QoE-Oriented 3D Video Transcoding for Mobile Streaming

YANWEI LIU, Institute of Acoustics, Chinese Academy of Sciences
SONG CI, Institute of Acoustics, Chinese Academy of Sciences, and University of Nebraska-Lincoln
HUI TANG, Institute of Acoustics, Chinese Academy of Sciences
YUN YE, University of Nebraska-Lincoln
JINXIA LIU, Zhejiang Wanli University

With advance in mobile 3D display, mobile 3D video is already enabled by the wireless multimedia networking, and it will be gradually popular since it can make people enjoy the natural 3D experience anywhere and anytime. In current stage, mobile 3D video is generally delivered over the heterogeneous network combined by wired and wireless channels. How to guarantee the optimal 3D visual quality of experience (QoE) for the mobile 3D video streaming is one of the important topics concerned by the service provider. In this article, we propose a QoE-oriented transcoding approach to enhance the quality of mobile 3D video service. By learning the pre-controlled QoE patterns of 3D contents, the proposed 3D visual QoE inferring model can be utilized to regulate the transcoding configurations in real-time according to the feedbacks of network and user-end device information. In the learning stage, we propose a piecewise linear mean opinion score (MOS) interpolation method to further reduce the cumbersome manual work of preparing QoE patterns. Experimental results show that the proposed transcoding approach can provide the adapted 3D stream to the heterogeneous network, and further provide superior QoE performance to the fixed quantization parameter (QP) transcoding and mean squared error (MSE) optimized transcoding for mobile 3D video streaming.

Categories and Subject Descriptors: H.5.1 [**Information Interfaces and Presentation**]: Multimedia Information Systems—*Evaluation/methodology*; I.4.2 [**Image Processing and Computer Vision**]: Compression(Coding)—*Transcoding*

General Terms: Algorithms, Design, Experimentation

Additional Key Words and Phrases: Mobile 3D video, transcoding, QoE, 3D video streaming

ACM Reference Format:

Liu, Y., Ci, S., Tang, H., Ye, Y., and Liu, J. 2012. QoE-oriented 3D video transcoding for mobile streaming. ACM Trans. Multimedia Comput. Commun. Appl. 8, 3s, Article 42 (September 2012), 20 pages.
DOI = 10.1145/2348816.2348821 http://doi.acm.org/10.1145/2348816.2348821

This work was supported in part by Important National Science & Technology Specific Project under contracts 2012ZX03003006-004 and 2011ZX03004-005, NSFC under grant Nos. 60972083 and 61102077, and Initial Fund of President Award of Chinese Academy of Sciences under contract Y129081621.

Authors' addresses: Y. Liu (corresponding author), S. Ci (corresponding author), and H. Tang, Institute of Acoustics, Chinese Academy of Sciences, No. 21, North 4th Ring West Road, Haidian District, Beijing 100190 China; email: liuyw@hpnl.ac.cn, cisong@ieee.org, tangh@hpnl.ac.cn; Y. Ye, University of Nebraska—Lincoln, 200B Peter Kiewit Institute, Omaha, NE 68182; email: yye@huskers.unl.edu; J. Liu (corresponding author), Zhejiang Wanli University, No. 8, South Qian Hu Road, Ningbo, Zhejiang 315100 China; email: liujinxia1969@126.com.

DOI 10.1145/2348816.2348821 http://doi.acm.org/10.1145/2348816.2348821

1. INTRODUCTION

As its very pleased depth effect, 3D video is a new multimedia form with novel visual experience. 3D video can provide the natural vision similar to the human feeling received from two eyeballs. It gradually goes into home by blu-ray disk, broadcasting, and Internet [Zou 2009] so that people can easily experience 3D impression at home. Currently, the broadband wireless communication is rapidly progressing, and the mobile multimedia services with the switchable 2D/3D display are also increasingly attractive. To provide the 3D experiences anywhere and anytime, 3D video is also gradually moving into mobile [Merkle et al. 2009].

Currently, stereoscopic 3D video and texture plus depth based 3D video are the appropriate data formats for 3D video applications. For texture plus depth format, it can provide a certain degree of parallex-shiftable effect in a limited range since depth can assist the receiver to generate the virtual view texture video [Liu et al. 2009]. However, the depth accuracy is not very satisfied in current stage [Vetro et al. 2011] so that the perception quality of generated stereoscopic 3D video is sometimes not well-pleasing. Comparably, the quality of stereoscopic 3D video format can meet most of commercial requirements for glass-based stereoscopic display and therefore it is preferred in the 3D market due to its best tradeoff among quality, simplicity, and easiness to compress. Especially, the frame-compatible stereoscopic 3D video accompanying high definition (HD) format has been adopted by most of commercial 3DTV applications [Vetro et al. 2011].

Though HD 3D video provides the vivid visual effects, it requires much more transmission bandwidth. To exploit the anywhere and anytime 3D viewing experience, 3D video deployed on the mobile device will be also increasingly popular. Since mobile 3D video is mostly transmitted over the wireless network, the limited bandwidth and time-varying characteristics of the wireless network have a great influence on the quality of mobile 3D video service [Liu et al. 2010]. To adapt to the heterogeneous network which consists of the wired and wireless channels, the stereoscopic HD 3D video must be transcoded with rate reduction to adapt to the wireless network, as well as down-sampling spatial resolution to match the mobile screen size [Liu et al. 2012].

In the past years, a lot of transcoding works [Vetro et al. 2005] were proposed to make the originally encoded 2D video source match the network status and terminal capabilities, and further promote the user-end visual experiences. The network adaptation and terminal adaptation are two primary goals for 2D video transcoding. For 3D video application, the virtual view adaptation based transcoding was also proposed to optimize the received video quality [Liu and Chen 2010]. In these works, whether 2D or 3D video transcoding, the common ground is that all of them utilize the mean squared error (MSE) as the adaptation metric to optimize the video quality. From the perspective of video compression, MSE is the very efficient assessment for video quality because it can quantize the signal fidelity by measuring error using an average way. However, when the video is transmitted over the network, the packet loss induced video distortion is random, and also it often has a local effect on the video quality after error concealment [Girod 1993]. Consequently, the image quality measurement in terms of MSE does not always correlate very well with the human visual perception in the presence of packet losses [Wang 2002].

To accurately characterize the user's final perception, the term of quality of experience (QoE) was proposed to evaluate the IPTV service quality. QoE denotes the overall quality perceived subjectively by the end users [ITU-T Rec. G.100 2007]. It is very important to Internet service provider. Unlike quality of service (QoS) providing the network-centric evaluation on service, QoE reflects the user-centric evaluation on the service. QoE is based on the subjective evaluation, and mean opinion score (MOS) is a very appropriate index (indicator) for QoE. For stereoscopic 3D video, the final 3D impression is the fusion result of two view signals with depth perception and thus the objective PSNR evaluation for 3D video does not efficiently match the subjective quality since it only reflects the image qualities of two views. Moreover, the current 2D video QoE measurement is not sufficient to provide the precise assessment for 3D video services. Hence, the 3D visual QoE metric [Jumisko-Pyykkö et al. 2011] with paying attention to depth perception needs to be developed to provide more efficient subjective evaluations.

As the key parts of video service, the source video encoding and networking (delivery) both have contributions to the final QoE. Therefore, mapping the video coding and networking to QoE and then seeking the solution of QoE-oriented video coding or networking is very important for video service. In recent literatures, significant progress has been made to the QoE-aware multimedia networking, such as rate adaptation [Piamrat et al. 2009; Thakolsri et al. 2010], cross-layer resource allocation [Ameigeiras et al. 2010], admission control [Piamrat et al. 2008], video quality adaptation [Khan et al. 2012, Rückert et al. 2012], and content-aware delivery [Worrall et al. 2010]. However, to the best of our knowledge, the previous research works pay little attention to the QoE-oriented video encoding or transcoding, which optimizes QoE from a perspective of source coding. Though some perceptual video coding techniques were proposed to cater for the characteristics of human visual system [Chen et al. 2010a], they did not deal with the networking and user-end feedback in the video encoding side. Hence, the perceptual video coding does not reflect the networking and user-end influence on QoE.

For high-quality mobile 3D video streaming, how to provide the network-adaptive transcoding to guarantee the user's QoE is an urgent problem to be solved. Especially, the current HD 3D video can not perfectly fit for the time-varying wireless network and small size mobile screen due to its large data amount and large image size. And also, the traditional MSE based transcoding can not very efficiently provide the satisfied QoE for the user. Hence, to guarantee the QoE-optimized mobile 3D video streaming over the wireless network, the 3D visual QoE needs to be characterized and further automatically measured in a real-time way to guide the 3D video coding or transcoding.

