Video QoE Models for the Compute Continuum

Lark Kwon Choi[1, 2], Yiting Liao[1], and Alan C. Bovik[2]
[1]Intel Labs, Intel Corporation, Hillsboro, OR, USA
[2]The University of Texas at Austin, Austin, TX, USA
{*lark.kwon.choi, yiting.liao*}*@intel.com, bovik@ece.utexas.edu*

1. Introduction

Video traffic is exponentially increasing over wireless networks due to proliferating video technology and the growing desire for anytime, anywhere access to video content. Cisco predicts that two-thirds of the world's mobile data traffic will be video by 2017 [1]. This imposes significant challenges for managing video traffic efficiently to ensure an acceptable quality of experience (QoE) for the end user. Since network throughput based video adaptation without considering user's QoE could lead to either a bad video service or unnecessary bandwidth waste, QoE management under cost constraints is the key to satisfying consumers and monetizing services [2].

One of the most challenging problems that needs to be addressed to enable video QoE management is the lack of automatic video quality assessment (VQA) tools that estimate perceptual video quality across multiple devices [2]. Researchers have performed various subjective studies to understand essential factors that impact video quality by analyzing compression or transmission artifacts [3], and by exploring dynamic time varying distortions [4]. Furthermore, some VQA models have been developed based on content complexity [5] [6]. In spite of these contributions, user QoE estimation across multiple devices and content characteristics, however, remains poorly understood.

Towards achieving high QoE across the compute continuum, we present recent efforts on automatically estimating QoE via a content and device-based mapping algorithm. In addition, we investigate temporal masking effects and describe a new dynamic system model of time varying subjective quality that captures temporal aspects of QoE. Finally, we introduce potential applications of video QoE metrics, such as quality driven dynamic adaptive streaming over HTTP (DASH) and quality-optimized transcoding services.

2. Improving VQA model for better QoE

VQA models can be generally divided into three broad categories: full-reference (FR), reduced-reference (RR), and no-reference (NR). Some representative high performing algorithms include: MultiScale-Structural SIMilarity index (MS-SSIM) [7] which quantizes "perceptual fidelity" of image structure; Video Quality Metric (VQM) [5] which uses easily computed visual features; Motion-based Video Integrity Evaluation (MOVIE) [8] which uses a model of extra-cortical motion processing; Video Reduced Reference spatio-temporal Entropic Differencing (V-RRED) [9] which exploits a temporal natural video statistics model; and Video BLIINDS [10] which uses a spatio-temporal model of DCT coefficient statistics and a motion coherence model.

The success of the above VQA metrics suggests that disruptions of natural scene statistics (NSS) can be used to detect irregularities in distorted videos. Likewise, modeling perceptual process at the retina, primary visual cortex, and extra-striate cortical areas are crucial to understanding and predicting perceptual video quality [11].

In addition, the quality of a given video may be perceived differently according to viewing distance or display size. Similarly, the visibility of local distortions can be masked by spatial textures or large coherent temporal motions of a video content. In this regard, modern VQA models might be improved by taking into account content and device characteristics. This raises the need to understand QoE for video streaming services across multiple devices, thereby to improve VQA models of QoE across the compute continuum.

3. Achieving high QoE for the compute continuum

How compression, content, and devices interact

To investigate perceived video quality as a function of compression (bitrate and resolution), video characteristics (spatial detail and motion), and display device (display resolution and size), we executed an extensive subjective study and designed an automatic QoE estimator to predict subjective quality under these different impact factors [2].

Fourteen source videos with a wide range of spatial complexity and motion levels were used for the study. They are in a 4:2:0 format with a 1920 × 1080 resolution. Most videos are 10~15 second long, except Aspen Leaves (4s). To obtain a desired range of video quality, the encoding bitrate and resolution sets for each video were chosen to widely range from 110kbps at 448 × 252 to 6Mbps at 1920 × 1080 based on assumed realistic video content and display devices. 80 and 96 compressed videos were displayed on a 42 inch HDTV and four mobile devices (TFT tablet, Amoled phone, Retina tablet, and Retina phone), respectively, and about 30 participants were recruited for each device to rate the videos by recording opinion score using the single-stimulus continuous quality evaluation (SSCQE) [12] method.

