Modeling of Perceptual Video Quality Considering Impact of Spatial, Temporal, and Amplitude Resolutions

DISSERTATION

for the Degree of

Doctor of Philosophy (Electrical Engineering)

Yen-Fu Ou

Jan. 2012

Modeling of Perceptual Video Quality Considering Impact of Spatial, Temporal, and Amplitude Resolutions

DISSERTATION

Submitted in Partial Fulfillment
of the REQUIREMENTS for the

Degree of

DOCTOR OF PHILOSOPHY (Electrical Engineering)

at the

POLYTECHNIC INSTITUTE OF NEW YORK UNIVERSITY

by

Yen-Fu Ou

Jan. 2012

Approved:

____
Department Head

____
Date

Copy No. __4__

Approved by the Guidance Committee:

Major: Electrical and Computer Engineering

Yao Wang
Professor of
Electrical and Computer Engineering

Yong Liu
Associate Professor of
Electrical and Computer Engineering

Chang Feng
Vice President, Technology and Innovation
ooVoo LLC.

Minor: Computer Science

Edward Wong
Associate Professor of
Computer Science

Microfilm or other copies of this dissertation are obtainable from

UMI Dissertations Publishing
Bell & Howell Information and Learning
300 North Zeeb Road
P.O.Box 1346
Ann Arbor, Michigan 48106-1346

VITA

Yen-Fu Ou received the B.S. and M.S. degrees in mechanical engineering from National Chiao Tung University, Hsinchu, Taiwan, in 2000 and 2002, respectively, and the M.S. degree in electrical and computer engineering from Columbia University, New York, NY, in 2006. He is currently working toward the Ph.D. degree in electrical engineering from the Polytechnic Institute of New York University, Brooklyn. He interned as a QoS/QoE researcher at ooVoo LCC, New Yrok, in 2009 and now he joins Wuawei R&D Core Network departement, Bridgewater, NJ, since 2012.

His current research interests include the perceptual video quality assessment, video bitstream adaptation, and QoS/QoE provisioning for video conferencing systems.

ACKNOWLEDGEMENT

During my journey toward the PhD degree, it would not be possible to keep forward without the support and help from many people. Now, it is my great pleasure to thank them for the encouragement, assistance and advice that I received over these five years.

First of all, I would like to address my deepest gratitude toward Prof. Dr. Yao Wang, my dissertation advisor, for offering this opportunity to follow her from ”Formosa Island” into this best video group in New York and pursue my doctoral studies under her supervision. I always esteem her ability of having a professional work attitude while being a considerate and open-minded advisor. I especially thank her for leading me into the profound research field and helping me develop a strong capability of doing research.

My professional development has further been influenced by some senior colleagues, Dr. Zhengye Liu, who gave me precious knowledge for my industry projects in OoVoo Inc, Dr. Tao Liu, who taught me how to set up the subjective quality test and his experiences regarding the quality assessment when I was new to this filed, and Dr. Zhan Ma, who was being an outstanding co-author with me for several published conference and journal papers and I learned a lot from his highly competent and inspiring research attitude. I would also like to thank Chang Feng, VP Technology & Innovation in ooVoo Inc, for partly funding my PhD program and who has been an creative and excellent mentor for the project of QoE/QoS over video streaming.

Special thanks also goes to my general assistants, Minyi Yang, who helped me implement my algorithm into a well-developed testbed so that we can build up our simulation smoothly; Wenzhi Lin and Huiqi Zeng, who helped me collect human subjects for the quality assessment of frame rate/quantization variation. My thanks are further extended to Yuanyi Xue for his excellent jobs on establishing the subjective experiment of our QSTAR modeling on mobile devices. My great appreciation also goes to my other collaborators, lab-mates and friends, Hao Hu, Xuan Zhou, Cagdas Bilen, Zhi Zhao, Yan Zhao, Xin Feng, Meng Xue, Zhili Guo, Ailing Song and Shufen Chan.

As my work is essentially based on data collected from many human subjects, I am

highly grateful to those who participate the quality evaluation and some of my friends, who spend even more than once to help us out for several experiments.

Even though my parents, Kim Ou and Su-O Yang, who never had this wonderful opportunity to move or study abroad, their support for me have always been unconditional and unreserved. Although I cannot be with you for many holidays, special occasions in Taiwan, you are definitely more than welcome to join with me in the future whenever necessary. Thank you mom and daddy always there for me.

I have a very grateful thank to this special person for being with me when I am pursuing my PhD degree, my beloved wife Ya-Chien Virginia Chang Chien. This ”journey” would have been far more difficult without a great partner encouraging me when I am mostly needed. Her continuous supporting, endless love, and carefully take care of my daily life sustain me to get where I am today. My sincerely appreciation to you Virginia.

AN ABSTRACT

Modeling of Perceptual Video Quality Considering Impact of Spatial, Temporal, and Amplitude Resolutions

by

Yen-Fu Ou

Advisor: Yao Wang

Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy (Electrical Engineering)

Jan 2012

In networked video applications where the sustainable bandwidths vary greatly among the receivers and in time, one must be able to code a video or adapt a precoded-bitstream to a wide range of bit rates. It is critical to choose appropriate frame size (FS), frame rate (FR) and quantization stepsize (QS) to optimize the perceptual quality for a target bit rate. Furthermore, it is important to understand the variation effect of these parameters, so that appropriate constraints can be imposed when adapting these parameters in response to bandwidth changes. However, well-established models that relate the perceptual quality with the spatial, temporal, and amplitude resolutions (STAR) and the variation of STAR do not exist today.

In this dissertation, we conducted three main subjective quality experiments. First one examines the impact of FR and QS on perceptual quality of a video for laptop devices. We propose to use the product of a spatial quality factor that assesses the quality of decoded frames without considering the frame rate effect and a temporal correction factor, which reduces the quality assigned by the first factor according to the actual frame rate. We find that the temporal correction factor follows closely an inverted falling exponential function of FR, whereas the quantization effect on the coded frames can be captured accurately by a sigmoid function of the PSNR or by an exponential function of QS. The complete model correlates well with the subjective ratings with a Pearson Correlation Coefficient (PCC) of 0.98 when parameters are obtained by least square fitting with the subjective ratings and a PCC of 0.97 when model parameters are predicted from the content features.

