IEICE TRANS. COMMUN., VOL.E102–B, NO.7 JULY 2019

INVITED PAPER *Special Section on Communication Technologies and Service Qualities in Various Access Networks*

Methods for Adaptive Video Streaming and Picture Quality Assessment to Improve QoS/QoE Performances

Kenji KANAI[†a)], ***Member***, **Bo WEI**[†b)], **Zhengxue CHENG**[†c)], ***Student Members***, **Masaru TAKEUCHI**[†d)], ***Member***, ***and*** **Jiro KATTO**[†e)], ***Fellow***

SUMMARY This paper introduces recent trends in video streaming and four methods proposed by the authors for video streaming. Video traffic dominates the Internet as seen in current trends, and new visual contents such as UHD and 360-degree movies are being delivered. MPEG-DASH has become popular for adaptive video streaming, and machine learning techniques are being introduced in several parts of video streaming. Along with these research trends, the authors also tried four methods: route navigation, throughput prediction, image quality assessment, and perceptual video streaming. These methods contribute to improving QoS/QoE performance and reducing power consumption and storage size.

key words: *video streaming, picture quality assessment, MPEG-DASH, machine learning*

1. Introduction

As pointed out by many researchers and engineers, Cisco forecasts Internet IP video traffic to reach more than 80% of the total traffic by 2021 [1]. Richer video content, ultra high-definition (UHD), 360-degree, and virtual reality/augmented reality (VR/AR) have been supported by large-scale video delivery services such as YouTube and Netflix. It was announced that Netflix and Amazon Prime subscribers worldwide crossed the 100 million mark.

Additionally, new technologies for video streaming are evolving. MPEG-DASH (Dynamic Adaptive Streaming over HTTP) was standardized in 2013 for supporting adaptive video streaming [2], [3]. In MPEG-DASH, video contents are encoded by multiple bitrate-resolution pairs, called "representations," and adaptive streaming is carried out by clients' selection of one representation according to their network conditions or capability of the receiving device. Furthermore, various machine learning techniques have been introduced into video streaming for improving/estimating streaming performance and picture quality [4]. Streaming quality prediction and image quality assessment are examples of the aforementioned methods, and the number of research papers in this area has increased in the last few years.

In this paper, a brief overview of recent trends in video streaming is provided in Sect. 2. Further, four proposals by the authors for improving Quality of Services/Quality of Experience (QoS/QoE) performances are introduced in subsequent sections. Section 3 introduces route navigation, which utilizes past connection history collected on clouds and recommends a moving route that will maximize throughput and/or minimize power consumption [5]. Section 4 presents throughput prediction that uses short-term and long-term connection records and attempts to predict future throughputs by applying machine learning methods [6]. This method is also extended to transportation-mode estimation. Section 5 demonstrates image quality assessment, which designs a no-reference type image quality predictor using convolutional neural networks (CNNs) and saliency maps [7]. Section 6 presents a perceptual video encoder in which neighboring DASH representations are perceptually discriminable, owing to a newly developed subjective quality estimator [8]. Finally, Sect. 7 concludes this paper.

2. Recent Trends in Video Streaming

2.1 MPEG-DASH

MPEG-DASH [3] is an international standard that is capable of continuous playback by changing the bitrate dynamically and adaptively while observing the network bandwidth. Video contents are encoded by multiple bitrate-resolution pairs and are divided into small segments of typical lengths of 2 to 5 s. The URL of each segment is written in the Media Presentation Description (MPD) file, which has information on encoded bitrates, resolutions, minimum buffer time, etc. Clients access the MPD file at the beginning of streaming session and refer to it for selecting the optimal bitrate/resolution pair according to their network conditions. Every segment can be accessed individually by the client via HTTP GET requests. Historically, MPEG-DASH standardization was triggered after proprietary HTTP streaming methods were proposed by Apple, Adobe, and Microsoft.

Previously, the "encoding recipe" which specifies bitrate/resolution pairs was empirically determined and fixed for all the video contents. However, this fixed encoding recipe suffers from following problems: higher resolution does not always show higher quality, excessive or insufficient bitrate allocation may occur, and perceptually redundant DASH representations which are not discriminable by the human eye are often generated. Therefore, Netflix indicates that the encoding recipe should be adaptively designed

Manuscript received August 1, 2018.
Manuscript publicized January 22, 2019.

[†]The authors are with Waseda University, Tokyo, 169-8555 Japan.
a) E-mail: k.kanai@aoni.waseda.jp
b) E-mail: weibo@aoni.waseda.jp
c) E-mail: zxcheng@asagi.waseda.jp
d) E-mail: masaru-t@fuji.waseda.jp
e) E-mail: katto@waseda.jp

DOI: 10.1587/transcom.2018ANI0003

according to input video characteristics [9], [10], although their concrete methods have not been shown yet.

