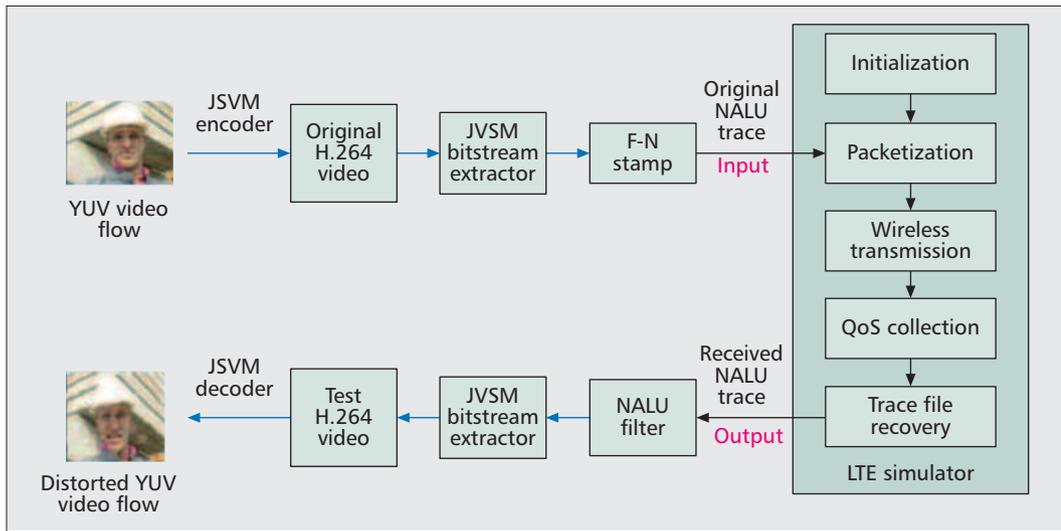
Figure 3. Illustration of creating distorted videos by a LTE network simulator.

source video, including the frame type, frame size, frame index, and so on.

Step 2: Packetization: For a given UE, the simulator generates virtual frames according to the trace file information in the application layer. Then each virtual frame is divided into several packets for transmission.

Step 3: Wireless Transmission: Given the scheduling scheme, the eNB allocates the available radio resource blocks (RBs) to different UEs based on the near-instantaneous wireless channel conditions. Each packet is further separated into multiple bursts and then transmitted over the assigned RBs. For each RB, the specific choice of modulation and code scheme (MCS) is selected for transmission, depending on the near-instantaneous channel condition [16]. Usually the objective of the adaptive modulation and coding (AMC) scheme is to maximize the transmission data rate without violating the burst loss rate target. For example, in the case of good channel conditions, a modulation/coding scheme with higher throughput is selected and vice versa. If a video burst is incorrectly received, the eNB has to retransmit this burst with the aid of a specific HARQ mechanism and Chase-Combining reception. The maximum number of retransmissions must not be excessive in order to avoid excessive delay.

Step 4: QoS Collection: The UE receives the bursts and checks whether they are decoded correctly or not. If all of the bursts in a packet are received without errors, the UE assembles these bursts into a packet. Otherwise, a packet is assumed to be lost and discarded. Meanwhile, a set of QoS parameters including the delay, delay-jitter, packet loss rate (PLR), and MLBS are recorded.

Step 5: Sequence Recovery: Following the downlink transmissions via the LTE network, the trace file can be recovered at the UE. If all the packets in a frame are assembled correctly, the UE removes the header of each packet and assembles them into the corresponding frame in accordance with the packet index. If any packet in a frame is lost or cannot be assembled in time for display, this frame is regarded as being lost.

The NALU-Filter records the NALUs in accordance with their frame-index order, and removes the NALUs of those specific frames that have not arrived at the receiver in time. After filtering, the trace file is fed to the JSVM Bit Stream Extractor together with the original compressed video, generating the received H.264 videos. Next, the JSVM decoder decodes them into a YUV video. In the decoding process, the dropped frames are replaced through copying their previous frames. In this way, a test video sequence is generated with the given QoS parameters. The simulation configurations are flexible, including the UE's location, and so on. Then the procedures from *Steps* 2 to 5 are repeated in order to obtain different test sequences associated with different sets of QoS parameters.