The second experiment investigates the impact of STAR on the perceptual quality of a compressed video. Subjective quality tests were carried out on the TI Zoom2 mobile development platform (MDP). Subjective data reveals that the impact of SR, TR and QS can each be captured by a function with a single content-dependent parameter. The joint impact of SR, TR and QS can be modeled by the product of these three functions with only three parameters. The complete model correlates very well with the subjective ratings with a PCC of 0.99 when parameters are obtained by least square fitting and a PCC of 0.98 when model parameters are predicted from the content features. We further validate our model on several datasets reported from other works and the accuracy of our model (in part or in whole) on these datasets is still promising.

The third experiment explores the impact of periodic frame rate and QS variation on perceptual video quality. Among many dimensions of FR/QS variation, as a first step we focus on videos in which two FR's, or QS's, alternate over fixed intervals. According to the observation and data analysis of the test results, we propose models that characterize the quality degradation with respect to FR, QS and bit rate variation. These quality models can help to make appropriate decisions for encoder adaptation when transmitting video over networks with fluctuating bandwidth.

List of Publications

Journal Publications

[1]. Yen-Fu Ou, W. Lin, H. Zeng, and Y. Wang, "Perceptual Quality of Video with Frame Rate and Quantization Variation : a subjective study and analytical modeling," Submitted to *Circuits and Systems for Video Technology (CSVT), IEEE Transactions on* , 2012.

[2]. Yen-Fu Ou, Y. Xue, Y. Wang, "Q-STAR: A Perceptual Video Quality Model for Mobile Platforms Considering Impact of Spatial, Temporal, and Amplitude Resolutions," Submitted to *IEEE Journal on Selective Areas in Communications : Special Issue on QoE-Aware Wireless Multimedia Systems*, 2011.

[3]. Z. Ma, M. Xu, Yen-Fu Ou and Y. Wang, "Modeling of Rate and Perceptual Quality of Video as Functions of Frame Rate and Quantization Stepsize and Its Applications," To appear in *Circuits and Systems for Video Technology (CSVT), IEEE Transactions on*, Nov. 2011.

[4]. Yen-Fu Ou, Z. Ma, T. Liu, Y. Wang "Perceptual Quality Assessment of Video Considering both Frame Rate and Quantization Artifacts", *Circuits and Systems for Video Technology (CSVT), IEEE Transactions on* ,Vol. 21, no. 3, pp. 286-298, March 2011.

Conference Publications

[1]. Yen-Fu Ou, H. Zeng, Y. Wang, Perceptual Quality of Video with Quantization Variation : a subjective study and analytical modeling, Submitted to, *IEEE International Conference on Image Processing (ICIP)*, 2012

[2]. Yen-Fu Ou, Y. Xue, Z. Ma, Y. Wang, "A Perceptual Video Quality Model for Mobile Platform Considering Impact of Spatial, Temporal, and Amplitude Resolutions," *IEEE*

Image, Video and Multidimendional Signal Processing Technical Committee (IVMSP) Workshop, Jun. 2011, pp. 117-122.

[3]. Yuanyi Xue, Yen-Fu Ou, Z. Ma and Y. Wang, "Perceptual Video Quality Assessment On A Mobile Platform Considering Both Spatial Resolution And Quantization Artifacts," *in Proc. of PacketVideo Workshop*, Hong Kong, December 2010, pp. 201-208.

[4]. Yen-Fu Ou, Y. Zhou and Y. Wang, "Perceptual Quality of Video with Frame Rate Variation : a subjective study," *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP10')*, Dallas, TX, March 14 - 19, 2010, pp. 2446 - 2449

[5]. Y. Wang, Z. Ma, and Yen-Fu Ou, "Modeling rate and perceptual quality of scalable video as functions of quantization and frame rate and its application in scalable video adaptation, *in Proc. of Packet Video Workshop*, 2009, pp. 1- 9.

[6]. Yen-Fu Ou, Z. Ma, and Y. Wang, "Modeling the Impact of Frame Rate and Quantization Stepsizes and Their Temporal Variations on Perceptual Video Quality: A Review of Recent Works," *in Proc. of IEEE Information Science and System (CISS)*, Princeton, NJ, March 2009, pp. 1-6.

[7]. Yen-Fu Ou, Z. Ma, Y. Wang "A Novel Quality Metric for Compressed Video Considering both Frame Rate and Quantization Artifacts," *International Workshop on Image Processing and Quality Metrics for Consumer (VPQM'08)*, Scottsdale, AZ, USA, Jan. 15-16, 2009.

[8]. Yen-Fu Ou, T. Luo, Z. Zhao, Z. Ma, Y. Wang "Modeling the Impact of Frame Rate on Perceptual Quality of Video", *IEEE International Conference on Image Processing (ICIP'08)*, San Diego, CA, USA, Oct. 12-15, 2008, pp. 689-692.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Multimedia applications and services are becoming more and more prevalent, such as video telephony, mobile video broadcasting, high definition television (HDTV), and Internet Protocol television (IPTV). Proper provisioning of the networked video applications, both in terms of video codec design and transport level control, depends on a reliable method that can predict the video quality automatically and accurately to assure a certain quality of service (QoS) or quality of experience (QoE).

There are basically two different ways to assess the perceptual video quality, i.e., subjective quality evaluation and objective quality prediction, each having its own merits and shortcomings. Subjective evaluation essentially mimics processing of human perception because it can accurately reflect the video quality ratings through standard experiments and measurements from a large amount of population, if it is well designed and performed. However, this kind of quality assessment is time- and effort-consuming due to the requirement of enormous hours or weeks planning and performing on plenty of subjects and it not even possible to conduct subjective assessment of a large number of test sequences by many subjects in a batch fashion, hence this quality assessment of a large number of test sequences in a batch fashion is not even possible. And it is even not feasible for some applications, e.g. real-time quality assessment during networked video applications. Therefore, a more efficient quality assessment solution that requires no involvement of subjects once developed is more practical. Objective quality prediction, as an alternative solution to subjective quality evaluation, has all these desirable features. Because it can emulate the humans judgment on video quality based on mathematic models and can be easily applied to any test video sequence, it receives more and more attention in both industrial and academic communities. However, lack of thorough understanding about human visual perception system, both psychologically and physiologically, and a large amount of possible video quality-affecting factors, both application dependent and content-dependent, make the design of effective video quality metrics a very intriguing task. In the past few years, a big effort in the scientific community has been devoted to the development of better video quality metrics that correlate well with the human perception of quality [9–15]. Although many metrics have been proposed, most of them are very complex and require the original video for estimating the quality. This makes their use in real-time transmission applications very difficult. Therefore, a robust and efficient objective metric that blindly estimates the quality of a video is still in need.