2.2 Machine Learning

Similar to other research fields, introduction of the machine learning into video streaming researches is ongoing. Gaussian Mixture Model (GMM), Hidden Markov Model (HMM), Support Vector Machine (SVM), Support Vector Regression (SVR), and reinforcement learning were introduced into many components such as throughput prediction [11], rate control [12], Multi-method and image quality assessment [10], [13]. Video Multimethod Assessment Fusion (VMAF) [10] is a picture quality predictor proposed recently by Netflix that integrates multiple image quality predictors and motion information by SVR, and outputs its own score.

Most recently, deep learning methods have been aggressively introduced into video streaming and image quality assessment. Pensieve [14] generates adaptive bitrate algorithms by applying deep reinforcement learning to observable QoS parameters of clients. Pensieve can be applied to the MPEG-DASH-based streaming system and adaptively selects future representations to avoid rebuffering while improving QoE. Regarding image quality assessment, several proposals, mainly using CNN, are being continued for no-reference image quality predictor since the publication of CVPR paper [15]. Correlation to ground truth subjective quality has increased although we still have scope for improving the prediction performance.

2.3 Mobile ICT Infrastructure

The authors have been executing the Mobile ICT Infrastructure project supported by JSPS Grant-in-Aid for Scientific Research (A) from 2015 to 2018. As shown in Fig. 1, in this project, we focus on (a) collection of connection history by smartphones, (b) generation of radio quality maps, (c) development of throughput prediction and consumption power prediction, (d) efficient rate control and route navigation, and (e) content delivery experiments. As extension techniques, we focus on utilization of cloud platforms and multiple sensors; investigation of new wireless communication techniques; and evaluations and improvement of QoE. Finally, we attempt to develop application prototypes such as smart route navigation, proactive/predictive content delivery, adaptive content off-loading, and rebuffering-free video streaming. The proposals put forth in this paper are mainly carried out under this project's direction along with considerations on recent trends in video streaming.

3. Route Navigation

3.1 Objectives

We previously proposed smart route navigation which maximizes communication quality (e.g., throughputs) and/or minimizes power consumption by utilizing connection history and radio quality maps [16]. We extend this work in the current study by incorporating adaptive playout buffer control.

3.2 Proposed Method and Evaluations

The basic scenario of the proposed method, which is shown in Fig. 2, is summarized as follows: (1) Connection history is collected, analyzed, and stored on a cloud system. (2) Our method determines an optimal travel route for high-speed and energy-efficient connections. (3) When a user enters into a high throughput area, our method temporarily extends a video playout buffer size and aggressively downloads video segments until the extended buffer is filled. After leaving this area, video contents are consumed smoothly till next connection spots are reached.

Recent video streaming adopts ON/OFF strategy, which is different from the classical video streaming in the 90s, that uses constant bit rate (CBR) packet delivery. This recent strategy can be classified into short (or zippy pacing) and long ON/OFF cycle (or sawtooth pacing) [17]. In our previous paper, we mentioned that the long ON/OFF cycle can contribute to reducing power consumption because the occurrences of tail energy can be reduced [18]. Based on these considerations, we consider an adaptive playout buffer control for reducing power consumption.

Figure 3 shows an example of our playout buffer control in which buffer size is increased when a user enters into a high-speed connection area. Figure 4 shows the throughput behaviors of the shortest and the optimal (good throughput)

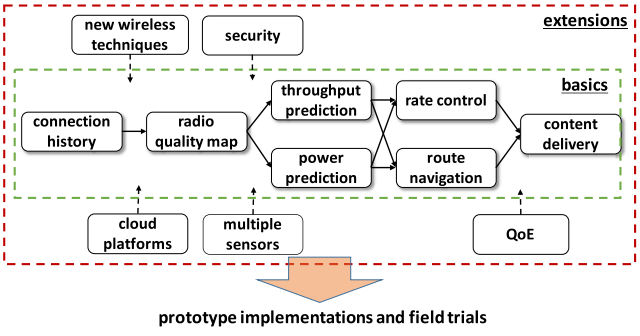

Fig. 1 Configuration of our mobile ICT infrastructure project.

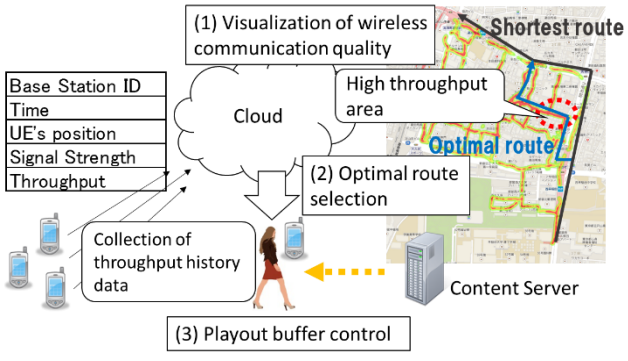

Fig. 2 Overview of smart route navigation.