For comparison, we store the traces of the source video streamed from the eNB as well as the test video streams recorded at the UEs during simulations. The QoS parameters including jitter, delay, the PLR, and the mean loss burst size (MLBS) are collected for QoE assessment. Given the different QoS parameters used in our simulator, in total 70 test sequences are generated in our experiments.

SUBJECTIVE TEST

In order to minimize the impact of random environmental factors on the results, there exist standardized methodologies for subjective quality evaluation. In our experiments, we present both the encoded and impaired video sequences to human subjects using a double stimulus (DS) absolute category rating (ACR) based subjective assessment methodology [4]. To elaborate on the DS studies, a pair of video clips are shown simultaneously on a split screen. The observers have to evaluate the impairments imposed on a test sequence by the network against the original sequence, using a five-point scale ranging from "very annoying" to "imperceptible." In our study the MOS is used to quantify the video impairments, which is a subjective metric of quantifying the video quality at the application level. The MOS is usually quantified on a scale ranging from 1 (worst) to 5 (best).

In each video presentation, a pair of video sequences including the reference video and a distorted one is presented to the observers only once. The play list of the 70 video pairs was presented to each observer in a random order using a random number generator. This way, no two observers evaluate the sequences in exactly the same order.

Figure 4. Illustration of the NN-based quality evaluation method.

Most observers in our subjective quality assessment are postgraduate students from Beijing University of Posts & Telecommunications (BUPT), aged between 20 and 28. As the observers are inexperienced in video quality assessment, specific instructions are provided to all the observers on how to evaluate the video sequences. First, a standard visual acuity chart is used for testing the visual acuity and normal vision of observers. Then we present three training sequences to indicate the typical impairments, which they might perceive during the experiment in order to familiarize the viewers with the test software, the type of assessment, the grading scale, reference sequence, test sequence, and voting period. The range and types of the impairments to be assessed are detailed in a questionnaire. Note that when analyzing the test results, the quality ratings that were assigned by the viewers to the training sequences are not taken into account.

Again, the test session comprises 70 video presentations, since the video database consists of 70 distorted video sequences and the original reference video sequence. In each video presentation, a pair of video sequences including the reference video and a distorted one is presented to the observers only once. The play list of the 70 video pairs was presented to each observer in a random order using a random number generator. This way, no two observers evaluate the sequences in exactly the same order. Each pair of the 300-frame sequences is displayed for 10 seconds at 30 frames per second. Afterward, the observers are required to complete a questionnaire with the aim of judging the difference between the original reference and the impaired video sequences on a discrete five-point scale. At the end of each presentation the observer is given five seconds of voting time to rate the quality of the stimulus based on the above-mentioned scale, ranging from "very annoying" to "imperceptible." The test for each viewer was limited to less than half an hour so that the observers do not experience any uneasiness or fatigue during the experiment. Note that the numerical values attached to the scale are used only for data analysis, but are not shown to the observers.

DATA POST-PROCESSING

Subjective evaluation according to the above steps produces integer values between 1 and 5. There may be variations in these distributions due to the subjective judgement differences between the observers. As a result, the overall mean score of the impaired video sequence is calculated as the arithmetic mean of all the individual scores.

PROPOSED QUALITY ASSESSMENT METHOD

In this section we propose a new quality assessment method based on neural networks (NN), which takes into account both objective factors and subjective factors. Our specific objective is to design and train an NN that is capable of estimating the MOS based on a number of LTE quality metrics, such as the delay, delay-jitter, packet loss rate (PLR), and so on. We will train this NN with the aid of our MOS scores recorded by the observers for specific sets of the above mentioned parameters so that in the future the operators will need no human-based training.

NN-Based Quality Assessment Method

As shown in Fig. 4, there are two main steps in our proposed method. First, a suitable NN architecture is chosen and its weights are trained using our data in the MOS database. In order to improve the accuracy of our QoE prediction technique, post-processing of the NN's weights may be invoked after the first step, if deemed necessary. After the training, the NN becomes capable of evaluating the video quality for arbitrary video clips and channel impairments. Further details of this method are described as follows.