1.1 Reviews of Subjective Quality Evaluation Methodology

Prior to developing accurate objective quality metrics, subjective assessment is necessary to be conducted as the preliminary gauge of perceived visual quality. Subjective quality assessment usually requires a subjective experiment where the quality index of each tested video sequence is produced by the mean of opinion scores (MOS) of the quality ratings from subjects. There are several testing methodologies are defined in Recommendation ITU-R BT.500-11 [16] and ITU-T Rec. P.910 [17], by International Telecommunication Union (ITU) and Video Quality Expert Group (VQEG). They explicitly provides specifications of how to perform different types of subjective video quality assessments, which include single stimulus, e.g., single stimulus continuous quality evaluation (SSCQE), Absolute Category Rating (ACR), and Double Stimulus (DS), e.g., double stimulus continuous quality scale (DSCQS) and Double-Stimulus Impairment Scale (DSIS). In DS, viewers are presented with two videos, one of which is a unimpaired source sequence, and the other is a processed version of that sequence. The sequence presentation orders are randomized. Viewers are asked to watch each video twice and evaluate the picture quality of both sequences using a grading scale in the second presentation. In SS, ACR is designed to test the subjective quality scores given by viewers without explicit references, since any of the DS methods cannot reproduce the real-world reference-free viewing conditions as a single stimulus method. In addition, ACR is a very efficient method and a large number of sequences can be tested in a relatively short time. Due to the lack of reference, it is assumed that all the references are perfect distortion-free video sequences. However, in most practical situations, some artifacts are inevitably introduced in the video capture phase, and hence these artifacts cannot be distinguished from the ones which are generated for testing purpose with this method. To solve this issue, later on VQEG introduced ACR with Hidden Reference (ACR-HR) method [17], where the original unimpaired versions of test sequences are inserted randomly into the test dataset, and then also judged by viewers, but the viewers are unaware of the existence of these references. Usually differential MOS (DMOS) between reference and test sequences is calculated to remove the reference effect, which is called reference removal. Some researchers have performed investigations on the relationships between subjective test results and subjective test protocols and proposed several approaches to improve the reliability and efficiency of the existing subjective quality assessment methods. A study [18] performed by National Telecommunications and Information Administration/The Institute for Telecommunication Sciences (NTIA/ITS) compared several aforementioned methods and concluded that SSCQE under proper design can produce quality estimates comparable to DSCQS. In a recent project report [19] of VQEG, subjective results obtained with ACR-HR method in different labs achieved very high consistency, which shows the effectiveness of this test method. It is found that humans tend to forget video contents displayed far enough from the current time instance due to the limitation of human memory capacity [20, 21]. A study on the impact of memory on SSCQE results [18] indicates the last 9-15 seconds of video content is critical for viewers to form their quality judgment on the entire clip. Properly designed SSCQE testing (with short 9-15 second sequences) maybe an effective substitute for more complicated

DSCQS testing. Therefore, the memory effect of human viewers is not only one of the concerns when designing and performing the subjective tests, but also a practical consideration when devising objective quality metrics, especially for quality assessment of longer video sequences. With respect to the issue of reusability of subjective results from different experiments, the work in [22] propose to use the subjective scores from the common set of test sequences from different tests to map all the subjective scores onto a single scale so that available subjective data is greatly increased and hence the inter-test comparisons are enabled.

1.2 Reviews of Objective Quality Assessment Methodology

Depending on the purpose for employing the quality metric, such as the quality monitoring, comparison of video processing systems, or optimization of the existing parameter settings and algorithms for a video system, it can be divided into three different categories according to accessibility of source (reference) video signal:

Full Reference (FR) metric - Both original and distorted videos are available. It require full access to all pixels of reference video signal. Both distorted and reference videos must be well calibrated before applying the quality metric.

Reduced Reference (RR) metric - Only partial information of original videos are available. Usually they first extract several features from both reference and distorted videos before applying the quality metric. It predicts the video quality based on the corresponding features of reference and distorted video signals.

No-reference (NR) metric - The original video is not available. It is also known as blind metrics, which can only access the distorted video signal. It is a very challenging task due to the lack of source information.

According to different approaches people use to estimate the impairment of a video, FR metrics are better utilized for offline video quality measurement, such as codec evaluation or laboratory simulation. This kind of metrics predict the quality of a video by comparing the differences between reference and distorted video signal in either pixel domain or some feature domain, requiring fully accessing the reference video. Most common and widely used FR metrics are mean square error (MSE) and peak signal-to-noise ratio (PSNR), i.e.,

$$\mathsf{MSE} = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\big(P_r(i,j) - P_d(i,j)\big)^2 \tag{1.1}$$

$$\mathrm{PSNR} = 10\log_{10}\Big(\frac{\mathsf{MAX}^2}{\mathsf{MSE}}\Big). \tag{1.2}$$

where M, N are the width and height of the image in pixel, MAX is the maximum possible pixel value of the image depending on its representing precision (it is 255 for 8 bits per sample) and i, j are i^{th} and j^{th} pixel in x, y coordinate. P_r, P_d are pixel values of reference and distorted video, respectively. This is a very simple and intuitive way to measure the video quality, but they do not always correlate well with subjective quality ratings [23–25]. There are many proposed FR metrics. One is the video quality metrics (VQM) [26] developed by NTIA/ITS, which provides an objective quality measurement for videos with variant encoding and transmission systems. It measures the perceptual effects of broad range of video impairments including blurring, jerky/unnatural motion, global noise, block distortion, color distortion, temporal distortion. Independent tests by VQEG have shown that the General Model of VQM on MPEG-2 and H.263 video has a high correlation with subjective video quality ratings [27]. This model has been recommended by ANSI as well as ITU-T as an objective video quality metric for secondary distribution of digitally encoded TV quality video [28, 29]. In the more recent works proposed by Z. Wang et. al [25], authors proposed the structure similarity (SSIM) quality index to predict the video quality. SSIM has been tested on the videos in LIVE database [30], which contains wide range of distortions, such as compression, wireless, and Gaussian noise. The results show that the quality metric can significantly improve the correlations with subjective data over

PSNR. Following their works of SSIM metric, Bovik's group further proposed the visual information fidelity (VIF) [31], and Motion-based Video Integrity Evaluation [32] also show the significant improvement over SSIM on LIVE database as well as the VQEG FRTV phase I pool [33].