In this article, we conduct the 3D visual QoE studies for mobile stereoscopic 3D video streaming. Two main contributions of our work are as follows. First, the QoE-oriented 3D video transcoding approach is proposed to optimize the mobile 3D video streaming performance. The proposed approach regulates the transcoding configurations in real-time by inferring the 3D visual QoE with the information feedbacks of network and user-end device. Second, based on the understanding of stereoscopic 3D vision theory, a 3D visual QoE inferring model is summarized to characterize the effect of encoding or transcoding parameters on human QoE in the actual networking environments. Through learning the pre-controlled distortion patterns of 3D contents, QoE inferring for mobile 3D video streaming can be automatically performed. Considering the problem that more distortion patterns for 3D contents are involved in the learning stage than those of 2D video, we propose a piecewise linear MOS interpolation method to reduce the cumbersome manual work of building QoE patterns.

The rest of the article is organized as follows. Section 2 presents the QoE-oriented mobile 3D video transcoding scheme. In Section 3, we first introduce the 3D visual QoE model by extensive subjective tests and then derive the mapping from network-adaptive transcoding parameters to QoE. In Section 4, the transcoding adaptation for the mobile streaming towards optimizing 3D visual QoE is interpreted in detail. Section 5 presents the experimental results. Finally, conclusions are provided in Section 6.

2. QOE-ORIENTED 3D VIDEO TRANSCODING SCHEME

In the Internet, each user's available bandwidth is highly heterogeneous because its last mile connection is mixed with wired and wireless portions. Generally, the bandwidth of wired portion is much higher than that of wireless portion so that we consider the video transmission adaptation problem for the transition from the wired channel to wireless channel. For 3D video streaming or 3D IPTV over the heterogeneous network, 3D video transcoding can provide the adaptive stream to the unsteady wireless network. To guarantee the QoE-optimized transmission over the heterogeneous network, we

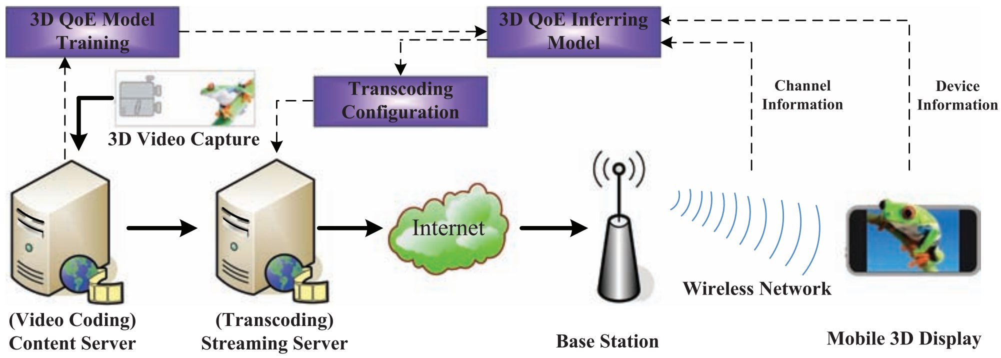


Fig. 1. QoE-oriented 3D video transcoding framework.

propose a 3D video transcoding framework, which consists of the content (encoding) server and the transcoder of the streaming server, as shown in Figure 1.

For stereoscopic 3D video, how to select the appropriate coding parameters for QoE-oriented transcoding with considering the network status is the core of the proposed framework. The content server compresses the captured HD 3D content, and generates the 3D video stream to be transmitted. The QoE model training is also performed with pre-designed transmission distortion patterns for the same HD 3D stream in the content server. The parameters of trained QoE model are transmitted to the transcoder of the streaming server to form the 3D QoE inferring model and then infer the optimal transcoding parameters with the real-time feedback of channel information and device information. According to the QoE-inferred configuration of coding parameters, the transcoder performs network-adaptive transcoding for mobile 3D video streaming. Currently, the transcoder is placed near the content server with a certain delay of real time channel information feedback. To reach a trade-off between the transcoding performance and the processing delay, the transcoder can also be placed at some positions close to the last mile of wireless channel.

The proposed H.264/AVC based 3D video transcoding scheme is shown in Figure 2. In current stage, the inter-view prediction based 3D video decoder is not broadly supported in the mobile devices, and moreover, the simulcast coding has the advantage of easily parallel performing. Hence, we adopt the way of simulcast coding to implement the stereoscopic 3D video transcoding. To decrease the streamed bit-rate, the cascade pixel-domain transcoding architectures are used for the left and right views, respectively.

As shown in Figure 2, the transcoder decodes the 3D bit-stream and fully re-encodes the decoded and spatially down-sampled signals with new QPs. Specifically for one view stream, the decoded picture is firstly spatially down-sampled and then encoded with motion compensated (MC) prediction or intra prediction to generate the prediction residuals. The residuals are discrete cosine transformed (DCT) and quantized with q_2 ($q_{2,l}$ or $q_{2,r}$) to satisfy the new output bit-rate, where $q_{2,l}$ and $q_{2,r}$ denote the quantization parameters of left and right views, respectively. During the stages of different transcoding operations, the optimal parameters can be obtained from the QoE inferring model. Currently, we mainly consider the the down-sample ratios, intra refresh rates and QPs of two views for the QoE-oriented 3D video transcoding, which will be explained in detail in Section. 3. In the transcoding loop, the quantized DCT coefficients can be inversely quantized with $q_{2,l}^{-1}$ or $q_{2,r}^{-1}$, and then inversely discrete cosine transformed (IDCT) to form the reconstructed reference picture to be stored in the frame buffer. The quantized DCT coefficients are variable length coded to generate the bit stream. Finally, the transcoded bit-streams of left and right views are multiplexed into one 3D stream to be transmitted.

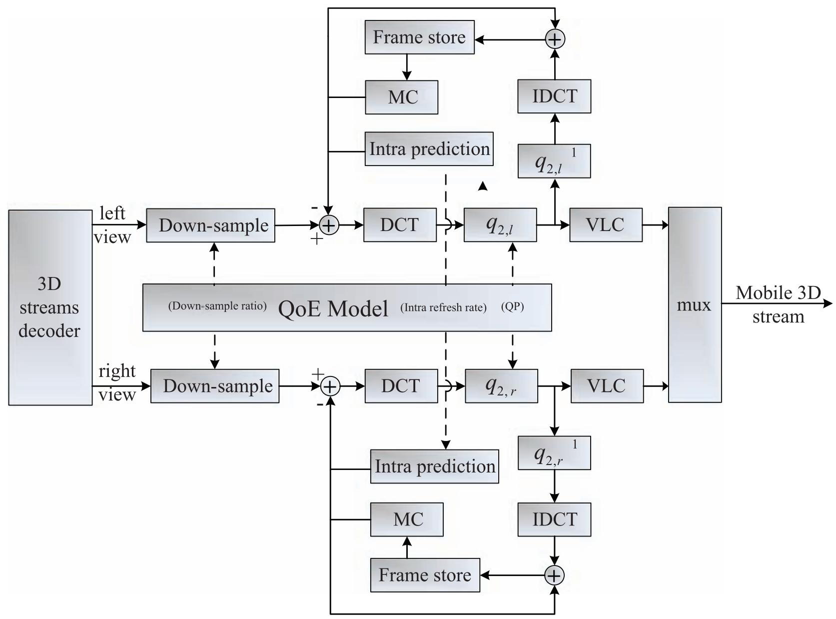


Fig. 2. Cascade 3D video transcoding.

3. 3D VISUAL QOE INFERRING MODEL

As the definition of QoE, we all know that many factors can affect the user's experiences, such as the network, equipment, protocols, terminals, etc. in the entire chain of video communication. Regarding 3D visual QoE, it is a multidimensional conception in which many complex human vision elements are involved and especially the depth perception context and psychological factors are included. The 3D visual QoE dimensions mainly include three key performance indicators, namely the image quality, depth perception, and visual comfort [Seuntïens 2006]. For a determinate user, if the conditions of the receiver terminal are fixed, the dominant factors affecting the 3D visual QoE are server-side processing operations (stereoscopic video content processing) and network-relevant elements. Therefore, for the mobile 3D video streaming over the heterogeneous network, we can analyze the QoE attributes (the parameters affecting QoE) in the transcoding process and establish a 3D QoE Inferring model, to optimally configure the transcoding parameters that have the dominating contributions to the viewing experience of human being. Currently, the focus of our work is on the QoE-oriented transcoding. The aspects of user-side, such as context, device (constraints of screen size, computational power, and battery life-time), display (display-specific artifacts, such as cross-talk and pseudo-stereo), social and psychological factors [Chen et al. 2010b] are not taken into account in the model.