MS-SSIM was used since it delivers excellent quality predictions and is faster than MOVIE or VQM. Figure 1 shows the plots of MS-SSIM against MOS for each device along with the best least-squares linear fit. The Pearson linear correlation coefficient (LCC) between MS-SSIM and MOS is 0.7234 for all data points, while LCC using device-based mapping is 0.8539, 0.8740, 0.7989, 0.8329, and 0.8169 for HDTV, TFT tablet, Amoled phone, Retina tablet, and Retina phone, respectively. Furthermore, device and content-specific mapping between MS-SSIM and MOS shows very high LCC (mean: ~ 0.98) as illustrated in Figure 2. To validate the proposed methods on a different VQA database (DB), we also analyzed the models using the TUM VQA DB [13]. LCC between MS-SSIM and MOS using the device and content-specific mapping for TUM VQA DB shows similar results (mean: ~0.98, standard deviation: 0.016). Results indicate that human perception of video quality is strongly impacted by device and content characteristics, suggesting that device and content-based mapping could greatly improve the prediction accuracy of video quality prediction models.

We then designed a MOS estimator to predict perceptual quality based on MS-SSIM, a content analyzer (spatial detail (S), motion level (M)), a device detector (display type (D), and resolution (R)) [3]. The predicted MOS is calculated as,

$$e\_MOS = \alpha \times \text{MS-SSIM} + \beta \qquad (1)$$

where α and β are functions of the four impact factors S, M, D, and R above. Using the proposed predictor and estimated values of α and β, LCC between the estimated MOS and actual MOS is 0.9861. Future work includes building a regression model to calculate α and β based on the impact factors and extending the video data set to better validate the designed predictor.

Temporal masking and time varying quality

The visibility of temporal distortions influences video QoE. Salient local changes in luminance, hue, shape or size become undetectable in the presence of large coherent object motions [14]. This "motion silencing" implies that large coherent motion can dramatically alter the visibility of visual changes/distortions in video. To understand why it happens and how it affects QoE, we have developed a spatio-temporal flicker detector model based on a model of cortical simple cell responses [15]. It accurately captures the observers'

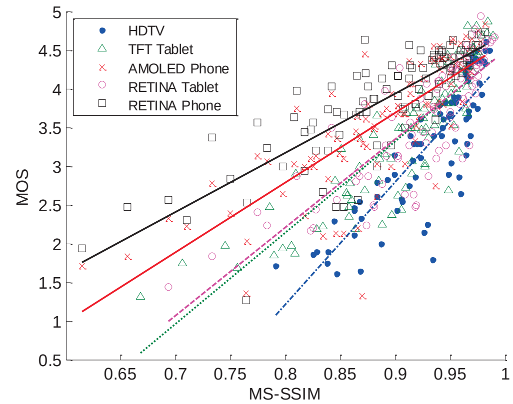


Figure 1 Device-based MS-SSIM and MOS

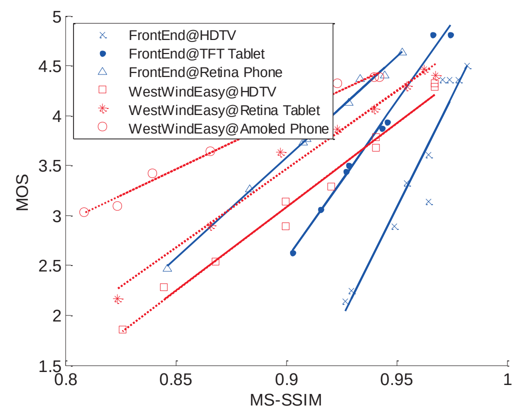


Figure 2 Device and content-based MS-SSIM and MOS mapping

perception of motion silencing as a function of object motion and local changes. In addition, we have investigated the impact of coherent object motion on the visibility of flicker distortions in naturalistic videos. The result of a human experiment involving 43 subjects revealed that the visibility of flicker distortions strongly depends on the speed of coherent motion. We found that less flicker was seen on fast-moving objects even if observers held their gaze on the moving objects [16]. Results indicate that large coherent motion near gaze points masks or 'silences' the perception of temporal flicker distortions in naturalistic videos, in agreement with a recently observed motion silencing effect [14].