Regarding the reduced reference metrics, it requires only partial information about the reference video. In general, certain features or physical measures are extracted from the reference and transmitted to the receiver as a side information to help evaluate the quality of the test video. One of the earliest reduced reference metrics was proposed by Webster et al. [34]. Their metric is a feature extraction metric that estimates the amount of impairment in a video by extracting localized spatial and temporal activity features using especially designed fliters. Other works include the work by Bretillon et al. [35] and the work by Carnec et al. [36]. Metrics in this class may be less accurate than the full reference metrics, but they are also less complex, and make real-time implementations more feasible. Nevertheless, synchronization between the original and impaired data is still necessary. Another merging hot topic recently is visual attention, and the related works [37–40] are proposed based on computational visual attention model (VAM). These works agreed that a unique saliency/importance map that topographically encodes for stimulus conspicuity over the visual scene is an efficient and plausible bottom-up control strategy. Among these VAMs, Ittis bottom-up saliency based visual attention model (SVAM) [37, 38] has demonstrated high correlation with human eye movements over static images and been used in various applications successfully and open an enlightening research direction for video quality evaluation [41, 42]. Note that NR metics is not presented in the dissertation because it is out of our scope.

1.3 Problem Statement

In wireless video streaming, due to the limited sustainable bandwidth of a receiver, a video often has to be coded (or transcoded or extracted from a scalable stream) at a reduced frame rate and/or frame size, so that each coded frame has adequate quality. A challenging problem is how to choose the appropriate spatial, temporal and amplitude resolutions (STAR), so as to achieve the best trade-off between picture quality and motion fluidity in the delivered video. Note that the amplitude resolution is controlled by the quantization stepsize (QS) or equivalently quantization parameter (QP). To solve this problem, one needs accurate models for both rate and quality, in terms of STAR. Another challenging problem is that the sustainable bandwidth of a wireless link often fluctuates in time, calling for adaptation of frame rate, frame size and QP. One naive approach would be to find the STAR that optimizes the perceptual quality over each short time duration based on the available instantaneous bandwidth. This will however create a video with rapidly fluctuating STAR, which may be annoying to the viewer. For example, variation in frame rate can cause visually annoying jitter artifacts. It is important to understand how does the variation of the STAR, individually and collectively, affect the perceived quality. Such understanding would enable us to impose proper constraints on the variation of the STAR, when adapting the STAR based on the time-varying bandwidth.

Prior work in video quality assessment is mainly concerned with applications where the frame rate and frame size of the video is fixed. The objective quality metric compares each pair of corresponding frames in deriving a similarity score or distortion between two videos with the same frame rate. The users in an application are very heterogeneous in their access link bandwidth, processing and display capabilities. The primary parameters of a video bitstream, which control the bandwidth requirement, include QS (controlling amplitude resolution), frame rate (controlling temporal resolution or TR) and frame size (controlling spatial resolution or SR). Given the bandwidth limitation and display resolution of a receiver, the encoder (as shown in Fig. 1.1, or a network transcoder or adaptor (as shown in Fig. 1.2) has to decide at which STAR to code, transcode or adapt a video, to achieve the best perceptual quality. Therefore, it is important to understand the impact of the STAR on the perceptual quality. On the other hand, studying the joint impact of all three dimensions on the perceptual quality is a complex and challenging task.

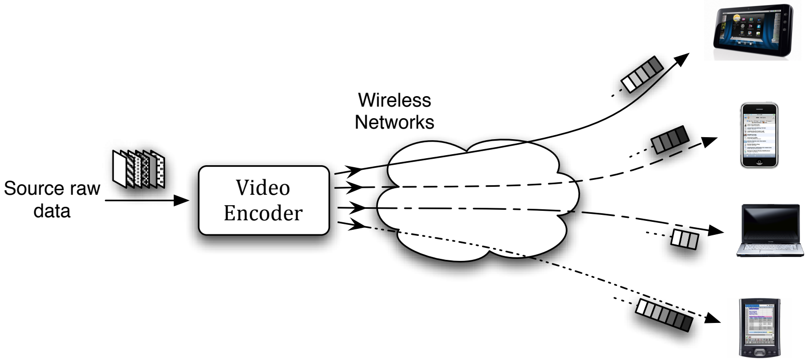


Figure 1.1: The multicast scenario using precode video bitstream.

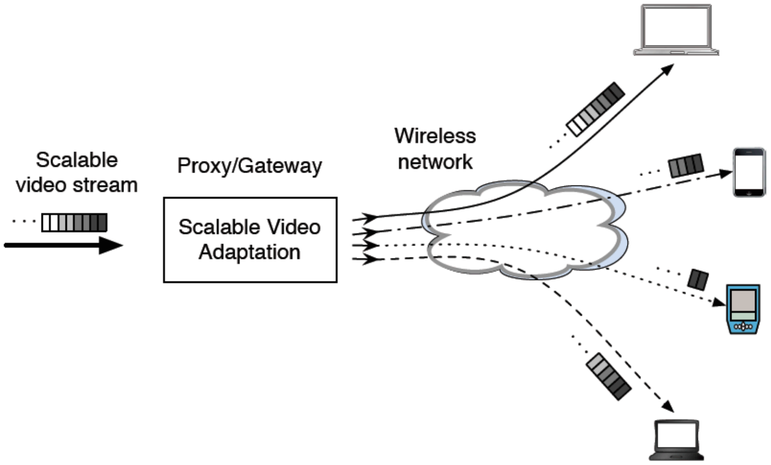


Figure 1.2: The multicast scenario using scalable video bitstream.