For stereoscopic 3D video, the disparity processing is the main line along the entire 3D video processing chain. Three key performance indicators of visual quality, depth perception and visual comfort are closely linked with the disparity processing. The 3D visual quality depends on the amount of the received and perceived disparity. In terms of the stereo vision theory [Mendiburu 2008], the disparity between two views is the core of 3D perception. The perceived depth for a world point in 3D space is related to its corresponding pixel disparity, which can be computed as Smolic et al. [2011]

$$Z = \frac{B \cdot D}{B - d}, \tag{1}$$

Table I. ITU-R Quality and Impairment Scales

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

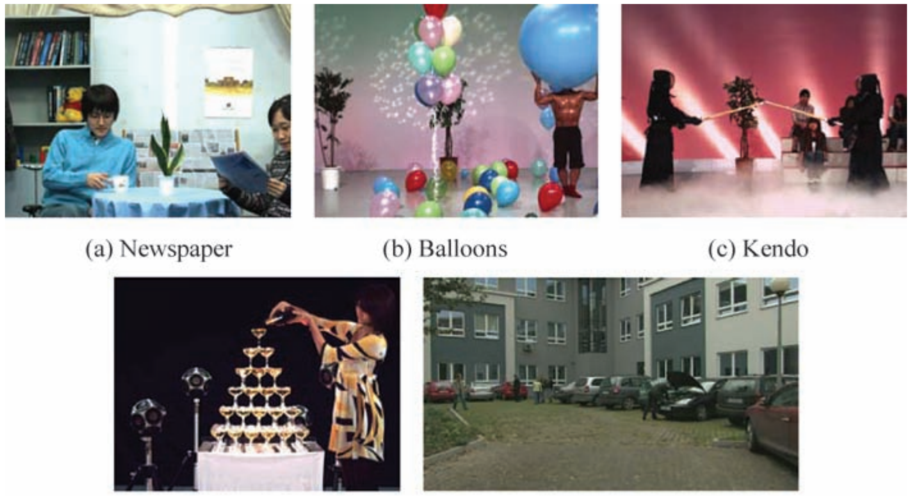
where Z, B, D, d denote the perceived depth, the baseline of two views, the viewing distance to the screen, and the horizontal pixel disparity, respectively. Since the disparity of the corresponding pixel between two view images is usually computed as a position shift to the left of an image pixel when viewed in the right image, it involves the matching of two corresponding pixels in the left and right images. In the end-to-end 3D video chain, any operation that changes the disparity will affect the final 3D visual QoE.

In the network-adaptive transcoding, there are several principal operations modifying the stereo disparities, such as the down-sample, quantization, packet loss, error resilience and concealment. The down-sample scaling operation can change the disparity in the course of reducing pixel samples. The lossy compression also can change (increase or reduce) the disparity because it can bring the mismatch between the corresponding pixels of a stereo pair. The quantization parameters (QP) of the left and right views have a significant impact on the disparity perception. Especially when the asymmetric stereo video coding mode is used, the QP selection of left and right views can affect the final quality of the disparity. The transmission behaviors, such as the transmission packet loss and packet delay also affect the received video quality by losing or changing the disparity. Since the error resilience coding in the transcoding alleviates the error-induced effect on the disparity perception so that it also has a contribution to the final QoE.

To analyze the effects of transcoding parameters on 3D visual QoE, we conduct extensive subjective tests on a 15.6 inch lenticular lenses based stereoscopic display (TOSHIBA Qosmio F750 laptop computer). The viewers sat in front of the screen with comfortable distance and the field of view was about 15^o. The subjective test adopts the SSIS (Single Stimulus Impairment Scale) method described in ITU-R BT. 500 [ITU-R Recommendation BT.500-11 2002], and five grade scales of MOS are used, as shown in Table I. A total of 21 subjects participated in the tests, consisting of 14 males and 7 females. Besides the lenticular lenses display, several glass-based anaglyph color (red-cyan) 3D evaluations are also performed on the general mobile device (such as Apple iPad 2).

Five 3D video sequences of Newspaper [Ho et al. 2008], Champagne, Kendo, Balloons [Tanimoto Lab at Nagoya University 2008], and Poznan_carpark [Domański et al. 2009] are chosen to evaluate QoE-oriented 3D video transcoding and their specific scene contents are shown in Figure 3. Their original spatial resolutions are HD or XGA (1024 × 768, close to HD) and the transcoding target screen sizes of mobile displays are set to 512 × 384, 480 × 360, and 480 × 272 for different sequences. The scene contents cover the indoor people dialog, outdoor sport and show, and different levels of object motions. In addition, these 3D video sequences include the contents captured by not only the static cameras but also moving cameras.

With the statistical analysis of subjective test results, we summarize a transcoding-relevant 3D visual QoE model to guide the network-adaptive transcoding for mobile 3D video streaming. Four prominent parameters affecting 3D visual QoE are currently taken into account in the model. Specifically, we only consider the down-sample ratio, quantization parameter, intra refresh rate, and packet loss rate in the model. The proposed 3D visual QoE model can be expressed as a function of $QoE(\gamma,\ q,\ i,\ \rho)$, where γ, q, i, and ρ denote the down-sample ratio, quantization parameter, intra refresh rate, and packet loss rate, respectively. Since 3D visual QoE is an integrative definition with many influence



(a) Newspaper (b) Balloons (c) Kendo

(d) Champagne (e) Poznan_carpark

Fig. 3. The visual contents of 3D video sequences. (a) Newspaper (©GIST), (b) Balloons, (c) Kendo, (d) Champagne (©Tanimoto Lab at Nagoya University), and (e) Poznan_carpark (©Poznan University of Technology).

factors, besides the above-mentioned four parameters there are other transcoding-related parameters may also contribute to it. If the readers want to investigate the effects of other parameters on the QoE, they can include more parameters in the model.

3.1 Transcoding Parameters Affecting QoE

3.1.1 *Down-Sample Ratio.* The down-sample ratios for two views of stereoscopic 3D video can be both selected to nicely fit for the actual mobile screen size. Of course, we also can select the bigger down-sample ratio for one view to further save the channel bandwidth, and thus form the mix-resolution (asymmetric) stereoscopic 3D video, one of whose views has a lower resolution compared to the other one [Brust et al. 2009]. In terms of binocular suppression theory, it is assumed that the human visual system (HVS) fuses the two views such that the perceived quality is close to that of the higher quality view [Perkins 1992]. Therefore, the down-sample ratios for left and right views can dominate the disparity quality in a certain degree and consequently affect the 3D visual QoE. If the down-sample ratio does not exceed the threshold ratio resulting in QoE slightly deterioration, we can control the down-sample ratios of two views to save the transmission rate, while keeping the acceptable QoE.

Assume that the down-sample ratio exactly matching mobile display size for one view is γ_d (horizontal down-sample ratio multiply vertical down-sample ratio), we define several ratios of γ_d, $\frac{9}{16}\gamma_d$, and $\frac{1}{4}\gamma_d$ for another view. Figure 4 shows the down-sample effect on 3D visual QoE and the corresponding spatial resolutions of the 3D video sequences are shown in Table II. The anaglyph color (red-cyan) 3D effects for different down-sampling ratio combinations of left and right views are shown in Figure 5. In the figure, the 3D depth effect with lower down-sample ratio is obviously superior to the other ones with higher down-sample ratios because the enough object structure and profile information are preserved with the lower down-sample ratio.

3.1.2 *Quantization Parameters.* When the spatial resolution of the stereoscopic video is determined, the QP selection, namely the rate selection has an influence on the final QoE. Thus, the question of what are the best coding QPs for the asymmetric stereo video coding needs to be considered to

Table II. The Spatial Resolutions of the 3D Video Sequences

Sequences	Mobile display size (after down-sample with γ_d)	Mobile display size (after down-sample with $\frac{9}{16}\gamma_d$)	Mobile display size (after down-sample with $\frac{1}{4}\gamma_d$)
Newspaper	512 × 384	384 × 288	256 × 192
Champagne	480 × 360	360 × 240	240 × 180
Kendo	512 × 384	384 × 288	256 × 192
Balloons	512 × 384	384 × 288	256 × 192
Poznan_carpark	480 × 272	360 × 152	120 × 68

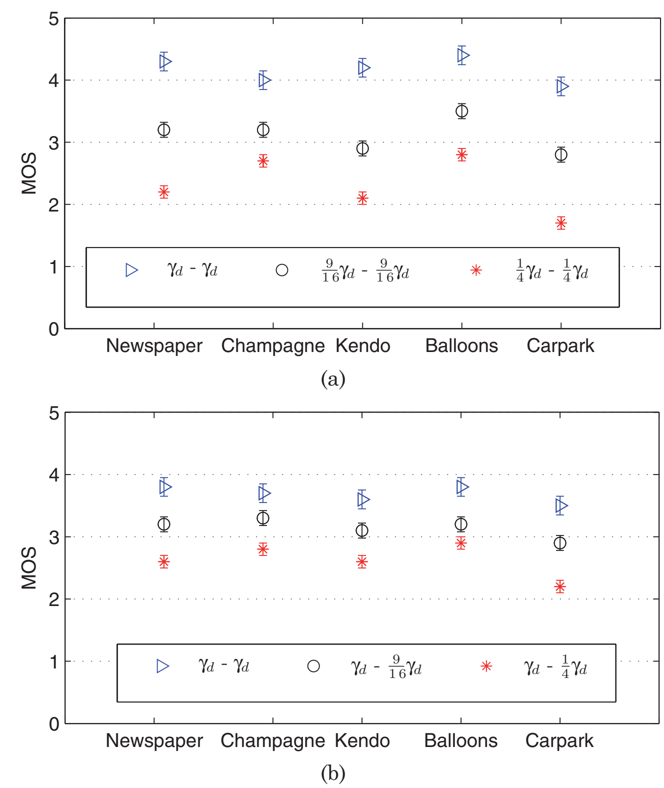


Fig. 4. 3D video quality ratings for different down-sample ratios (95% confidence interval). (a) The uncompressed 3D video with same spatial resolution of left and right views and (b) the compressed 3D video with different spatial resolutions of left and right views for the fixed rate of 800 kbps. (The resolution combination of two views in the figure is represented with the form of "left view down-sample ratio - right view down-sample ratio").

optimize the 3D visual QoE. On the one hand, the coding QP reflects the visual quality impairment levels (compression artifacts) of left and right views, and on the other hand, it also affects the perceived disparity due to the fidelity difference between two views. The disparity directly reflects the depth perception. Hence, we can take QP as a parameter for QoE-oriented transcoding.