Time varying video quality has a definite impact on human judgment of QoE. Although recently developed HTTP-based video streaming technology enables flexible rate adaptation in varying channel conditions, the prediction of a user's QoE when viewing a rate adaptive HTTP video stream is not well understood. To solve this problem, Chao *et al.* have proposed a dynamic system model for predicting the time varying subjective quality (TVSQ) of rate adaptive videos [17]. The model first captures perceptual relevant spatio-temporal features of the video by measuring short time subjective quality using a high-performance RR VQA model called V-RRED [9], and then employs a Hammerstein-Wiener model to estimate the hysteresis effects in human behavioral responses. To validate the model, a video database including 250 second long time varying distortions was constructed and TVSQ was measured via a subjective study. Experimental results show that the proposed model reliably tracks the TVSQ of video sequences exhibiting time-varying level of video quality. The predicted TVSQ could be used to guide online rate-adaptation strategies towards maximizing the QoE of video streaming services.

Application of video QoE models

Recently developed QoE models open up opportunities to improve cooperation between different ecosystem players in end-to-end video delivery systems, and to deliver high QoE using the least amount of network resources. We have shown that in an adaptive streaming system, DASH clients can utilize quality information to improve streaming efficiency [18]. The quality-driven rate adaptation algorithm jointly optimizes video quality, bitrate consumption, and buffer level to minimize quality fluctuations and inefficient usage of bandwidth, thus achieving better QoE than bitrate-based approaches [18]. Another usage of the QoE model is to allow transcoding services to determine the proper transcoding quality on a content-aware and device-aware fashion. The QoE metric helps the transcoder to achieve the desired QoE without over consuming bandwidth. Furthermore, content-specific and device-specific video quality information may facilitate service providers to design more advanced multi-user resource allocation strategies to optimize overall network utilization and ensure a good QoE for each end user.

Acknowledgment

The authors thank Philip Corriveau and Audrey Younkin for collaboration on the subjective video quality testing.

References

[1] Cisco Systems, Inc., “Cisco visual networking index: Global mobile data traffic forecast update, 2012-2017,” Feb. 2013.

[2] Y. Liao, A. Younkin, J. Foerster, and P. Corriveau, “Achieving high QoE across the compute continuum: How compression, content, and devices interact,” in *Proc. of VPQM*, 2013.

[3] K. Seshadrinathan, R. Soundararajan, A.C. Bovik, and L. K. Cormack, “Study of subjective and objective quality assessment of video,” *IEEE Trans. Image Process.*, vol.19, no.6, pp.1427-1441, June 2010.

[4] K. Moorthy, L. K. Choi, A. C. Bovik, and G. de Veciana, “Video quality assessment on mobile devices: Subjective, behavioral and objective studies,” *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 6, pp. 652-671, Oct. 2012.

[5] M. H. Pinson and S. Wolf, “A new standardized method for objectively measuring video quality,” *IEEE Trans. Broadcast.*, vol. 50, no. 3, pp. 312-322, Sep. 2004.

[6] J. Korhonen and J. You, “Improving objective video quality assessment with content analysis,” in *Proc. of VPQM*, 2010.

[7] Z. Wang, E. Simoncelli, and A.C. Bovik, “Multi-scale structural similarity for image quality assessment,” *Ann. Asilomar Conf. Signals, Syst., Comput.*, 2003.

[8] K. Seshadrinathan and A.C. Bovik, “Motion-tuned spatio-temporal quality assessment of natural videos,” *IEEE Trans. Image Process*, vol. 19, no. 2, pp. 335-350, Feb. 2010.

[9] Soundararajan, and A. C. Bovik, “Video quality assessment by reduced reference spatio-temporal entropic differencing,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 4, pp. 684-694, Apr. 2013.

[10] M. Saad and A.C. Bovik, “Blind quality assessment of videos using a model of natural scene statistics and motion coherency,” Invited Paper, *Ann Asilomar Conf Signals, Syst, Comput.*, 2012.

[11] A.C. Bovik, “Automatic prediction of perceptual image and video quality,” *Proceedings of the IEEE*, to appear, 2013,

[12] ITU-R Rec. BT. 500-11, “Methodology for the subjective assessment of the quality of television pictures,” 2002.

[13] C. Keimel, A.Redl and K. Diepold, “The TUM high definition video data sets,” in *Proc. of QoMEX*, 2012.