1.4 Reviews of Related Works

There have been several works studying the impact of frame rate artifacts on perceptual video quality. In a recent review of frame rate effect on human perception of video [43], it is found that frame rate around 15 Hz seems to be a threshold of humans' satisfaction level, but the exact acceptable frame rate varies depending on video content, underlying application, and the viewers. In addition, the authors of [44] proposed that the preferred frame rate decreases as video bandwidth decreases, and two switching bandwidths corresponding to the preferred frame rates were derived. The work in [45] investigated the preferred frame rate for different types of video. In [46], a particular high-motion type of coded video sequences (sports game) was explored. It was found that high spatial quality is more preferable than high frame rate for small screens. However, no specific quality metric, which can predict the perceived video quality, were derived in these works [43–46].

The works in [1, 2, 47] proposed quality metrics that consider the effect of frame rate. The work in [1] used logarithmic function of the frame rate to model the negative impact of frame rate dropping on perceptual video quality in the absence of compression artifacts. The model was shown to correlate well with subjective ratings for both CIF and QCIF videos. However, this model requires two content-dependent parameters, which may limit its applicability in practice. The metric proposed in [47] explores the impact of regular and irregular frame drop. The quality of each video scene is determined by weighting and normalizing a logarithm function of temporal fluctuation and the frame dropping severity. Finally, the overall quality of the entire video is the average of the quality indices over all video scene segments. The work in [2] also considers the impact of both regular and irregular frame drops and examines the jerkiness and jitter effects caused by different levels of strength, duration and distribution of the temporal impairment. However, [47] did not provide a single equation, which can predict the perceptual quality of regular frame drops, and even though [2] did, the proposed quality model has four parameters, and the authors did not consider how to derive these parameters from the underlying video.

Besides the study of frame rate impact on perceptual quality, Feghali et al. proposed a video quality metric [3] considering both frame rate and quantization effects. Their metric uses a weighted sum of two terms, one is the PSNR of the interpolated sequences from the original low frame-rate video, another is the frame-rate reduction. The weight depends on the motion of the sequences. The work in [48] extended that of [3] by employing a different motion feature in the weight. Besides Feghali's works, several works e.g., [4–6, 49, 50, 50, 51], have explored the impact of SR, TR and quantization artifacts (not QS directly) fully or partially on perceived video quality. Although the quality assessment in [4] and [6] include 3 and 6 different spatial resolutions, respectively, these works only involve an SR range from QCIF to CIF. The works in [5, 49] only include two SR's, QCIF and CIF. Hauske et. al [50] proposed the video quality metrics as a function of PSNR and

FR. Authors in [6,49] also proposed quality metrics considering STAR. The quality model in [49] is a function of the bitrate and a so called truncated bitrate ratio of SNR-scalability, while the quality model in [6], similar to the work in [50], is a function of PSNR, TR and SR. However, none of the tests reported in [4–6,49] were carried out on mobile devices. But even though the work [50] is for mobile devices, they claimed that model parameters are independent of video content, while on the contrary, authors in [51] believe that the model parameters depend on video contents. Nevertheless, the proposed [51] does not automatically estimate the model parameters. The work in [52] proposed a quality metric considering block-fidelity, content richness fidelity, spatial-textural, color, and temporal masking. They combined all these components into a quality index to predict the perceptual quality. This model involves sophisticated processing to extract content components from video sequences. Hence, it may not be applicable for practical application.

In our previous works [53,54], we investigated the impact of TR and QS on perceptual video quality, which was evaluated on larger screen size of laptop monitor, and proposed the video quality model considering the effect of TR and QS under a fixed SR (CIF). However, we believe that the form factor and the display screen may affect the viewing experiences. It is important to use a display environment similar to the actual mobile device during the subjective test. Therefore, in this work, we conduct the test on a mobile platform (Zoom2 from TI) with a screen size of 4.1-inch at a resolution of WVGA (854x480). In a preliminary study [55], we conduct a subjective test to explore the impact of SR and QS for mobile devices. By extending this study we further investigate the quality assessment considering the interaction of SR, TR, and QS, and propose a complete quality model in terms of SR, TR, and QS. Preliminary results of this study were reported in [56]. At the end, together with a rate model, which is also a function of STAR, the proposed quality model can help to determine the optimal STAR at which to encode a video or adapt a scalable video, given a target rate

However, in wireless video streaming, due to the limited sustainable bandwidth of a receiver, a video often has to be coded (or transcoded or extracted from a scalable stream) at a reduced frame rate and/or frame size, so that each coded frame has adequate quality. A critical issue is how to choose the appropriate spatial, temporal and amplitude resolutions (STAR), so as to achieve the best trade-off between picture quality and motion fluidity in the delivered video There have been several studies regarding the influence of temporal and amplitude resolutions, individually or jointly, on the perceptual quality [2,3,47,57,58]. Some of these works (e.g. [3,57]) consider the case where the FR and QS are fixed in the entire video, whereas others (e.g. [2,47]) consider the impact of FR variation, due to non-uniform and bursty packet losses, while authors in [58] proposed the variable frame rate control scheme based on the jerkiness of the video to adapt the frame rate simultaneously under fluctuate bandwidth environment. Nevertheless, they don't explore the quality impact while varying the frame rate.

1.5 Organization of the Dissertation

This dissertation is organized as follows:

In Chapter 2, we investigate the impact of frame rate and quantization on perceptual quality of a video for laptop devices. We first describe subjective tests conducted to evaluate the quality degradation due to temporal and quantization artifacts. We then describe the proposed quality model. The model uses the product of a spatial quality factor that assesses the quality of decoded frames without considering the frame rate effect and a temporal correction factor, which reduces the quality assigned by the first factor according to the actual frame rate. The complete model correlates well with the subjective ratings.

In Chapter 3, we conduct the subjective quality test and address the objective quality model considering the impact of SR, TR and QS on the TI Zoom2 mobile development platform (MDP). Subjective data reveal that the impact of SR, TR and QS can each be captured by a function with a single content-dependent parameter. The complete model correlates very well with the subjective ratings. Alternatively, we also investigate the relation between quantization-induced quality impact with respect to bit rate and PSNR as QS is unavailable.

In Chapter 4, we explore the impact of periodic frame rate or quantization variation on perceptual video quality. According the observation and data analysis of the test results, we propose to use different analytical models to characterize model the quality degradation due to variations in FR, QS and bit rate.