Figure 6(a) shows the subjective evaluations for different QP combinations of two views without the total bit-rate constraint. Figure 6(b) shows the subjective evaluations for different QP combinations of two views under the fixed bit-rate constraint of 1150 kbps for two views. The same spatial resolutions for two views are 512 × 384, 512 × 384 and 480 × 272 for Kendo, Newspaper and Poznan_carkpark, respectively. It can be seen from Figure 6 that 3D video quality mainly depends on the compound effects

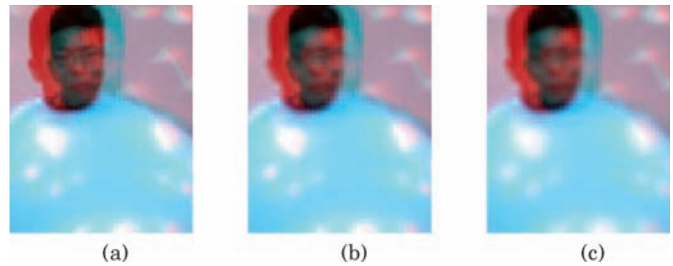


Fig. 5. The subjective anaglyph color (red-cyan) 3D effects for different spatial resolution combinations of left and right views (The part image of the 10th frame for Balloons sequence). (a) left: 512 × 384, right: 512 × 384; (b) left: 512 × 384, right: 256 × 192; (c) left: 256 × 192, right: 256 × 192.

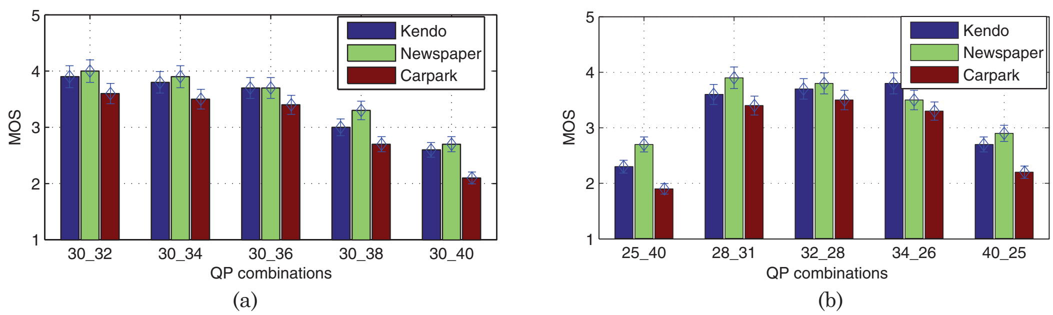


Fig. 6. Subjective evaluations for different QP combinations of two views (95% confidence interval).

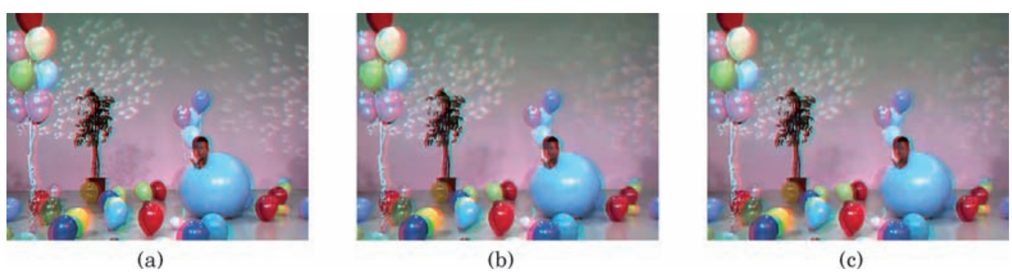
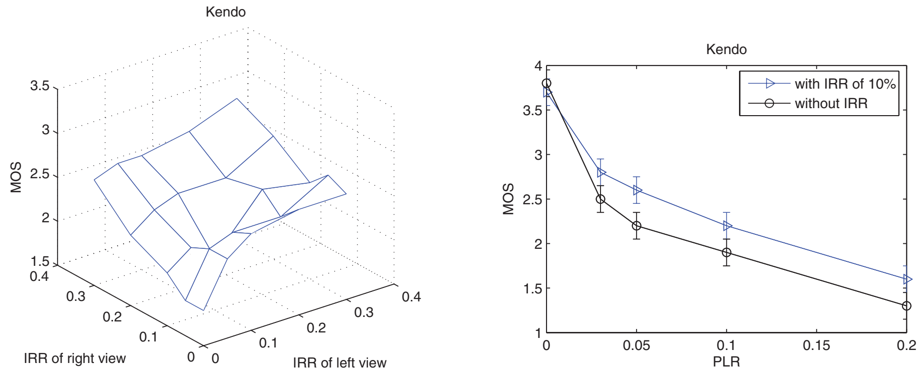


Fig. 7. The subjective anaglyph color (red-cyan) 3D effects for different QP combinations of left and right views (The 230th frame of Balloons sequence). (a) left QP:28, right QP: 28; (b) left QP:28, right QP: 40; (c) left QP:40, right QP: 40.

of the higher quality view's video quality and the QP difference between two views. The subjective anaglyph color (red-cyan) 3D effects for different QP combinations of left and right views for Balloons sequence are shown in Figure 7. It can be seen from the figure that the different visual artifacts occurred in the image of Figure 7(c) so that they result in an inferior 3D visual perception to the other two images of Figures 7(a) and 7(b).

3.1.3 *Intra Refresh Rate.* For video coding, the error resilience technique is often used to alleviate the error influence on video quality. Intra refresh coding is an effective tool to prevent the error propagation. For the inter-coded frames (P frames), only parts of the MBs are encoded as intra mode after the rate distortion optimization. For the successive inter-coded frames, if one frame is split into N regions, forcing the intra coding of one region in one frame, and thus after N frames the whole frame



(a) (b)

Fig. 8. The effect of IRR on QoE (95% confidence interval). (a) 3D video subjective evaluation with different IRRs (PLR of 5% and the rate of 400 kbps for each view are used in the experiment) and (b) 3D video subjective evaluation with IRR of 10%.

intra refreshing will be completed. The more forced intra MBs, the better error-resilience performance can be obtained with the penalization of higher coding bit-rate.

The intra refresh rate (IRR) controls the trade-off between the error-resilience and coding bit-rate. With the limited transmission bandwidth, the different intra refresh rates lead to different perceptual video qualities. Figure 8 shows the MOS values with different intra refresh rates and different packet loss rates (PLRs). In the figure, the candidate intra refresh rates are set to 0, 0.1, 0.2, 0.3, and 0.4. It can be seen from Figure 8 that intra refresh rate is an important encoding parameter which significantly affects the visual quality during the error-resilient video transmission.

3.2 Network-Adaptive Transcoding to QoE Mapping

Since QoE is the subjective perception in nature, MOS is a common metric for the realistic QoE evaluation which involves many actual subjects scoring the experience after watching the video. However, for real-time monitoring or predicting user's QoE under a certain network setting, it is impossible to select many subjects to score the perceived qualities. We must seek a mapping model or prediction model to let the QoE evaluation be automatically performed without human's intervention. In other words, the subjective evaluation must be transformed to the objective procedures to measure the video quality in an automatic, quantitative and repeatable way. Though we found that many factors contribute to the user's 3D visual QoE, it is very difficult to establish a simple and effective function mapping these factors to the QoE. A hybrid assessment tool pseudo subjective quality assessment (PSQA) [Mohamed et al. 2002] has been proposed to cope with the limitations of the objective and subjective evaluations. The main idea of PSQA is first to collect the subjective feedbacks for distorted videos with pre-controlled parameters, and then use the corresponding results to train a random neural network (RNN) to learn the relation between the parameters that cause the distortion and the perceived experience. Since RNN can simulate the real biological nervous system of a human-being, it effectively gives the computed MOS value with an objective manner.