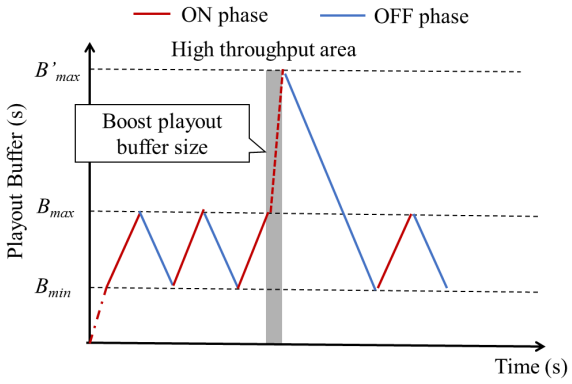

Fig. 3 Behavior of our adaptive playout buffer control.

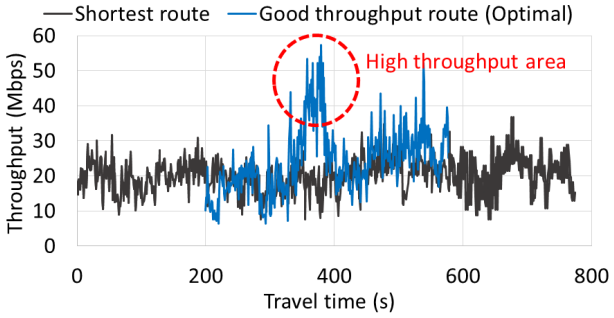

Fig. 4 Throughput comparison of the shortest and optimal route selected by the proposed method.

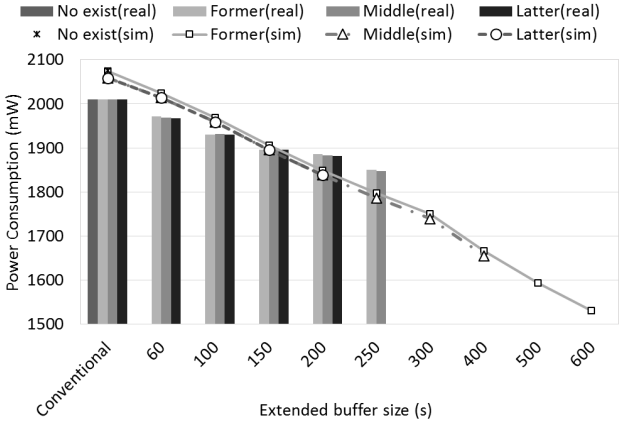

Fig. 5 Comparison of power consumption according to extended buffer sizes.

route; the shortest route is provided by Google Maps. We can confirm that our route can achieve higher throughput as expected.

Figure 5 shows the result of comparison of power consumption when the content bitrate is 20 Mbps, where *No exist* represents the case of no high-speed area, whereas *Former*, *Middle*, and *Latter* represent the location of high-speed areas near to the starting point, in the middle of the route and near the destination, respectively. *Conventional* represents the case when maximum and minimum buffer sizes are 30 s and 20 s respectively, and *real* and *sim* represent field experiment and simulation, respectively. We can save 250 mW when the maximum buffer size is 250 s.

4. Throughput Prediction

4.1 Objectives

We have been collecting throughput data in addition to location and time information by using a smartphone to generate radio quality maps. The data are used for prediction experiments. We focused on both short-term and long-term prediction and confirmed that the latter contributes to reduction in rebuffering in video streaming when users are not stationary [19], [20]. For the short-term prediction, we attempted linear prediction, GMM-HMM, a hybrid of linear and GMM-HMM [6], and most recently, long short-term memory (LSTM) [21]. We also investigated transportation-mode estimation by using throughputs, received signal strength indication (RSSI), and Cell ID [22]. This contributes to switching throughput prediction models according to moving patterns of users.

4.2 Proposed Method and Evaluations

Figure 6 shows the basic structure of our hybrid throughput predictor. We adopt a hybrid prediction model that switches linear prediction assuming autoregressive (AR) model and GMM-HMM-based state transition model. An SVM is also introduced as a classifier for selecting each of the prediction models according to throughput records. This classifier is pre-trained by using the dataset of past throughput records.

Figure 7 demonstrates the comparison of root mean square relative errors (RMSREs) among several prediction methods in different scenarios. In prediction methods, we compare harmonic mean (HM), last sample (LS), moving average (MA), stochastic [23], linear prediction (AR), GMM-HMM, and our hybrid method (HOAH). We use seven datasets: three LTE cases (static, walk, and bus), two HSPDA cases (ferry and train), and two WiFi cases (personal 5 GHz and enterprise WPA2). HSPDA datasets are published in [24], whereas LTE and WiFi datasets are collected by us. AR performs better in static cases when throughput is stationary. In moving cases, GMM-HMM performs better because HMM can follow state changes. From these results, we can conclude that our hybrid method is effective for the abovementioned seven scenarios.