Step 1: NN Training: Since the feed-forward (FF) back propagation (BP) NN is of acceptable computational complexity, it is used in the first step of our quality assessment method. In this FF NN, the information flows only along the forward direction (called forwards), i.e. from the input nodes data to the so-called hidden NN nodes (if any) and then to the output nodes. There are no cycles or loops in the NN. As described in the previous section, we have constructed a video quality database, storing the MOS values and the corresponding QoS parameters. This database can then be used for assessing the quality of video transmitted over the LTE wireless network. The database is divided into two sets, one of which is used for training the NN, while the other is for testing our techniques outside this training set by comparing its objective scores with the MOS scores.

A three-layer fully connected BP neural network is trained for estimating the MOS with the aid of four input nodes as the main objective QoS parameters, namely, the delay, delay-jitter, PLR, and MLBS. The NN only has a single output node corresponding to the QoE metric, namely, the MOS. Moreover, five hidden nodes are invoked so as to strike a trade-off between the MOS-estimation accuracy and the complexity imposed. As a result, a N1 x N2 x N3 = 4 x 5 x 1 BP neural network structure is formulated in the first step. The mean squared error (MSE) between the desired and actual neural network output averaged over all training data is chosen as the fitness metric of the neural network, that is,

$$\overline{\eta} = \frac{1}{N_d} \sum_{n=1}^{N_d} \left[z(n) - y(n) \right]^2, \tag{1}$$

where N_d denotes the number of data points used for training the BP-NN, $y(n)$ is the output of the NN, and $z(n)$ is the desired output (i.e. MOS) given by the database. Taking the square root of the MSE yields the root-mean-square error (RMSE).

In the BP-NN, a gradient-based approach is used for optimizing the NN weights, for the sake of minimizing the MSE between the MOS scores rated by the observers and those estimated by the NN. The steepest decent minimization of the MSE adjusts the NN weights proportionately to the MSE gradient w.r.t. the above mentioned delay, delay-jitter, PLR, and MLBS parameters so that the MSE is reduced after each weight adjustment. After a given number of iterations are carried out, for example 10, the training process is completed and the resulting MOS-learning model is ready for use in the next step.

Step 2: NN Weight Post-Processing using Particle Swarm Optimization (PSO): Whether PSO should be invoked at all depends on the specific performance requirements. As shown in Fig. 4, the MSE between the desired and actual MOS-outputs of the NN relying on the NN-weights trained during the first step is calculated and compared to the target MSE $\overline{\eta}_t$. If the MSE is no larger than $\overline{\eta}_t$, the NN weights of the model are confirmed and ready to be used for quality assessment. Otherwise, the proposed PSO-aided post-processing is applied to the weights with the objective of further reducing the model's MSE. It is plausible that the post-processing will remain futile, if the initial MSE is excessive. On the other hand, a reduced MSE is achieved at the expense of a higher computational complexity. Therefore, a suitable value of the target MSE is based on a trade-off between the achievable MOS-prediction performance and the complexity imposed, which can be found by simulations.

The PSO algorithm was originally conceived for simulating the social behavior of bird flocks. In PSO, a particle refers to a point in the design space that changes its position based on both the direction and velocity of the NN-weights updates. Initially, a set of particles randomly propagates in the design space, moving in randomly defined directions over a number of consecutive PSO iterations. The direction of a particle then gradually changes in the direction of the best previous positions of both itself and of its peers in its vicinity for the sake of discovering even better positions quantified in terms of a certain fitness measure or objective function.

The dimension of the search space for each particle is given by the number of NN weights and bias-values used in the BP-NN, which is $d = (N_1 N_2 + N_2) + (N_2 N_3 + N_3) = 31$. Without loss of generality, the position of the first particle is initialized in accordance with the weights and bias values of the model, which are obtained by the training in the first step. Meanwhile, the positions of the other particles are initialized to random values obeying a given uniform distribution. The velocities of the particles are randomly generated by using the specific upper and lower bounds of V_{MAX} and V_{MIN}. In each iteration, each particle tries to adjust its velocity according to both its own personal best position ever visited, as well as the best previous global position attained by any of its neighbor particles in order to minimize the error function of the BP-NN.