In Chapter 5, we discuss how to predict the parameters for the models derived in Chapters 2 and 3. First, we introduce several content features extracted from source video signals and utilized the GLM and CVE criteria to build up the optimum weighted linear combination of features to estimate the model parameters.

In Chapter 6, we summarize major contributions of this dissertation and discuss possible future work.

Chapter 2

Perceptual Quality of Video Considering both Frame Rate and Quantization Artifacts for Laptop Devices

In this chapter we explore the impact of frame rate and quantization on perceptual quality of a video. We propose to use the product of a spatial quality factor that assesses the quality of decoded frames without considering the frame rate effect and a temporal correction factor, which reduces the quality assigned by the first factor according to the actual frame rate. We find that the temporal correction factor follows closely an inverted falling exponential function, whereas the quantization effect on the coded frames can be captured accurately by a sigmoid function of the PSNR or by an exponential function of quantization stepsize (QS). The proposed model is analytically simple, with each function requiring only a single content-dependent parameter. The proposed overall metric has been validated using both our subjective test scores as well as those reported by others. For all seven data sets examined, our model yields high Pearson correlation (higher than 0.9) with measured MOS.

This chapter is organized as follows. We first address the subjective test configurations and the test results in Sec. 4.1. Section 2.2 presents the proposed objective metric, and validates its accuracy with our subjective test data. Section 2.3 compares our metric with those proposed in [1–3] on several datasets reported by others. Finally Section 2.4 concludes the chapter.

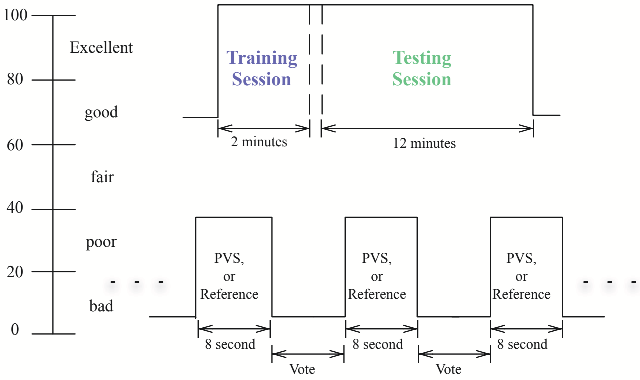


Figure 2.1: Subjective quality test setup.

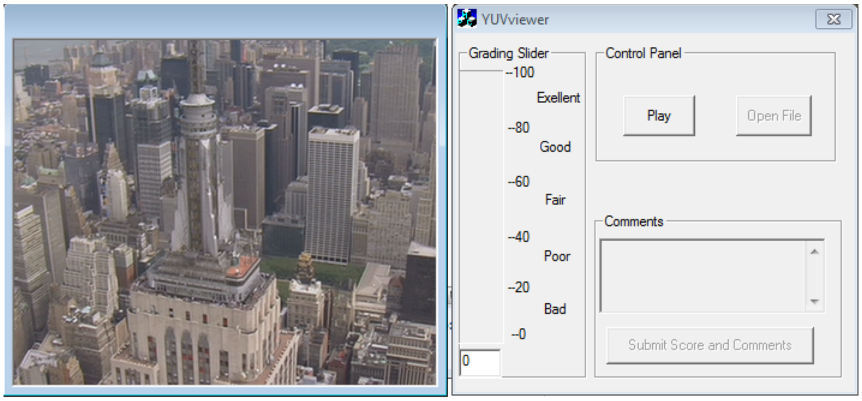


Figure 2.2: The multicast scenario using scalable video bitstream.

2.1 Subjective Quality Assessment

2.1.1 Test Sequence Pool

Seven video sequences, Akiyo, City , Crew, Football, Foreman, Ice, Waterfall, all in CIF (352×288) resolution at original frame rate 30 fps, are chosen from JVT (Joint Video Team) test sequence pool [59]. All these sequences are coded using scalable video model (JSVM912) [60], which is the reference software for the scalable extension of H.264/AVC (SVC) developed by JVT. For each sequence, one scalable bitstream is generated with four temporal layers corresponding to frame rates of 30, 15, 7.5, 3.75Hz, and each temporal layer in turn has four CGS quality layers created with QP equal to 28, 36, 40, and 44[1], respectively, using the coarse grain scalability (CGS). A processed video sequence (PVS) is created by decoding a scalable bitstream up to a certain temporal layer and a quality layer.

The subjective rating test for the seven sequences were done in two separate experiments. In the first experiment, 64 PVSs from four sequences ("Akiyo", "City", "Crew", and "Football") were rated, varying among four frame rates (30, 15, 7.5 and 3.75Hz) and four QP levels (28, 36, 40, and 44). In the second experiment, 60 PVSs from five sequences ("Akiyo", "Football", "Foreman", "Ice" and "Waterfall") are rated. In this case, we still test among four frame rates but only among 3 QP levels (28, 35, 40). This is because the results from the first session show that it is very hard for the viewers to tell the difference between QP=40 and 44. We included the two common sequences ("Akiyo", "Football") in both experiments, so that we can determine an appropriate mapping between the subjective ratings from two experiments, following the algorithm described in [22].

2.1.2 Test Configuration

The subjective quality assessment, illustrated in Figure 2.1, is carried out by using a protocol similar to ACR (Absolute Category Rating) described in [17]. In the test, a subject is shown one PVS at a time, and is asked to provide an overall rating at the end of the clip. The rating scale ranges from 0 (worst) to 100 (best) with text annotations shown next to the rating numbers as shown in Fig. 2.1 and the user interface is shown in Fig. 2.2. Most of the viewers for both of the subjective test are engineering students from Polytechnic Institute of New York University, with age 23 to 35. Other details regarding each experiment are given below.

1. The first experiment:

   In order to shorten the duration of the test, the experiment is divided into two subgroups. Each of them contains 38 processed video sequences and lasts about 14

[1] Different from JSVM default configuration utilizing different QPs for different temporal layers, the same QP is chosen among all temporal layers at CGS layer.

minutes. Each subgroup test consists of two sessions, a training session and a test session. The training session (about 2 minutes) is used for the subject to accustom him/herself to the rating procedure and ask questions if any. The training clips including PVSs from 'Soccer' and 'Waterfall' are chosen to expose viewers to the types and quality range of the testing clips. The PVSs in the test session (about 12 minutes) are ordered randomly so that each subject sees the video clips in a different order. Thirty one non-expert viewers who had normal or corrected-to-normal vision acuity participated in one or two subgroup tests. There are on average 20 ratings for each PVS.