We also attempt transportation mode estimation by using communication quality parameters, throughputs, RSSI, and Cell ID, which are available inline during session [22]. Figure 8 shows the concept of transportation mode estimation combined with subsequent throughput prediction. Because the throughput behaviors are different according to moving patterns, we can expect better prediction by preparing different prediction models per transportation mode. Furthermore, if we can achieve sufficient estimation performance without using dedicated sensors like GPS and accelerometers, power reduction of mobile devices can be expected.

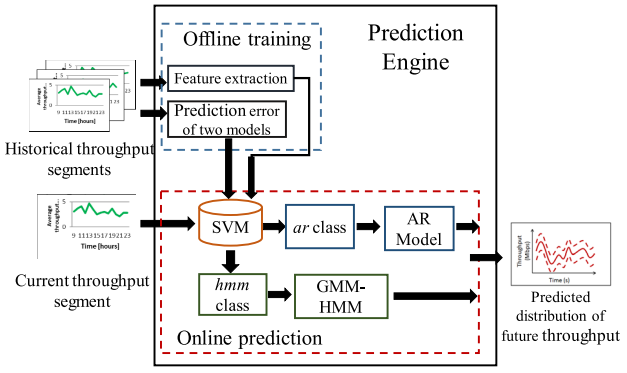

Fig. 6 Configuration of proposed throughput predictor.

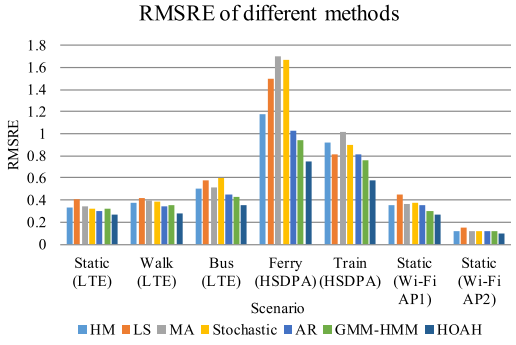

Fig. 7 Comparison of RMSREs among various prediction methods.

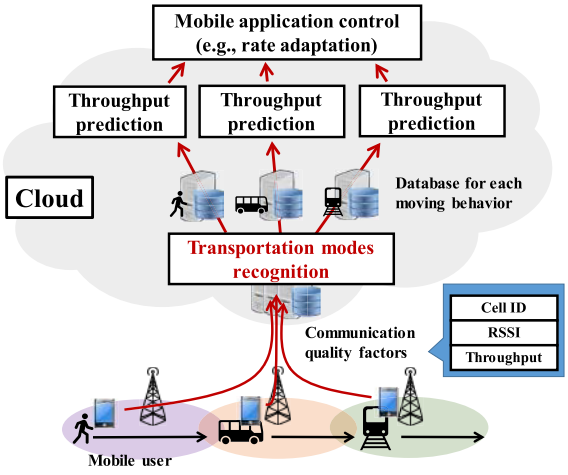

Fig. 8 Transportation mode estimation using communication quality factors.

Figure 9 shows the results of recognition accuracy of three transportation modes (static, walking, and riding a train) for three different datasets. In this experiment, we use random forest as a classifier among three candidates: SVM, nearest neighbor, and random forest, based on comparison experiments. As shown in the figure, results conclude that typical transportation modes for daily commuting can be accurately recognized. The contribution of Cell IDs is significant for this task.

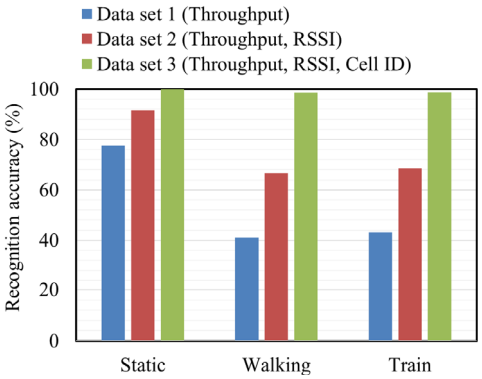

Fig. 9 Results of recognition accuracy of three transportation modes.

5. Image Quality Assessment

5.1 Objectives

Research on image quality assessment has a long history because it is well known that MSE and PSNR do not necessarily represent subjective image quality precisely. SSIM [25] is a popular technique that offers better quality prediction than PSNR; however, its performance is limited and reference images are necessary for evaluation. No-reference type (or blind) image quality assessment methods, which do not need reference images, have also been proposed including the one proposed by the authors, namely the one using SVR [26]. However, their performances are not impressive owing to the limited capability of classical machine learning techniques.

In recent years, however, deep learning-based image quality assessment has been focused on by many researchers owing to its high prediction accuracy. In this research, we propose a blind and fast image quality predictor using convolutional neural networks. Our method introduces a saliency map [27] into the predictor and devises acceleration techniques for reducing computational complexity of the proposed method.