Discussions

For the video services considered, the NN-based model can be used for quality assessment after its training is completed. In order to ensure the accuracy of the model, it is important to choose suitable QoS parameters as the input nodes of the proposed model. In other words, we have to find out which specific QoS parameters the QoE of the given video service depends mostly on. The QoS parameter values invoked for training are expected to be sufficiently diverse to include almost all practical possibilities. On the other hand, the subjective test used for establishing the reference QoE must be performed carefully according to a pre-defined subjective assessment methodology.

The QoS parameter values invoked for training are expected to be sufficiently diverse to include almost all practical possibilities. On the other hand, the subjective test used for establishing the reference QoE has to be performed carefully according to a pre-defined subjective assessment methodology.

Naturally, the accuracy of the proposed model based on the neural network and particle swarm optimization (PSO) may be improved at the cost of imposing an increased complexity, for example by increasing either the number of iterations in the NN or the number of particles in the PSO. However, the accuracy typically cannot be proportionately increased as a function of the complexity. Indeed, there is a trade-off between the accuracy achieved and the complexity imposed, which will be discussed in the next section.

Although only CIF-resolution videos were encoded by the H.264/SVC codec in our experiments, the proposed method can be readily invoked for the quality assessment of other resolutions, including High-Definition (HD) video frames. Even if the video frames are encoded with the aid of other codecs, our technique remains applicable, but the model must be re-trained.

NUMERICAL RESULTS AND ANALYSIS

In this section the performance of the proposed assessment method is evaluated and compared to other methods in terms of the quality metrics

Parameter	Value
Cellular layout	3-sector sites
Carrier frequency	2.0 GHz
Bandwidth	10 MHz
Size of resource bock (RB)	180 KHz
Number of resource bock (RB)	50
Fast fading model	SCME
Antenna configuration (Tx:Rx)	1:1
Inter-eNB distance	500 m
Number of UEs	10
Path loss model	$PL(R) = 128.1 + 37.6\log10(R)$, R in km
Shadowing standard deviation	8 dB
Antenna pattern (horizontal)	$A(\theta) = -\min[12\theta/\theta 3dB)^2, A_m]$ $\theta_{3dB} = 70°$, $A_m = 20$ dB
Transmit power	46 dBm
TTI length	0.5 ms
CQI feedback	2 TTI
ACK/NACK delay	8 TTI
AMC scheme	27
HARQ mode	Chase combining
Max # of retransmission	3

Table 1. Main parameters in the LTE simulator.

considered. The video quality impairments imposed by wireless transmissions were varied by appropriately adjusting the parameters of the 3G LTE wireless network simulator. The main parameters employed in our experiments are shown in Table 1 [17].

EFFECTS OF THE NUMBER OF ITERATIONS/PARTICLES

Most parameters used during our NN-aided estimation are assumed to be fixed. Then we need to determine the number of PSO particles and iterations used in the PSO algorithm. It is apparent that the higher the number of particles and iterations, the better the MSE performance at the expense of computational complexity. Hence, from the viewpoint of practical implementation, it is necessary to trade off the achievable performance against the computational costs. Figure 5a quantifies the attainable MSE performance versus complexity for our assessment method as a function of the number of PSO iterations, which is assumed to be 30. Naturally, when the number of PSO iterations increases, the MSE of MOS-estimation improves. However, an MSE floor may emerge when the number of iterations becomes higher than 60, even though the complexity quantified in terms of the number of floating-point operations per second (FLOPS) continues to increase linearly. Therefore, the number of iterations was chosen to be 60 in our experiments. On the other hand, Fig. 5b presents the MSE performance versus the complexity as a function of the number of PSO particles. It can be observed that the number of particles can be selected to be around 30 in order to strike an elegant trade-off between the attainable performance and the imposed complexity.