To guarantee the transcoded stream to adapt to the network, the network status needs to be considered in the QoE model. For mobile 3D video streaming, the packet loss behaviors dominate most of the networking influence on the 3D visual QoE. During the streaming, different packet loss patterns have different effects on the QoE. The burst loss and random loss [Liang et al. 2008] are two mainly packet loss patterns for packet video streaming. The random loss usually denotes the random isolated packet loss. Comparably, the burst loss denotes the consecutive packet loss and it has different loss

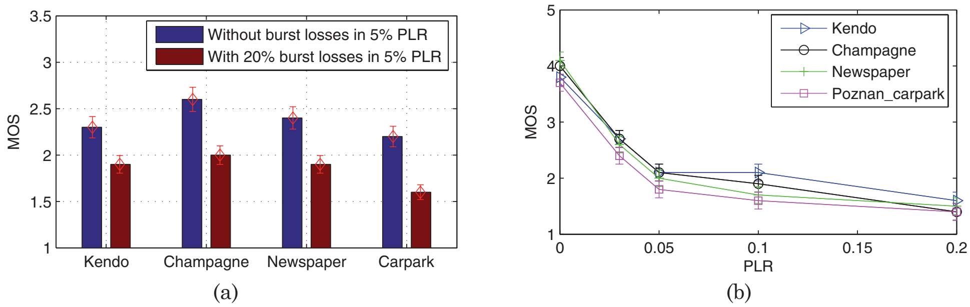


Fig. 9. (a) The burst loss effects on the 3D visual QoE (95% confidence interval). (b) 3D visual QoE with different PLRs (95% confidence interval).

(a) (b) (c)

Fig. 10. Subjective evaluations of anaglyph red-cyan 3D with different PLR combinations of left and right views. (the 213th frame for Balloons sequence). (a) left PLR: 0%, right PLR: 0%; (b) left PLR: 0%, right PLR: 10%; (c) left PLR: 10%, right PLR: 10%.

lengths (the number of continuing lost packets). In the experiment, we set the length of a burst packet loss with one random number between 2 and 5, and the packet loss effects on the QoE are shown in the Figure 9(a). It can be seen from the figure that the distinct loss patterns can result in different degrees of damages on the visual quality. Especially, the error concealment is not very efficient for more successive packet losses, the burst loss makes a compromising perceived realism of 3D. To more accurately assess the QoE, the burst loss rate is also considered in the QoE model. In Figure 9(b), 20% burst losses are included in the total PLR. The comparison result of Figure 9(b) shows that the total packet loss rate and the specific packet loss pattern have a combined impact on the final video QoE.

Figure 10 shows the subjective anaglyph (red-cyan) 3D evaluations with different PLR combinations of left and right views. It can be seen from the figure that the 3D visual quality gradually becomes worse with the increasing PLR. In Figure 10(c), the packet loss artifacts lead to the inferior image quality and specifically the boundary between the object and background becomes very blur due to the error corruption. Consequently, the 3D visual quality also becomes slightly annoying and the content seems slightly unnatural. Moreover, the packet loss induced binocular disparity mess causes a little visual fatigue and destroys the stereopsis.

Except the packet loss, the streaming traffic delay and varying delay (jitter) also degrade the 3D visual QoE. Though the transmission delay does not have any guidance to the QoE-oriented transcoding, they also affect the accuracy of the prediction of final QoE. Currently, the delay and delay variations are implicitly considered in the 3D visual QoE model. Actually, when the playback caching buffer is used, the jitter effect can be almost neglected. And also, the packets arriving later than a maximal threshold can be taken regarded as lost. Hence, the delay effect on the QoE is indirectly included into the influence of packet loss in the QoE model.

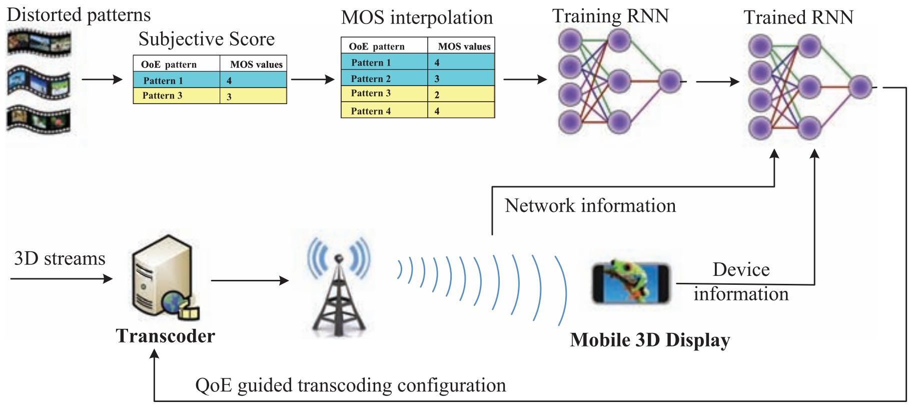
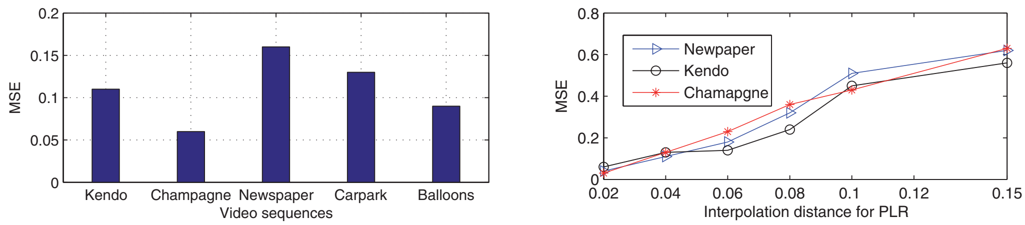


Fig. 11. The mapping from the network-adaptive transcoding to QoE.

To perform the fast transcoding for optimizing QoE, we can train RNN with the pre-controlled QoE patterns in an offline way in the content (encoding) server. Once the RNN training is completed, it can be used repeatedly in the future streaming. Currently, spatial down-sampling is adopted in the transcoding. Thus, the training of RNN for inferring QoE is performed with stereo videos of different down-sample ratios. We utilize several distorted patterns with different combinations of QoE parameters as the input variables of RNN. In the preparation of video distortion patterns, we also record their rate and QP (R-Q) information to form the corresponding R-Q models, which will be used to decide the possible QP range in the transcoding. We invited 21 subjects to perform the subjective tests. The subjective scores of 15 subjects are used to train the RNN, and then the scores of other 6 subjects are used to validate the RNN. After getting the trained RNN, the transcoder can collect the channel and device information as the actual inputs to obtain the optimal transcoding configuration. The specific mapping flow from the network-adaptive transcoding to QoE is shown in Figure 11.

Currently, the RNN training needs many distortion patterns. Especially for stereoscopic 3D video, more distortion cases for the combinations of two views need to be considered. Building these large number of patterns will consume great manpower and material resources. To reduce the cost of RNN training, we proposed a piecewise linear MOS interpolation method to build the distortion patterns. According to the statistical analysis of experimental results, we can take regard the changing trend of one QoE parameter within a limited range as linear when the other parameters keep constant. For example, in Figure 9(b), the MOS values with increasing PLRs are approximately linear changing within the range of [0.1 0.2], and we can interpolate the MOS values of positions between 0.1 and 0.2. Therefore, we can build the sparse distortion patterns with little manpower and then interpolate MOS values of some patterns to increase the density of distortion patterns. Specifically, we first interpolate the MOS values of patterns with changing only one parameter while keeping other parameters fixed, and then interpolate the next parameter pattern with fixing other parameters. When the interpolation for the patterns of all parameters is completed, the number of patterns will be greatly increased. Figure 12(a) shows the overall mean squared errors of the interpolated MOS values compared to the actual MOS values which human subjects give.

Since the QoE training adopts the piecewise linear MOS interpolation method, the complexity for constructing dense patterns is greatly reduced. Currently, the accuracy of the MOS interpolation depends on the interpolation distance which denotes the range of parameter value within which the MOS keeps approximately linear. When the interpolation distance is very large, the linear approximation will result in the non-precise MOS values. Figure 12(b) shows the relationship between the MOS



(a) (b)

Fig. 12. (a) The accuracy of MOS interpolation and (b) MOS interpolation accuracy vs.the interpolation distance for PLR.

Table III. The Summary of Some Notations

Variables	Definitions
W	The channel bandwidth for current 3D video streaming
γ	The candidates set of down-sample ratios
I	The candidates set of intra refresh rate
Q	The candidates set of transcoding QP
Q_l	The transcoding QP of left view
Q_r	The transcoding QP of right view
$R_l(Q_l)$	The transcoding rate of left view
$R_r(Q_r)$	The transcoding rate of right view
γ_l	The down-sample ratio for left view
γ_r	The down-sample ratio for right view
I_l	The intra refresh rate for left view
I_r	The intra refresh rate for right view
ρ	The packet loss rate
ρ_b	The burst loss rate in the total packet losses

interpolation accuracy and the interpolation distance. It can be seen that there exists a threshold of interpolation distance that limits the MOS interpolation accuracy. When the interpolation distance exceeds the threshold, the accuracy of the interpolation is greatly decreased. Therefore, we build the sparse pattern to guarantee the interpolation distance satisfy the accuracy threshold.