5.2 Proposed Method and Evaluations

Figure 10 shows the configuration of our proposed image quality predictor. During pre-processing, local contrast normalization is applied to input images, and the normalized image is split into 32×32 patches with RGB components. In the training process, we apply a distortion clustering strategy which consists of three steps: CNN for distortion recognition, posterior observation, and distortion clustering. The distortion recognition tries to recognize 13 distortion types such as compression, blur, and white noise. The posterior observation tries to describe distorted images by two parameters of gamma function. The distortion clustering tries to group the distortion types into four clusters (distortion mapping table) based on the abovementioned two parameters. We then design and train another CNN for each distortion

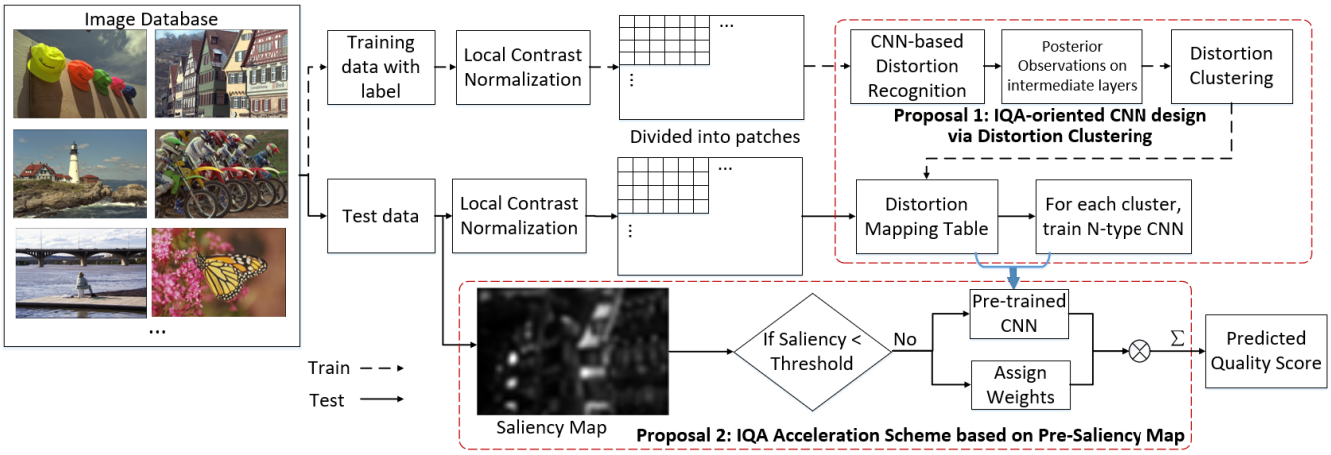

Fig. 10 Configuration of the proposed image quality predictor.

Table 1 Comparison of prediction accuracy among existing methods and our proposal.

Methods	PCC	SROCC
PSNR	0.868	0.873
SSIM	0.913	0.906
VSI	0.948	0.952
Kang, CVPR 2014	0.953	0.956
Li, DSP 2016	0.956	0.935
Sun, VCIP 2016	0.958	0.959
Bosse, ICIP 2016	0.972	0.96
Pan, VCIP 2016	0.969	0.968
Zuo, ICIP 2016	0.967	0.964
Proposed Method	**0.978**	**0.974**

Algorithm 1 Pre-Saliency Map based Quality Aggregation

Input: Tested Image $I(X,Y)$ and Saliency Map $S(X,Y)$
Split I into $N \times N$ patches $\{P_1, P_2, \cdots, P_M\}$
while $i \in [1, M]$ **do**
  Calculate the average saliency value in P_i:
  $\overline{S_i} = \frac{\sum_{x=1}^{N}\sum_{y=1}^{N} S(x,y)}{N \times N}$
  if $\overline{S_i} \leq \epsilon$ **then**
    Skip the CNN computation for P_i;
  else
    $q_i = f_{CNN}(P_i)$;
    Assign weights by $\omega_i = Norm(\sum_{j=1}^{N \times N} S(j))$;
  end if
  $i = i + 1$;
end while
Final quality score for I is $Q = \frac{\sum_{i=1}^{M} \omega_i \times q_i}{\sum_{i=1}^{M} \omega_i}$.

cluster to predict image quality score.

To achieve the final image quality score, we use a saliency map. By analyzing the relation between saliency information and prediction errors, we found that non-salient regions are likely to have larger prediction errors than salient regions. Therefore, we remove non-salient patches from image quality calculation and assign weights to salient patches. This removal of non-salient regions contributes to the acceleration of image quality score calculation. A summarized algorithm description of our image quality predictor is presented below.