MSE PERFORMANCES WITH/WITHOUT WEIGHT POST-PROCESSING

PSO-based weight post-processing is invoked only for $\overline{\eta} \le \overline{\eta}_t$. The MOS-estimation MSE performance of the proposed method is characterized in Fig. 5c with different values of the threshold $\overline{\eta}_t$, where the relative ratio of activating the post-processing is also shown. As expected, upon increasing the threshold value, the MSE is reduced as a function of the post-processing activation ratio. However, the MSE improvement is insignificant with large threshold values. Therefore, we opted for $\overline{\eta}_t = 0.2$ as a good trade-off between the performance achieved and the complexity imposed.

To elaborate further, Fig. 5d plots the cumulative distribution function (CDF) of the MOS-estimation MSE, when the proposed assessment method is applied both with and without weight post-processing. It is clear from Fig. 5d that our weight post-processing is capable of achieving performance gains, especially with a relatively high initial MSE. For example, MSE ≤ 0.25 may be achieved without PSO in 62 percent of the cases, which increases to 91 percent with the assistance of PSO.
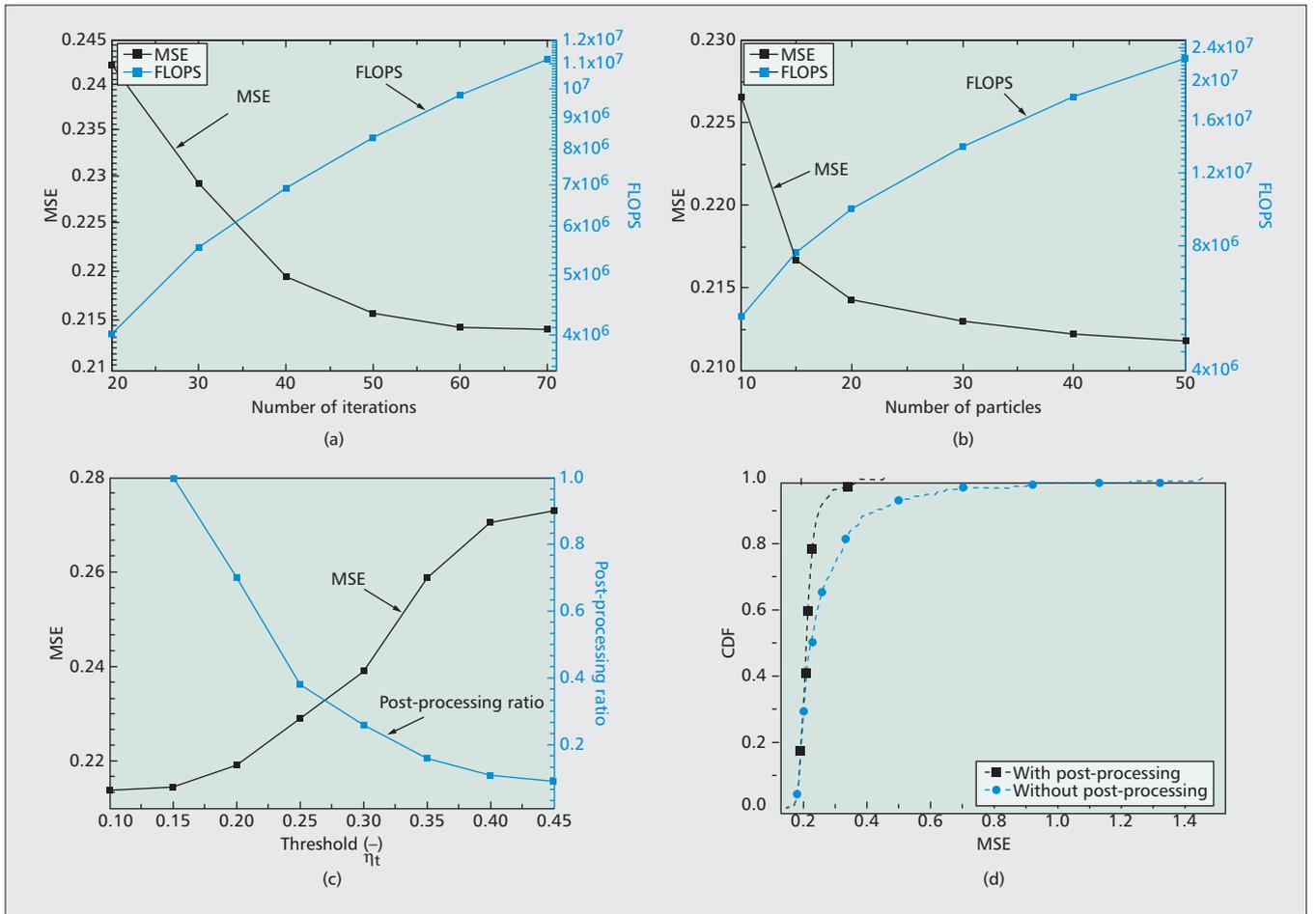
Figure 5. MSE performance under different conditions. a) MSE and complexity versus the number of PSO particles; b) MSE and complexity versus the number of PSO particles; c) MSE and post-processing ratio versus threshold value; d) MSE CDF performance with or without post-processing.