4. 3D VIDEO TRANSCODING CONFIGURATION TOWARDS MAXIMIZING QOE

According to the 3D QoE model, the user's QoE can be inferred with the real-time feedback of network status and then the transcoder can dynamically regulate the transcoding parameters to cater for the time-varying network. To facilitate understanding, we summarize some notations in Table III and then present the network-adaptive transcoding configuration procedures. Based on the idea of QoE-oriented transcoding, the network-adaptive QoE optimization under the channel bandwidth W can be mathematically formulated as

$$\left(Q_l^{opt},\ Q_r^{opt}, \gamma_l^{opt}, \gamma_r^{opt}, I_l^{opt}, I_r^{opt}\right) = \underset{\substack{(Q_l \in Q,\ Q_r \in Q,\\ \gamma_l \in \gamma,\ \gamma_r \in \gamma,\\ I_l \in I,\ I_r \in I)}}{\arg\max} \{QoE(Q_l, Q_r, \gamma_l, \gamma_r, I_l, I_r, \rho,\ \rho_b)\}$$
$$\text{subject to } R_l\left(Q_l^{opt}\right) + R_r\left(Q_r^{opt}\right) < W \tag{2}$$

where Q_l^{opt}, Q_r^{opt}, γ_l^{opt}, γ_r^{opt}, I_l^{opt} and I_r^{opt} denote the selected optimal transcoding parameters, and their specific meanings can refer to Table III.

In the transcoding, the appropriate QP range, down-sample ratio range and IRR range which meet the channel limitation need to be firstly decided. Only obtaining the appropriate ranges of parameters

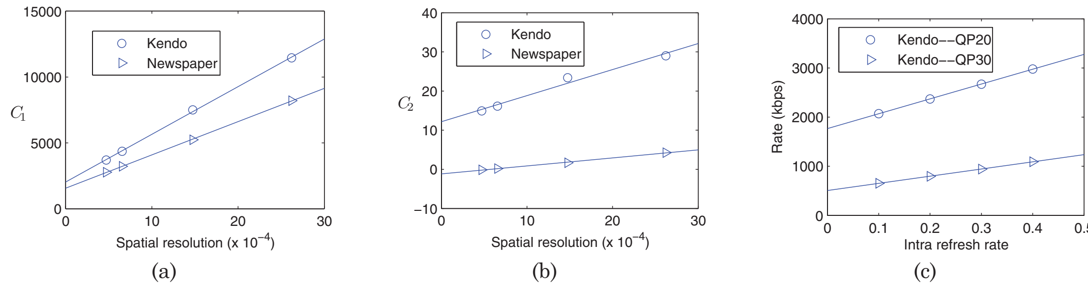


Fig. 13. (a) The linear relationship between C_1 and video spatial resolution; (b) The linear relationship between C_2 and video spatial resolution; (c) The linear relationship between coding bit-rate and intra refresh rate (the fixed coding with QP = 20 and QP = 30).

can the optimal parameters be selected according to (2) for optimizing the QoE. Given the channel bandwidth limitation W, the QP range can be computed in terms of the rate-quantization (R-Q) model. The R-Q models for the videos with different spatial resolutions are different. However, there exists a relationship between the different R-Q models for different spatial resolution videos since these videos are represented with different pixel sampling densities for the same scene content. As we all known, the linear R-Q model (the linear relationship between rate R and reciprocal of the quantization step size Q_{step}) [Ma et al. 2005] for the video with original resolution can be expressed as

$$R = \frac{C_1}{Q_{step}} + C_2, \tag{3}$$

where C_1 and C_2 are constants. Assume the obtained spatial resolution is $s(m, n)$ after down-sampling. Via extensive experiments, we found that the R-Q model for the video with size of $s(m, n)$ can be approximately characterized as,

$$R = \frac{k_1 * s(m, n) + k_2}{Q_{step}} + k_3 * s(m, n) + k_4, \tag{4}$$

where k_1, k_2, k_3 and k_4 are constants, and m and n are the horizontal and vertical image sizes, respectively. According to (3), C_1 and C_2 can be written as

$$C_1 = k_1 * s(m, n) + k_2 \tag{5}$$

and

$$C_2 = k_3 * s(m, n) + k_4. \tag{6}$$

The linear relationship between C_1 and $s(m, n)$, and linear relationship between C_2 and $s(m, n)$ have been verified in Figures 13(a) and 13(b). When the QP is fixed, the intra refreshing also changes the transcoding bit-rate. Especially for low bit-rate transcoding, the intra refresh rate has a major contribution to the coding bit-rate. Statistical analysis of experimental result shows that the coding bit-rate keeps a linear relationship with intra refresh rate for the fixed QP coding, as shown in Figure 13(c). Thus, (4) can be changed as

$$R = \frac{k_1 \cdot g(m, n) + k_2}{Q_{step}} + k_3 \cdot g(m, n) + k_4 + k_5 \cdot I_R + k_6, \tag{7}$$

where k_5 and k_6 are constants, and I_R denotes the intra refresh rate. After merging the constants, the new R-Q model with considering the down-sample ratio and intra refresh rate can be approximately

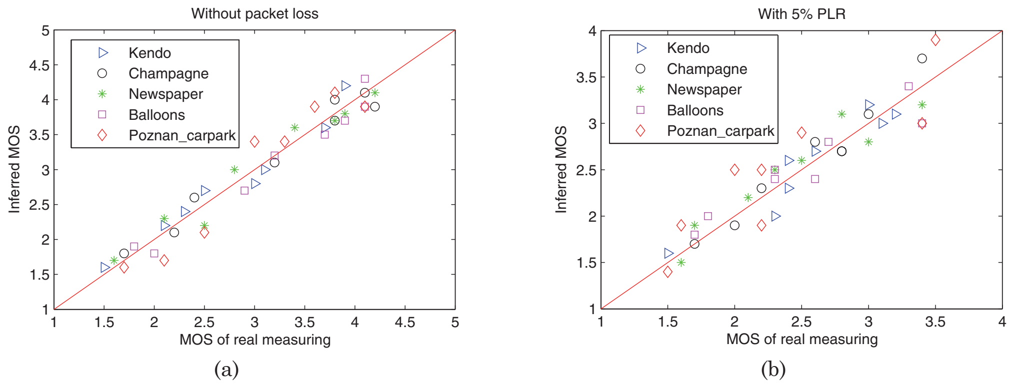


Fig. 14. Inferred MOS vs. realistic measured MOS, (a) without packet loss and (b) with 5% PLR (including 20% burst losses).

expressed as

$$R = \frac{k_1 \cdot g(m,n) + k_2}{Q_{step}} + k_3 \cdot g(m,n) + k_4 \cdot I_R + k_5, \tag{8}$$

where the constants of k_1, k_2, k_3, k_4 and k_5 can be previously obtained and recorded during the period of QoE model training. Hence, the appropriate QP range, down-sample ratio range and intra refresh rate range for each view can be achieved by (8) within the limitation of the channel bandwidth.

Currently, the QP selection is dynamically performed in the unit of group of pictures (GOP) and it has already fully considered the effect of dynamic QP changing between two successive GOPs on the user's QoE. Usually, the QP changing for the next GOP is smoothed by considering the QP value of the previous GOP so that it does not bring the QoE deterioration introduced by the excessive temporal quality variation.

5. EXPERIMENTAL RESULTS

The mobile stereoscopic 3D video streaming with the network topology in Figure 1 is simulated in NS2 using the real networking traffic trace. We offline reproduce the real-time playback effect on 3D display to evaluate the final QoE. In the experiment, it is assumed that only one wireless hop exists in the heterogeneous network topology. Generally, this assumption is reasonable since a wireless link is usually used only in the last hop of the connection in actual network. Since the inter-view prediction based 3D video decoder is currently not very ubiquitous in the practical mobile video players, we adopt the way of simulcast encoding to implement the 3D video transcoding. Based on the H.264/AVC reference software JM16.1, we have implemented a spatial down-sample and rate reduction based 3D video transcoder. The H.264/AVC baseline profile with IPPP coding structure is used because it is suitable for the mobile applications. In the transcoding, the GOP size is set to 30 for error resilience, and the frame rate is set to 25.

In the RNN training stage, 15 users give their MOS values of the pre-controlled 972 training patterns. The piecewise linear MOS interpolation with appropriate interpolation distances of different parameters is used to increase the pattern densities. After MOS interpolation, 8100 MOS values are used to train the RNN. The MOS values of another 6 viewers for the distortion patterns are used to verify and optimize the trained RNN parameters. To evaluate the 3D visual QoE inferring model for mobile 3D video streaming, we select 16 viewers to give their real MOS values to evaluate the accuracy of QoE estimation. Figure 14 shows the scatter plot with inferred MOS vs. real MOS measured from the subjective tests. If the estimated values are close to that of actually measured, the more dots

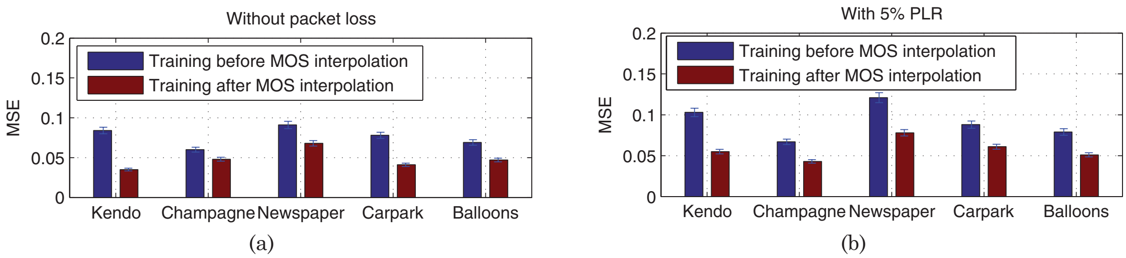


Fig. 15. QoE inferring accuracy comparison between the training before MOS interpolation and that after MOS interpolation.

would be located near the diagonal line in the figure. In terms of this judgment, it can be seen from the figure that the MOS values inferred from the proposed 3D visual QoE model correlate very well with the subjective ratings under the practical transmission environments. In the figures, the overall mean squared errors of QoE inferring for Figures 14(a) and 14(b) are 0.049 and 0.067, respectively. They also verified the precision of QoE estimation.