Table 1 shows the comparison results of prediction accuracy to ground truth subjective quality among existing methods and our proposal for LIVE database [28]. We recognize that our proposal yields the best result in both Pearson correlation coefficient (PCC) and Spearman rank-order correlation coefficient (SROCC). We also confirmed that by using fast saliency map model [29] and reducing the number of salient regions (i.e., increasing threshold ε in Algorithm 1), the computational complexity can be successfully reduced while maintaining prediction accuracy.

6. Perceptual Video Streaming Using JND Estimation

6.1 Objectives

Current video streaming uses a fixed encoding recipe (bitrate-resolution pair) as shown in Table 2 for providing network adaptivity and/or device scalability by using MPEG-DASH or similar techniques. However, it has been indicated that this fixed encoding recipe suffers from several problems such as improper resolution selection and stream redundancy; therefore, the necessity of "per-title encode optimization" which adaptively generates an encoding recipe according to input video characteristics is advocated [9], [10].

Moreover, Just-Notable Difference (JND) is known as a subjective quality measure which quantifies the number of people noticing quality differences. Table 3 [30] shows the relationship between JND score and the number of people when people compare two pictures: reference picture and degraded picture; the ratio of people who prefer the reference video determines the JND score. As per the definition, when JND score is zero, people do not notice picture difference, but when JND score is one, many people start to notice the picture difference.

Therefore, in this research, we develop a perceptual quality driven video streaming based on JND scores. We develop a JND estimator by using SVR and generate an encoding recipe in which neighboring representation has one JND distance. This approach can contribute to avoiding redundant encoding caused by the fixed encoding recipe.

Table 2 Conventional fixed encoding recipe.

Bitrate(kbps)	Resolution
5800	1920x1080
4300	1920x1080
3000	1280x720
2350	1280x720
1750	720x480
1050	640x480
750	512x384
560	512x384
375	384x288
235	320x240

Table 3 Relationship between JND score and the number of people who noticed quality differences.

JND Score	Number of people	Percentage to think "reference" is better
0	2	50%
1	4	75%
2	8	87.5%
3	16	93.75%
4	32	96.875%

6.2 Proposed Method and Evaluations

Figure 11 shows our encoder structure for 2K video input. *Pre-encoding* carries out trial encoding of four resolutions (1080p, 720p, 540p, and 360p) and three QP points (15, 30, and 5) per 5-s segment. *Distortion measures* calculate the following three distortion measures of the encoded pictures: PSNR, SSIM, and VMAF. *Curve fitting* approximates rate-distortion curves by equation $y = a \cdot x^b + c$ from three trial encoding points per resolution. *JND estimation* estimates JND scores on the approximated rate-distortion curves by using a developed JND estimator. *Recipe generation* generates an encoding recipe in which neighboring representation has one JND distance. Finally, *Encode* carries out video encoding to produce multiple representations according to the recipe.

The JND estimator is pre-trained by using VideoSet database [31], which has ground truth JND scores of one, two, and three achieved by subjective assessment for 220 sequences in addition to using our own subjective assessment results for compensating JND scores more than four. SVR inputs are QP, bitrate, resolution and, distortion measures, and SVR output is the estimated JND score.

Figure 12 shows an example of rate-JND curves, and Fig. 13 illustrates the comparison of averaged storage sizes for storing 10 representations by the fixed encoding recipe and our adaptive encoding recipe. From Fig. 12, we can recognize that better rate-JND curves (i.e., selection of better bitrate-resolution pair) are achieved. From Fig. 13, we notice that storage sizes can be reduced by more than half while keeping similar picture quality. Although omitted, we also verified that better QoE performances were achieved over congested networks implemented by a network emulator owing to smaller storage sizes.

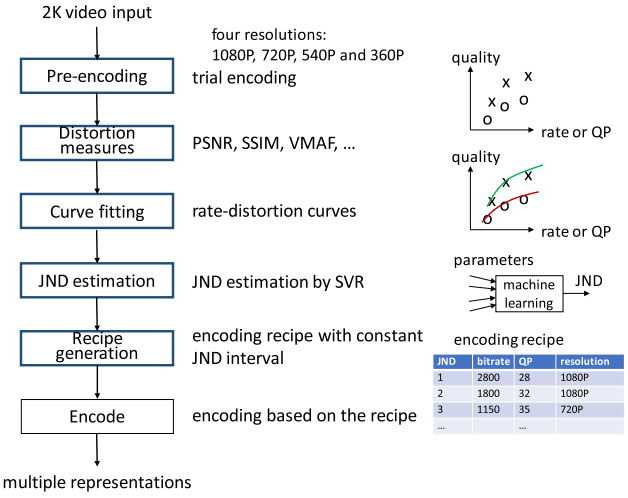

Fig. 11 Proposed encoder structure.