2. The second experiment:

   Each subgroup contains 24 PVSs. The training clips (6 PVSs) are picked from the entire PVS pool except the sequences included in the testing session and the selections of testing points are uniformly distributed among the entire range. The sequences in the test session are also ordered randomly. Thirty three non-expert viewers who had normal or corrected-to-normal vision acuity participated in one or two subgroup tests. There are on average 16 ratings for each PVS.

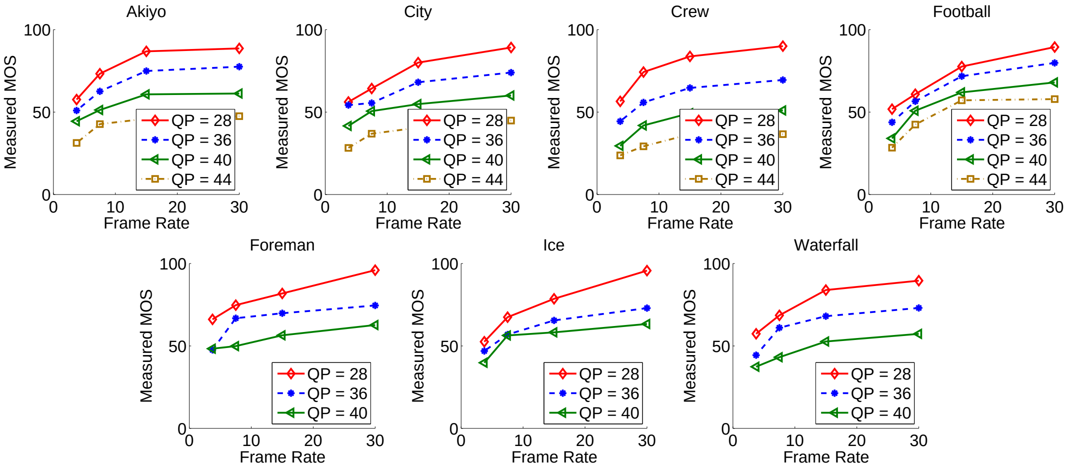


Figure 2.3: Measured MOS against frame rate at different QP. 95% confident interval of the first four training sequences is 20.29, and 21.38 is for the last three new sequences.

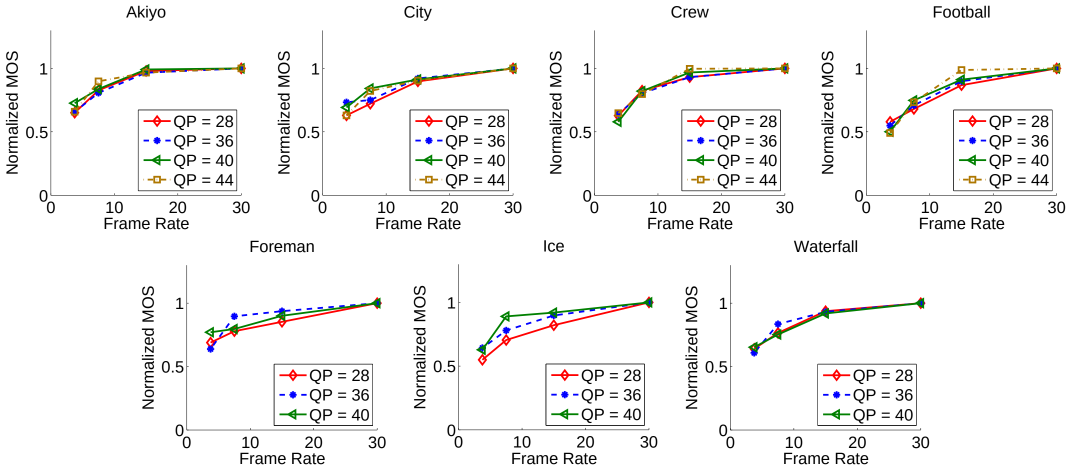


Figure 2.4: Normalized MOS against frame rate at different QP.

2.1.3 Data Post-Processing

Given the rating range from 0 to 100, different viewers' scores tend to fall in quite different subranges. The raw score data should be normalized before analysis. We first find the minimum and maximum scores given by each viewer for a specific source sequence, then normalize all viewers' score for this sequence by the average of minimum scores and the average of maximum scores among all subjects. We then average normalized viewer ratings for the same processed video sequence to determine its mean opinion score (MOS).

Let $u_{v\varsigma}$ denote the score of viewer v for each processed video sequence ς and V is the total number of viewers. As is often the case with subjective testing, some users' ratings are inconsistent either with other viewers' ratings for the same PVS, or with ratings for the other PVS's by the same viewer. We adopted, with some modification, the screening method recommended by BT.500-11 [16] designed for Single Stimulus Continuous Quality Evaluation (SSCQE) to screen our collected data. Our modification makes use of the fact that our test contains sequences that are different in frame rates under the same QP. If a viewer is consistent, then his/her rating for a lower frame-rate video should not be better than that for a higher frame-rate video. For each original sequence, we try to identify viewers who give lower ratings for higher frame-rate videos and do not consider the ratings by these viewers. Specifically, following [16] , we first determine the mean, standard deviation, and Kurtosis coefficients for each PVS using $\overline{u}_\varsigma$, σ_ς, and $\beta_{2\varsigma}$, respectively. Then we use the following procedure to identify viewers who give scores that are far from the average score by all viewers, as well as those viewers who give lower scores to higher frame-rate videos. Here, the mean and standard deviation for each PVS is defined as $\overline{u}_\varsigma = \frac{1}{V}\sum\limits_{v=1}^{V} u_{v\varsigma}$, and $\sigma_\varsigma = \left(\sum\limits_{v=1}^{V} \frac{(u_{v\varsigma}-\overline{u}_\varsigma)^2}{V-1}\right)^{1/2}$. The Kurtosis coefficient is obtained via β_2 test [16] for PVS ς, i.e., $\beta_{2\varsigma} = \frac{m_{4\varsigma}}{(m_{2\varsigma})^2}$, where $m_{n\varsigma} = \frac{1}{V}\sum\limits_{v=1}^{V}(u_{v\varsigma} - \overline{u}_\varsigma)^n$.