QoE Performance Evaluation using Different Assessment Methods

Figure 6 presents the true subjective scores, the objective PSNR-to-MOS results, the VQM results, and the MOS results attained by the proposed method, respectively. The MOS vs. PSNR mappings are MOS = 5 if PSNR $\geqslant$ 37, MOS = 4 if $31 \leqslant$ PSNR < 37, MOS = 3 if $25 \leqslant$ PSNR < 31, MOS = 2 if PSNR $20 \leqslant$ PSNR < 25, and MOS = 1 if PSNR $\leqslant$ 20. Then the PSNR metric can be mapped to the integer MOS scores between 1 and 5. However, these MOS results do not accurately reflect the actual user experience reported by the observers. Moreover, the VQM as a standardized objective method of measuring video quality is also used as a reference [8]. In Fig. 6 we observe that compared to the PSNR-to-MOS and VQM methods, the MOS estimated by the proposed method is in more close agreement with those given by the viewers, indicating that the proposed method succeeds in accurately reflecting the users' perception. Since the objective methods only consider a simple mapping between the QoS and QoE, most of the predicted MOS values result in a large deviation from the subjective results in the case of the MOS values, indicating a high video quality. For example, almost all the MOS values estimated by the

PSNR-to-MOS mapping method are larger than the subjective results. For the VQM method, when the MOS is lower than 1.5, the differences between the subjective and estimated MOS become fairly low. When increasing the MOS values, the gap widens, and the MOS values of the VQM become significantly lower than those suggested by the subjective results. Furthermore, the RMSEs of the proposed VQM and PSNR-based methods are 0.4629, 0.8025, and 1.7726, respectively. It can be seen that the RMSE of our proposed QoE model is lower than those of the other two methods, which further emphasizes the ability of the proposed system to quantify the grade of user satisfaction in the presence of transmission-induced video impairments.

Conclusions

Reliable quality assessment of video services over wireless networks is essential for optimizing their performance. In this article we first constructed a subjective quality assessment database for characterizing the video impairments imposed by a 3G LTE network simulator. The QoS-related parameters of the simulator are adjustable to meet the requirements. Then a new quality assessment method based on neural networks was proposed, where the weights of the
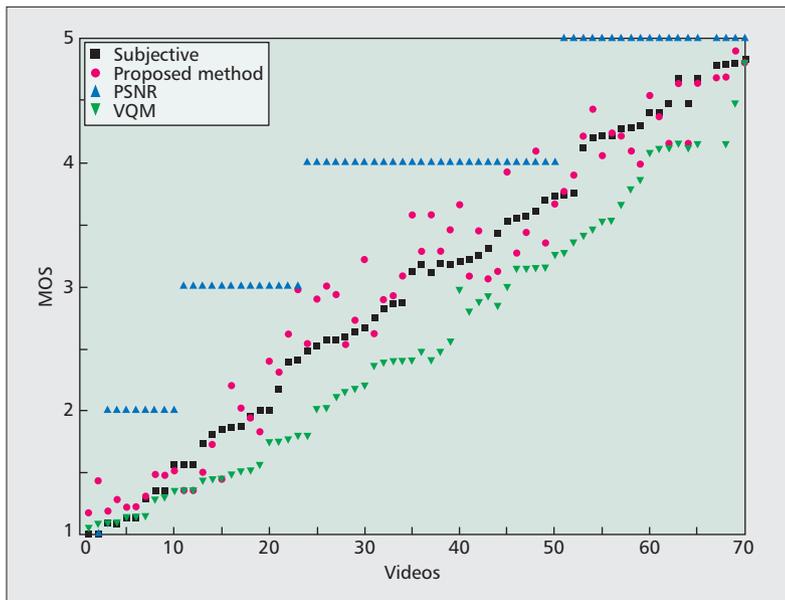
Figure 6. MOS performance comparison among different assessment methods.