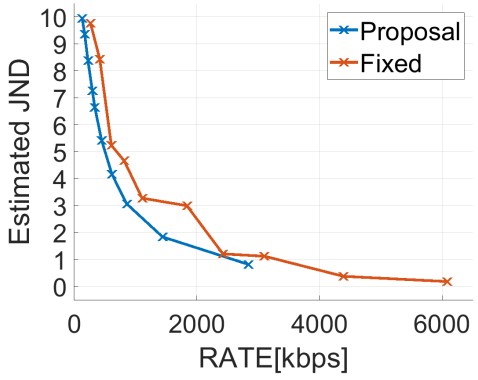

Fig. 12 Example of rate-JND curves for VideoSet No.8 sequence.

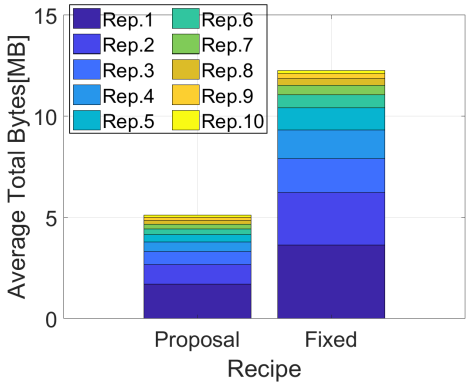

Fig. 13 Comparison of averaged total storage sizes for VideoSet sequences.

7. Conclusion

This paper has presented a brief overview of recent trends in video streaming and introduced four methods proposed by the authors. Route navigation using past connection records successfully provided higher throughput and lower power

consumption. In throughput prediction, adaptively switching between linear prediction and GMM-HMM model resulted in better prediction performance than existing methods. Transportation-mode estimation using communication quality parameters was also described. Image quality assessment based on CNN and saliency maps demonstrated the best prediction performance among the compared methods, and its acceleration performance was discussed. Finally, perceptual video streaming using a newly developed JND estimator was proposed, and an improvement in rate-JND curves and a reduction in storage sizes were achieved.

As future work, integration of the proposed methods into a single framework will be considered. Furthermore, improvement in each method will be considered by incorporating recent evolution in machine learning.

Acknowledgments

This work was partially supported by JSPS KAKENHI Grant Numbers 15H01684, 15H02688 and 17K12681.