To verify the efficiency of the proposed piecewise linear MOS interpolation method, the comparison of QoE inferring accuracy between the training before MOS interpolation and that after MOS interpolation is shown in Figure 15. In the figures, the overall mean squared errors of QoE inferring for actual transmission environments are compared for different 3D video sequences. It can be seen that the QoE inferring accuracy can be improved by increasing QoE patterns via piecewise linear MOS interpolation.

Table IV provides the verification of transcoding parameters selection under the channel bandwidth constraint of 800kbps. To evaluate the accuracy of QoE-inferred transcoding parameters, the full-searching selection of transcoding parameters is adopted as the comparison reference. For the full-searching selection, the subjects observe the distorted videos with all possible candidates of the parameters and then select the optimal parameter to obtain the maximal MOS value. Given the transmission status with the packet loss rate of 5% (including 20% burst losses), and the target mobile screen size of 512×384, the precisions of down-sample ratio selection, QP selection, and intra refresh rate selection for both views of a stereoscopic 3D video pair are summarized in Table IV.

In the table, the parameters for left and right views are correspondingly compared with the form of "left view parameter_right view parameter". Since two groups of different QP combinations for two views under one total rate constraint are possible to result in almost the same 3D visual QoE, the QP inferring sometimes can find a little different result from the full searching method. Though the QoE inferring sometimes finds a little different QP combination, it does not decrease the final 3D visual QoE. In other words, it also can make the transcoded and transmitted 3D stream reach the optimal QoE. In the proposed QoE-oriented transcoding, the QoE model training uses the same sequences with different spatial resolutions so that the transcoding parameters can be accurately chosen. In the current experiment, the accuracy of QoE training depends on the abundance of the distortion patterns in a certain degree. Since the distortion patterns after MOS interpolation are enough to cover most of distortion cases in the practical transmission environment, the optimal transcoding parameters are very accurately inferred.

Since our proposed QoE-oriented transcoding can regulate the transcoding configuration parameters in real-time, it can provide the superior streaming performance to the fixed QP transcoding. Figure 16 shows the QoE comparison between the QoE-oriented transcoding and the fixed QP transcoding for mobile 3D video streaming. In the experiment, the fixed QP is carefully selected in line with meeting the channel bandwidth constraint. When the channel bandwidth constraint is time-varying, the fixed

Table IV. The Precisions of QP, Down-Sample Ratio, and Intra Refresh Rate Selections with the Form of "Left View Parameter_Right View Parameter"

Kendo						
Test index	Horizontal image size		QP		Intra refresh rate	
	Inferred	Full searching	Inferred	Full searching	Inferred	Full searching
1	512_512	512_512	22_23	23_23	0.1_0.1	0.1_0.2
2	512_384	512_384	28_28	28_29	0.3_0.2	0.3_0.3
3	512_512	512_512	34_33	34_33	0.4_0.3	0.4_0.3
4	512_256	512_256	38_40	38_40	0.2_0.2	0.2_0.2
5	512_384	512_384	20_25	19_26	0_0.1	0_0.1
Newspaper						
Test index	Horizontal image size		QP		Intra refresh rate	
	Inferred	Full searching	Inferred	Full searching	Inferred	Full searching
1	512_384	512_384	22_28	23_27	0.1_0	0_0.1
2	512_256	512_256	25_30	25_30	0.2_0.3	0.2_0.3
3	512_512	512_512	29_30	29_30	0.4_0.1	0.4_0.1
4	512_384	512_384	33_38	33_37	0.3_0.1	0.3_0.1
5	512_512	512_512	38_40	38_40	0.2_0.1	0.1_0.1
Balloons						
Test index	Horizontal image size		QP		Intra refresh rate	
	Inferred	Full searching	Inferred	Full searching	Inferred	Full searching
1	512_256	512_256	19_30	20_30	0.2_0.2	0.2_0.2
2	512_384	512_384	23_26	23_27	0.1_0.1	0.1_0.1
3	512_512	512_512	28_30	28_30	0.3_0.3	0.2_0.3
4	512_256	512_256	33_41	32_42	0.4_0.3	0.3_0.3
5	512_512	512_512	38_22	38_22	0.1_0.1	0_0.1

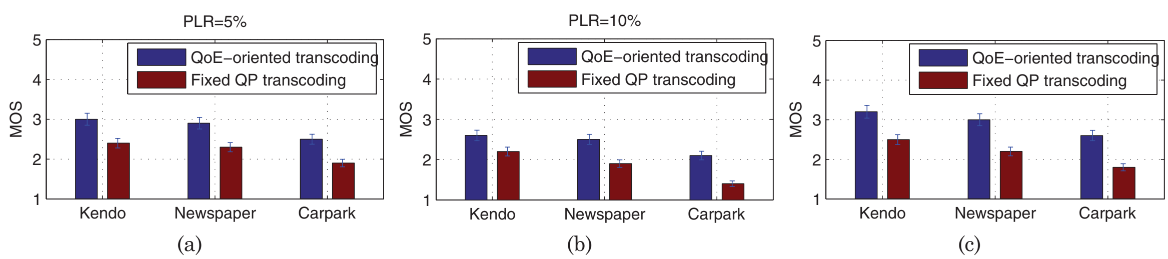


Fig. 16. The transcoding streaming performance comparisons between the QoE-oriented and fixed QP transcodings with (a) PLR = 5% (including 20% burst losses) under fixed rate constraints of 1000kbps, (b) PLR = 10% (including 20% burst losses) under fixed rate constraints of 1000 kbps and (c) time-varying rate constraints and PLRs (including 20% burst losses) (95% confidence interval).

QP is then selected to meet the average value of the varying rates in the entire test period. To extend the viewing time to reach 35 seconds, the test sequences are duplicated with 5 times. Figures 16(a) and 16(b) show the transcoding streaming performances with fixed rate constraint of 1000 kbps and Figure 16(c) shows the transcoding streaming performances with time-varying rate constraints from 500 kbps to 1500 kbps, time-varying PLRs from 0 to 15% (including 20% burst losses in the total packet losses).

Traditional video transcoding is usually based on the MSE or PSNR evaluation [Yin et al. 2003]. The parameters of conventional network-adaptive transcoding are often determined by minimizing the overall video distortion in terms of MSE (MSE-optimized transcoding). Figure 17 shows the performance comparison between the QoE-oriented transcoding, MSE-optimized transcoding, and the fixed

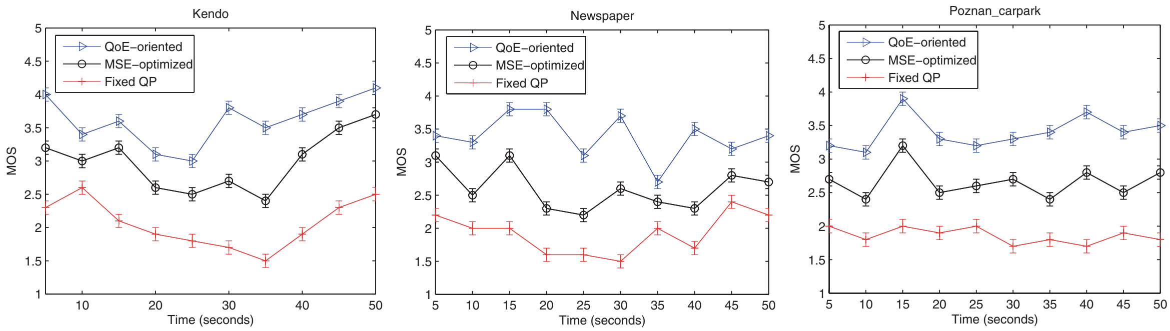


Fig. 17. Performance comparison between the QoE-oriented transcoding, MSE-optimized transcoding and fixed QP transcoding (95% confidence interval).

QP transcoding for mobile 3D video streaming over the dynamic heterogeneous network. In the figure, the test sequence was repeated 10 times to combine into one long sequence for extending the viewing time. At each time-interval of 5 seconds, the realistic QoE is measured. All of QoE-oriented, MSE-optimized, and fixed QP transcodings use the same transmission condition of time-varying rate constraints and time-varying PLRs. In the figures, the streaming performance curves have an obvious decline at the 25th second. According to the packet loss statistics, the packet loss rate is very high and some burst losses are occurred at that period and consequently some frames are frozen at that time for display. From the figure, it can be seen that the MSE-optimized and QoE-oriented transcodings can correspondingly regulate the video quality to alleviate the frame frozen effects. But the fixed QP transcoding can not make any response to the bad network status, so that it results in the worst streaming performance.