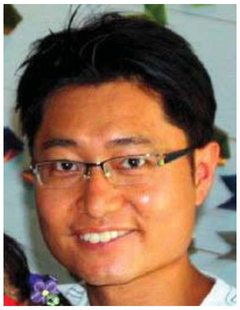
[14] J. W. Suchow and G. A. Alvarez, “Motion silences awareness of visual change,” *Curr. Biol*., vol. 21, no. 2, pp.140-143, Jan. 2011.

[15] L. K. Choi, A. C. Bovik and L. K. Cormack, “A flicker detector model of the motion silencing illusion,” *Ann. Mtng. Vision Sci. Soc.*, 2012.

[16] L. K. Choi, L. K. Cormack, and A. C. Bovik, “On the visibility of flicker distortions in naturalistic videos,” in *Proc. of QoMEX*, 2013.

[17] C. Chen, L. K. Choi, G. de Veciana, C. Caramanis, R. W. Heath Jr, A. C. Bovik, “A dynamic system model of time-varying subjective quality of video streams over HTTP,” in *Proc. of ICASSP*, 2013.

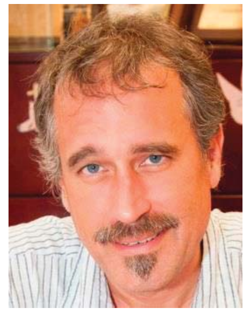
[18] Y. Liao, J. Foerster, O. Oyman, M. Rehan, Y. Hassan, “Experiment results of quality driven DASH,” *ISO/IEC JTC1/SC29/WG11*, M29247, 2013.



Lark Kwon Choi received his B.S. degree in Electrical Engineering from Korea University, Seoul, Korea, in 2002, and the M.S. degree in Electrical Engineering and Computer Science from Seoul National University, Seoul, Korea, in 2004, respectively. He has worked for KT (formerly Korea Telecom) as a senior engineer from 2004 to 2009 on IPTV platform research and development. He has contributed on IPTV standardization in International Telecommunication Union (ITU-T) and Telecommunications Technology Association (TTA). He is currently pursuing his Ph.D. degree as a member of the Laboratory for Image and Video Engineering (LIVE) at The University of Texas at Austin under Dr. Alan. C. Bovik's supervision. His research interests include image and video quality assessment, spatial and temporal visual masking, video QoE, and motion perception.

Yiting Liao received the B.E. and M.S. degrees in Electronic Engineering from Tsinghua University, China in 2004 and 2006, respectively, and the Ph.D. degree in Electrical and Computer Engineering from the University of California, Santa Barbara, in 2011.

She is currently a Research Scientist with Intel Corporation, Hillsboro, OR. Her current research interests include video quality assessment and user experience evaluation, and video optimization techniques over wireless networks.



Alan C. Bovik (S'80–M'81–SM'89–F'96) holds the Keys and Joan Curry/Cullen Trust Endowed Chair at The University of Texas at Austin, where he is a Professor in the Department of Electrical and Computer Engineering and in the Institute for Neuroscience, and Director of the Laboratory for Image and Video Engineering (LIVE).

He has received a number of major awards from the IEEE Signal Processing Society, including: the Best Paper Award (2009); the Education Award (2008); the Technical Achievement Award (2005), the Distinguished Lecturer Award (2000); and the Meritorious Service Award (1998). Recently he was named Honorary Member of IS&T (2012) and received the SPIE Technology Achievement Award (2012). He was also the IS&T/SPIE Imaging Scientist of the Year for 2011.

Professor Bovik has served in many and various capacities, including Board of Governors, IEEE Signal Processing Society, 1996-1998; Editor-in-Chief, *IEEE Transactions on Image Processing*, 1996-2002; Overview Editor, *IEEE Transactions on Image Processing*, 2009-present; Editorial Board, The Proceedings of the IEEE, 1998-2004; Senior Editorial Board, *IEEE Journal on Special Topics in Signal Processing*, 2005-2009; Associate Editor, IEEE Signal Processing Letters, 1993-1995; Associate Editor, *IEEE Transactions on Signal Processing*, 1989-1993; He founded and served as the first General Chairman of the *IEEE International Conference on Image Processing*, held in Austin, Texas, in November, 1994.

Dr. Bovik is a registered Professional Engineer in the State of Texas and is a frequent consultant to legal, industrial and academic institutions.