References

[1] "Complete Visual Networking Index (VNI) Forecast," https://www.cisco.com/c/en/us/solutions/service-provider/visual-networking-index-vni/index.html
[2] I. Sodagar, "The MPEG-DASH standard for multimedia streaming over the Internet," IEEE Multimedia, vol.18, no.4, pp.62–67, April 2011.
[3] ISO/IEC 23009-1, "Dynamic adaptive streaming over HTTP (DASH) – Part 1: Media presentation description and segment formats," 2014.
[4] Netflix Tech Blog, "Using Machine Learning to Improve Streaming Quality at Netflix," https://medium.com/netflix-techblog/using-machine-learning-to-improve-streaming-quality-at-netflix-9651263ef09f, March 2018.
[5] K. Kanai, S. Takenaka, J. Katto, and T. Murase, "Energy-efficient mobile video delivery utilizing moving route navigation and video playout buffer control," IEICE Trans. Commun., vol.E101-B, no.7, pp.1635–1644, July 2018.
[6] B. Wei, K. Kanai, W. Kawakami, and J. Katto, "HOAH: A hybrid TCP throughput prediction with autoregressive model and hidden Markov model for mobile networks," IEICE Trans. Commun., vol.E101-B, no.7, pp.1612–1624, July 2018.
[7] Z. Cheng, M. Takeuchi, K. Kanai, and J. Katto, "A fully-blind and fast image quality predictor with convolutional neural networks," IEICE Trans. Fundamentalsm vol.E101-A, no.9, pp.1557–1566, Sept. 2018.
[8] M. Takeuchi, S. Saika, Y. Sakamoto, T. Nagashima, Z. Cheng, K. Kanai, J. Katto, K. Wei, J. Zengwei, and X. Wei, "Perceptual quality driven adaptive video coding using JND estimation," PCS 2018, June 2018.
[9] Netflix Tech Blog, "Per-Title Encode Optimization," https://medium.com/netflix-techblog/per-title-encode-optimization-7e99442b62a2, Dec. 2015.
[10] J.D. Cock, Z. Li, M. Manohara, and A. Aaron, "Complexity-based consistent-quality encoding in the cloud," IEEE ICIP 2016, Sept. 2016.
[11] Q. Xu, Z.M. Mao, S. Mehrotra, and J. Li, "PROTEUS: Network performance forecast for real-time, interactive mobile applications," ACM MobiSys 2013, June 2013.
[12] V. Menkovski and A. Liotta, "Intelligent control for adaptive video streaming," IEEE ICCE 2013, Jan. 2013.
[13] A.K. Moorthy and A.C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," IEEE Trans. Image Process., vol.20, no.12, pp.3350–3364, April 2011.
[14] H. Mao, R. Netravali, and M. Alizadeh, "Neural adaptive video streaming with pensieve," ACM SIGCOMM 2017, Aug. 2017.
[15] L. Kang, P. Ye, Y. Li, and D. Doermann, "Convolutional neural networks for no-reference image quality assessment," CVPR 2014, June 2014.
[16] K. Kanai, J. Katto, and T. Murase, "Performance evaluations of comfort route navigation providing high-QoS communication for mobile users," ITE Trans. Media Technology and Applications, vol.2, no.4, pp.327–335, Oct. 2014.
[17] A. Rao, A. Legout, Y.S. Lim, D. Towsley, C. Barakat, and W. Dabbous, "Network characteristics of video streaming traffic," ACM CoNEXT 2011, Dec. 2011.
[18] Y. Ishizu, K. Kanai, J. Katto, H. Nakazato, and M. Hirose, "Energy-efficient video streaming over named data networking using interest aggregation and playout buffer control," IEEE Greencom 2015, Dec. 2015.
[19] H. Konishi, K. Kanai, and J. Katto, "Improvement of throughput prediction accuracy for video streaming in mobile environment," IEEE GCCE 2014, Oct. 2014.
[20] K. Kanai, H. Konishi, Y. Ishizu, and J. Katto, "A highly-reliable buffer strategy based on long-term throughput prediction for mobile video streaming," IEEE CCNC 2015, Jan. 2015.
[21] Bo Wei, W. Kawakami, K. Kanai, J. Katto, and S. Wang, "TRUST: A TCP throughput prediction method in mobile networks," IEEE Globecom 2018, Dec. 2018.
[22] W. Kawakami, K. Kanai, B. Wei, and J. Katto, "Machine learning based transportation modes recognition using mobile communication quality," IEEE ICME 2018, July 2018.
[23] H. Yoshida, K. Satoda, and T. Murase, "Constructing stochastic model of TCP throughput on basis of stationarity analysis," IEEE Globecom 2013, Dec. 2013.
[24] HSDPA, http://home.ifi.uio.no/paalh/dataset/hsdpa-tcp-logs/
[25] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," IEEE Trans. Image Process., vol.13, no.4, pp.600–612, April 2004.
[26] T. Kumekawa, M. Wakabayashi, J. Katto, and N. Wada, "Blind PSNR estimation of compressed video sequences supported by machine learning," ITE Trans. Media Technology and Applications, vol.2, no.4, pp.353–361, Oct. 2014.
[27] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," IEEE Trans. Pattern Anal. Mach. Intell., vol.20, no.11, pp.1254–1259, Nov. 1998.
[28] H.R. Sheikh, M.F. Sabir, and A.C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," IEEE Trans. Image Process., vol.15, no.11, pp.3440–3451, Nov. 2006.
[29] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," IEEE Trans. Image Process., vol.20, no.8, pp.2378–2386, Aug. 2011.
[30] VideoClarity: "Understanding MOS, JND, and PSNR," available at http://videoclarity.com/PDF/WPUnderstandingJNDMOSPSNR.pdf, viewed at Jan. 2018.
[31] H. Wang, I. Katsavounidis, J. Zhou, J. Park, S. Lei, X. Zhou, M.O. Pun, X. Jin, R. Wang, X. Wang, Y. Zhang, J. Huang, S. Kwong, and C.C. Kuo, "Videoset: A large-scale compressed video quality dataset based on JND measurement," J. Vis. Commun. Image R., vol.46, pp.292–302, 2017.

Kenji Kanai received the B.E., M.E., and Ph.D. degrees from Waseda University, Tokyo, Japan, in 2010, 2012, and 2015, respectively. He is currently an Assistant Professor at Waseda University. He is a member of IEEE, IEICE, and IPSJ.

Bo Wei is currently working toward her Ph.D. degree at the Graduate School of Fundamental Science and Engineering, Waseda University. She received her B.E. and M.E. degrees from Tianjin University, Tianjin, China in 2012 and 2015, respectively. She is a student member of the IEEE and IEICE.

Zhengxue Cheng received the B.E. degree from Shanghai Jiao Tong University, Shanghai, China, in 2014, and received the M.E. degree from Waseda University and Shanghai Jiao Tong University, in 2015 and 2017, respectively, through a double-degree program. She is currently pursuing the Ph.D. degree in Waseda University.

Masaru Takeuchi received the B.E. and M.E. degrees from Waseda University, Tokyo, Japan, in 2010 and 2012, respectively. He joined Sharp Corporation in 2012 and then joined Waseda University in 2015. He is currently pursuing the Ph.D. degree in Waseda University. He is a member of IEEE and IEICE.

Jiro Katto received the B.S., M.E., and Ph.D. degrees from the University of Tokyo in 1987, 1989, and 1992, respectively; all in electrical engineering. He joined NEC Corporation in 1992 and then joined Waseda University in 1999. He is a fellow of IEICE and a member of ITE, IPSJ, IEEE, and ACM.