model were trained and the match between the objective and subjective results was verified. The parameters of the proposed method were carefully selected with the aid of simulations, considering the trade-off between the attainable performance and the computational complexity imposed. Our method attains better QoE prediction accuracy than the traditional objective assessment methods. Future investigations will be conducted in the context of other multimedia services.

References

[1] ITU-T Recommendation P.10/G.100 (2006) Amendment 1 (01/07), P.10 : New Appendix I — Definition of Quality of Experience (QoE), Jan. 2001.
[2] P. Brooks and B. Hestnes, "User Measures of Quality of Experience: Why Being Objective and Quantitative is Important," *IEEE Network*, vol. 24, no. 2, Mar.–Apr. 2010, pp. 8–13.
[3] ITU-R Recommendation P.800, "Methods for Subjective Determination of Transmission Quality," 1996.
[4] ITU-T Recommendation P.910, "Subjective Video Quality Assessment Methods for Multimedia Applications," International Telecommunication Union, Geneva, Switzerland, 1996.
[5] J.-S. Lee, "On Designing Paired Comparison Experiments for Subjective Multimedia Quality Assessment," *IEEE Trans. Multimedia*, vol. 16, no. 2, 2014, pp. 564–71.
[6] J. Klaue, B. Tathke, and A. Wolosz, "Evalvid— A Framework for Video Transmission and Quality Evaluation," *Proc. Int'l. Conf. Modelling Techniques and Tools for Computer Performance Evaluation*, 2003, pp. 255–72.
[7] Z. Wang *et al.*, "Image Quality Assessment: From Error Visibility to Structural Similarity," *IEEE Trans. Image Process.*, no.13, 2004, pp. 600–12.
[8] T. Wang, A. Pervezand, and H. Zou, "VQM-Based QoS/QoE Mapping for Streaming Video," *Proc. IEEE Int'l*
Conf. Broadband Network and Multimedia Technology (IC-BNMT), 2010, pp. 807–12.
[9] H. Wu-Hsiao and L. Chi-Hsiang, "QoS/QoE Mapping and Adjustment Model in the Cloud-Based Multimedia Infrastructure," *IEEE Syst. J.*, vol. 8, no. 1, 2014, pp. 247–55.
[10] S. Mohamed and G. Rubino, "A Study of Real-Time Packet Video Quality Using Random Neural Networks," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 12, Dec. 2002, pp. 1071–83.
[11] W. Cherif *et al.*, "A_PSQA: PESQ-like Non-Intrusive Tool for QoE Prediction in VoIP Services," *Proc. IEEE Int'l. Conf. Communications* (ICC), June 2012, pp. 2124–28.
[12] J. Kennedy and R. C. Eberhart, "Particle Swarm Optimization," *Proc. IEEE Int'l Conf. Neural Networks*, 1995, pp. 1942–48.
[13] P. Seeling and M. Reisslein, "Video Transport Evaluation with H.264 Video Traces," *IEEE Commun. Surveys Tutorials*, vol. 14, no. 4, 2012, pp. 1142–65.
[14] B. Gorkemil, Y. Sadi, and A. Tekalp, "Effects of MGS Fragmentation, Slice Mode and Extraction Strategies on Performance of SVC with Medium-Grained Scalability," *Proc. IEEE Int'l. Conf. Image Processing*, Sep., 2010, pp. 4201–04.
[15] C.-H. Ke, "myEvalSVC: An Integrated Simulation Framework for Evaluation H.264/SVVC Transmission," *KSII Trans. Internet and Information Systems*, vol. 6, no. 1, Jan 2012.
[16] K. Zheng *et al.*, "Stochastic Performance Analysis of a Wireless Finite-State Markov Channel," *IEEE Trans. Wireless Commun.*, vol. 12, no. 2, Feb., 2013, pp. 782–93.
[17] 3GPP TR 36.814, v2.0.0, "Further Advancements for E-UTRA, Physical Layer Aspects," Mar. 2010.