During the streaming, the left view and right view are sometimes not synchronized for display due to the packet transmission congestion. Especially, this situation is sometimes very serious for the fixed QP transcoding. Since the QoE-oriented and MSE-optimized transcoding can decrease the spatial resolution and transmitted bit-rate (increasing QP), they slightly mitigate the transmission delay, and consequently avoiding the nonsynchronization of the stereoscopic video to some extent.

6. CONCLUSION

This article has presented an efficient QoE-oriented 3D video transcoding approach for mobile stereoscopic 3D video streaming. Through extensive experimental and statistical analyses, we propose a 3D visual QoE model to characterize the nonlinear relationship between 3D visual QoE and the transcoding-relevant parameters that significantly affect the user's perceived visual experience under the actual network status. Based on the 3D visual QoE model, the proposed QoE-oriented 3D video transcoding dynamically regulates the transmitted 3D video stream in real time to adapt to the network status as well as terminal screen size. Thus, it can alleviate the impacts of bandwidth changing and error-prone characteristics of the time-varying wireless network on the user's QoE. The superior performance of the proposed QoE-oriented 3D video transcoding was demonstrated by extensive experimental results. The QoE improvement was also observed when compared with the conventional fixed QP transcoding and MSE-optimized transcoding for mobile 3D video streaming over the heterogeneous network.

ACKNOWLEDGMENTS

The authors would like to thank Gwangju Institute of Science and Technology for providing Newspaper sequence; Nagoya University for providing the sequences of Champaign, Kendo, and Balloons, and Poznan University of Technology for providing Car Park sequence.

REFERENCES

AMEIGEIRAS, P., RAMOS-MUNOZ, J. J., NAVARRO-ORTIZ, J., AND MOGENSEN, P. 2010. QoE oriented cross-layer design of a resource allocation algorithm in beyond 3G systems. *Elsevier Computer Commun. 33*, 5, 571–582.

BRUST, H., SMOLIC, A., MÜLLER, K., TECH, G., AND WIEGAND, T. 2009. Mixed resolution coding of stereoscopic video for mobile devices. In *Proceedings of 3DTV-CON*. 1–4.

CHEN, Z., LIN, W., AND NGAN, K. N. 2010a. Perceptual video coding: Challenges and approaches. In *Proceedings of IEEE International Conference on Multimedia & Expo*. 19–23.

CHEN, W., FOURNIER, J., BARKOWSKY, M., AND LE CALLET, P. 2010b. New requirements of subjective video quality assessment methodologies for 3DTV. In *Proceedings of VPQM.*

DOMAŃSKI, M., GRAJEK, T., KLIMASZEWSKI, K., KURC, M., STANKIEWICZ, O., STANKOWSKI, J., AND WEGNER, K. 2009. Poznan multiview video test sequences and camera parameters. ISO/IEC JTC1/SC29/WG11, MPEG 2009/M17050.

GIROD, B. 1993. What's wrong with mean-squared error. In *Digital Images and Human Vision*. The MIT Press, 207–220.

HO, Y.-S., LEE, E.-K., AND LEE, C. 2008. Multiview video test sequence and camera parameters. ISO/IEC JTC1/SC29/WG11, MPEG2008/m15419.

ITU-R. 2002. *Methodology for the Subjective Assessment of the Quality of Television Pictures*. Recommendation BT.500-11.

ITU-T. 2007. *New appendix I -Definition of quality of experience (QoE).* Recommentation G.100/P.10 Amendment 1.

JUMISKO-PYYKKÖ, S., HAUSTOLA, T., BOEV, A., AND GOTCHEV, A. 2011. Subjective evaluation of mobile 3D video content: Depth range versus compression artifacts. *Proc. SPIE*, vol. 7881, 78810C.

KHAN, A., SUN, L., AND IFEACHOR, E. 2012. QoE Prediction Model and its Application in Video Quality Adaptation over UMTS Networks. *IEEE Trans. Multimedia 14*, 2, 431–442.

LIANG, Y. J., APOSTOLOPOULOS, J. G., AND GIROD, B. 2008. Analysis of packet loss for compressed video: effect of burst losses and correlation between error frames. *IEEE Trans. Circ. Syst. Video Tech. 18*, 7, 861–874.

LIU, S. AND CHEN, C. W. 2010. 3D video transcoding for virtual views. In *Proceedings of ACM Multimedia*. 795–798.

LIU, Y., HUANG, Q., MA, S., ZHAO, D., AND GAO, W. 2009. Joint video/depth rate allocation for 3D video coding based on view synthesis distortion model. *Signal Process. Image Commun. 24*, 8, 666–681.

LIU, Y., PENG, G., HU, Y., CI, S., AND TANG, H. 2010. A multi-pass VBR rate control method for video plus depth based mobile 3D video coding. In *Proceedings of 11th Pacific-Rim Conference on Multimedia (PCM'10)*.

LIU, Y., CI, S., TANG, H., AND YE, Y. 2012. *Application-Adapted Mobile 3D Video Coding and Streaming—A Survey. 3D Res. J. 3*, 01(2012)5, Springer.

MOHAMED, S. AND RUBINO, G. 2002. A study of real-time packet video quality using random neural networks. *IEEE Trans. Circ. Syst. Video Tech. 12*, 12, 1071–1083.

MA, S., GAO, W., AND LU, Y. 2005. Rate-distortion analysis for H.264/AVC video coding and its application to rate control. *IEEE Trans. Circ. Syst. Video Tech. 15*, 12, 1533–1544.

MENDIBURU, B. 2008. *3D Movie Making-Stereoscopic Digital Cinema from Script to Screen.* Elsevier, New York, 2008978-0-240-81137-6.

MERKLE, P., WANG, Y., MÜLLER, K., SMOLIC, A., AND WIEGAND, T. 2009. Video plus depth compression for mobile 3D services. In *Proceedings of IEEE 3DTV Conference*. Potsdam, Germany.

PIAMRAT, K., KSENTINI, A., BONNIN, J.-M., AND VIHO, C. 2009. Q-DRAM: QoE based dynamic rate adaptation mechanism for multicast in wireless networks. In *Proceedings of IEEE GLOBECOM.*

PERKINS, M. G. 1992. Data compression of stereopairs. *IEEE Trans. Comm. 40*, 4, 684–696.

PIAMRAT, K., KSENTINI, A., VIHO, C., AND BONNIN, J.-M. 2008. QoE-aware admission control for multimedia applications in IEEE 802.11 Wireless Networks. In *Proceedings of IEEE VTC*.

RÜCKERT, J., ABBOUD, O., ZINNER, T., STEINMETZ, R., AND HAUSHEER, D. 2012. Quality adaptation in P2P video streaming based on objective QoE metrics. IFIP Networking.

SEUNTÏENS, P. J. H. 2006. Visual experience of 3DTV. Ph.D. Thesis, Technische Universiteit Eindhoven.

SMOLIC, A., KAUFF, P., KNORR, S., HORNUNG, A., KUNTER, M., MÜLLER, M., AND LANG, M. 2011. Three-dimensional video postproduction and processing. *Proc. IEEE 99*, 4, 607–625.

THAKOLSRI, S., KELLERER, W., AND STEINBACH, E. 2010. QoE-based rate adaptation scheme selection for resource-constrained wireless video transmission. In *Proceedings of ACM Multimedia*. 783–786.

TANIMOTO LAB AT NAGOYA UNIVERSITY. 2008. http://www.tanimoto.nuee.nagoya-u.ac.jp/.

VETRO, A., XIN, J., AND SUN, H. 2005. Error resilience video transcoding for wireless communications. *IEEE Wirel. Comm. 12*, 4, 14–21.

VETRO, A., TOURAPIS, A. M., MÜLLER K., AND CHEN, T. 2011. 3D-TV Content Storage and Transmission. *IEEE Trans. Broadcast. 57*, 2, part 2, 348–394.

WANG, Z., BOVIK, A. C., AND LU, L. 2002. Why is image quality assessment so difficult?. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*.

WORRALL, S. T., KONDOZ, A. M., DRIESNACK, D., TEKALP, M., KOVACS, P., ADARI, T., AND GOKMEN, H. 2010. DIOMEDES: Content Aware Delivery of 3D Media Using P2P and DVB-T2. 2010 NEM Summit: Towards Future Media Internet, Barcelona, Spain.

YIN, P., VETRO, A., XIA, M., AND LIU, B. 2003. Rate-distortion models for video transcoding. In *Proceedings of the Conference on Image and Video Communications and Processing*. 479–488.

ZOU, W. 2009. An overview for developing end-to-end standards for 3-D TV in the home. *Inf. Display, 25*, 7, 14–19.

Received January 2012; revised April 2012; accepted May 2012