Biographies

KAN ZHENG [SM'09] received B.S., M.S., and Ph.D. degrees from Beijing University of Posts and Telecommunications (BUPT), China, in 1996, 2000, and 2005, respectively, where he is currently a professor. He has worked in companies including Siemens and Orange Labs R&D. His current research interests are in the field of wireless communications, with an emphasis on resource allocation in heterogeneous networks, M2M networks, and vehicular communication networks. He holds 12 patents, and he has published more than 200 journal and conference papers and books.

XIAOLI ZHANG received her B.S. degree from the School of Information and Communication Engineering, Beijing University of Posts and Telecommunications (BUPT), China, in 2012. She is currently pursing her M.S. degree in the Key Lab of Universal Wireless Communications, Ministry of Education, BUPT. Her research interests include radio resource allocation in wireless networks.

QIANG ZHENG received his B.S. degree from the College of Computer Science and Technology, Shandong University of Technology (SDUT), China, in 2010. He is currently a Ph.D. candidate in the Key Lab of Universal Wireless Communications, Ministry of Education, Beijing University of Posts and Telecommunications (BUPT). His research interests include radio resource allocation, performance analysis, and optimization in heterogeneous vehicular networks.

WEI XIANG [M'04, SM'10] received the B.Eng. and M.Eng. degrees, both in electronic engineering, from the University of Electronic Science and Technology of China, Chengdu, China, in 1997 and 2000, respectively, and the Ph.D. degree in telecommunications engineering from the University of South Australia, Adelaide, Australia, in 2004. Since January 2004 he has been with the School of Mechanical and Electrical Engineering, University of Southern Queensland, Toowoomba, Australia, where he currently holds a faculty post of associate professor. His research interests are in the broad area of communications and information theory, particularly coding and signal processing for multimedia communications systems.

LAJOS HANZO, FREng, FIEEE, FIET, Fellow of EURASIP, DSc, received his degree in electronics in 1976 and his doctorate in 1983. In 2009 he was awarded the honorary doctorate "Doctor Honoris Causa" by the Technical University of Budapest. During his 38-year career in telecommunications he has held various research and academic posts in Hungary, Germany, and the UK. Since 1986 he has been with the School of Electronics and Computer Science, University

of Southampton, UK, where he holds the chair in telecommunications. He has successfully supervised approximately 100 Ph.D. students, co-authored 20 John Wiley/IEEE Press books on mobile radio communications totalling in excess of 10,000 pages, published 1400+ research entries in IEEE Xplore, acted both as TPC and General Chair of IEEE conferences, presented keynote lectures, and has been awarded a number of distinctions. Currently he is directing a 100-member academic research team, working on a range of research projects in the field of wireless multimedia communications sponsored by industry, the Engineering and Physical Sciences Research Council (EPSRC) UK, the European Research Council's Advanced Fellow Grant, and the Royal Society's Wolfson Research Merit Award. He is an enthusiastic supporter of industrial and academic liaison, and he offers a range of industrial courses. He is also a Governor of the IEEE VTS. From 2008 to 2012 he was the Editor-in-Chief of the IEEE Press and a Chaired Professor also at Tsinghua University, Beijing. His research is funded by the European Research Council's Senior Research Fellow Grant.