Recall that for each original video sequence α (e.g., $\alpha \in \{$Akiyo, City, Crew, Football$\}$), there are 4 frame rates and 4 QP's tested, with a total of 16 PVS. For each viewer v and original sequence α, we determine $P_{v\alpha}$, $Q_{v\alpha}$ and $R_{v\alpha}$ by the following procedure:

1. Starting with $P_{v\alpha} = 0$, and $Q_{v\alpha} = 0$, for each PVS ς of the same original sequence α;

```
if 2 < $\beta_{2\varsigma}$ < 4, then
  if $u_{v\varsigma} \geq \overline{u}_\varsigma + 2\sigma_\varsigma$, then $P_{v\alpha} = P_{v\alpha} + 1$;
  if $u_{v\varsigma} \leq \overline{u}_\varsigma - 2\sigma_\varsigma$, then $Q_{v\alpha} = Q_{v\alpha} + 1$;
else,
  if $u_{v\varsigma} \geq \overline{u}_\varsigma + \sqrt{20}\sigma_\varsigma$, then $P_{v\alpha} = P_{v\alpha} + 1$;
  if $u_{v\varsigma} \leq \overline{u}_\varsigma - \sqrt{20}\sigma_\varsigma$, then $Q_{v\alpha} = Q_{v\alpha} + 1$.
```

2. For the same original video α and for all PVS at the same QP, we compare the ratings obtained for different frame rates by viewer v, and count the numbers of times the viewer's rating for a lower frame rate PVS is higher than for a higher frame rate PVS. Specifically, let $u_{v\varsigma(f,QP)}$ indicate the rating given by viewer v, for a sequence ς with frame rate f and quantization parameter QP. Starting with $R_{v\alpha} = 0$, for each PVS ς belongs to the same original sequence α:

```
For all $f$ and $QP$,
  if $u_{v\varsigma(f/2,QP)}/u_{v\varsigma(f,QP)} \geq$ T, then $R_{v\alpha} = R_{v\alpha} + 1$;
  else if $u_{v\varsigma(f/2,QP)} > u_{v\varsigma(f,QP)}\Big|_{f=30,\ 15 \text{ and } 7.5}$,
```

then $R_{v\alpha} = R_{v\alpha} + 1$.

T is set to 1.2 based on our observation.

3. We reject ratings for sequence α by viewer v,
   if $R_{v\alpha} > 2$, $P_{v\alpha} > 1$ or $Q_{v\alpha} > 1$ [16]

Above process allows us to discard, for each original sequence, all the ratings from a viewer when his/her ratings are significantly distant from the average scores for at least 2 PVSs. In addition, it also excludes all ratings by a viewer for each original sequence, when his/her ratings for a lower frame rate video is better than for a higher frame rate video at least 3 times, among all PVSs for this sequence.

After screening there are on average 15 and 14 user ratings for each PVS in first and second experiments, respectively. Figure 2.3 presents the subjective test results. We see that no matter what QP level is, MOS reduces consistently as the frame rate decreases. In order to examine whether the reduction trend of the MOS against the frame rate is independent of the quantization parameter, we plot in Figure 2.4, the normalized MOS, which is the ratio of the MOS with the MOS at the highest frame rate (30Hz in our case), at the same QP. We see that these normalized curves corresponding to different QPs almost overlap with each other, indicating that the reduction of the MOS with frame rate is quite independent of the QP.

2.2 Proposed Quality Metric

As described earlier, results in Figures 2.3 and 2.4 suggest that the impact of frame rate and that of quantization is separable. Based on this observation, we propose the following metric consisting of the product of two functions:

$$\mathrm{VQMTQ}(\mathsf{PSNR}, f) = \mathrm{SQF}(\mathsf{PSNR}, f_{\max})\mathrm{TCF}(f; q) \tag{2.1}$$

where f represents the frame rate and PSNR is the average of PSNRs of decoded frames. As described earlier, the first term $\mathrm{SQF}(\mathsf{PSNR})$ measures the quality of encoded frames without considering the frame rate effect. The second term models how the MOS reduces as the frame rate decreases. The specific forms of the function $\mathrm{TCF}(f)$ and $\mathrm{SQF}(\mathrm{PSNR})$ are described in Sec. 2.2.1 and 2.2.2, respectively.

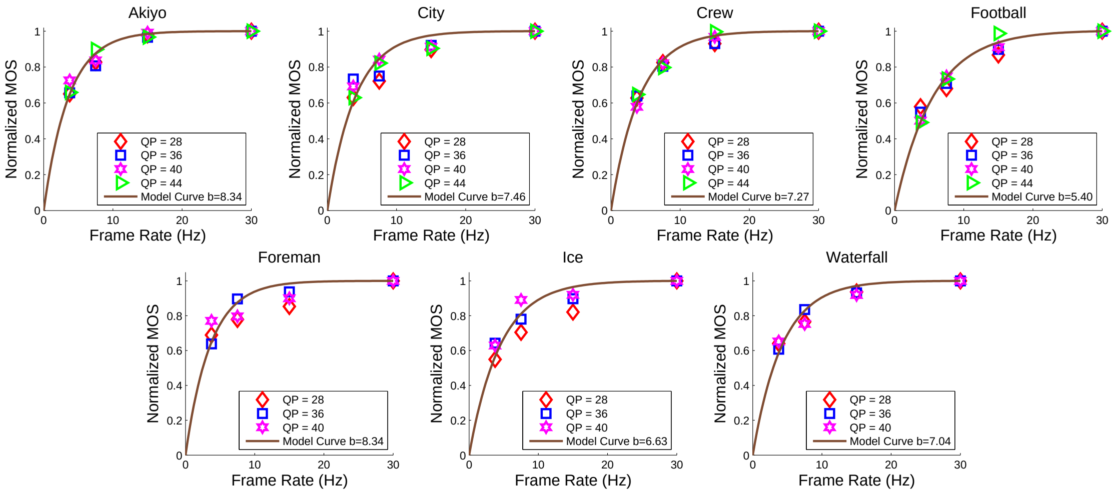


Figure 2.5: The measured normalized MOS and temporal correction factor (TCF) against frame rate. PCC=0.95

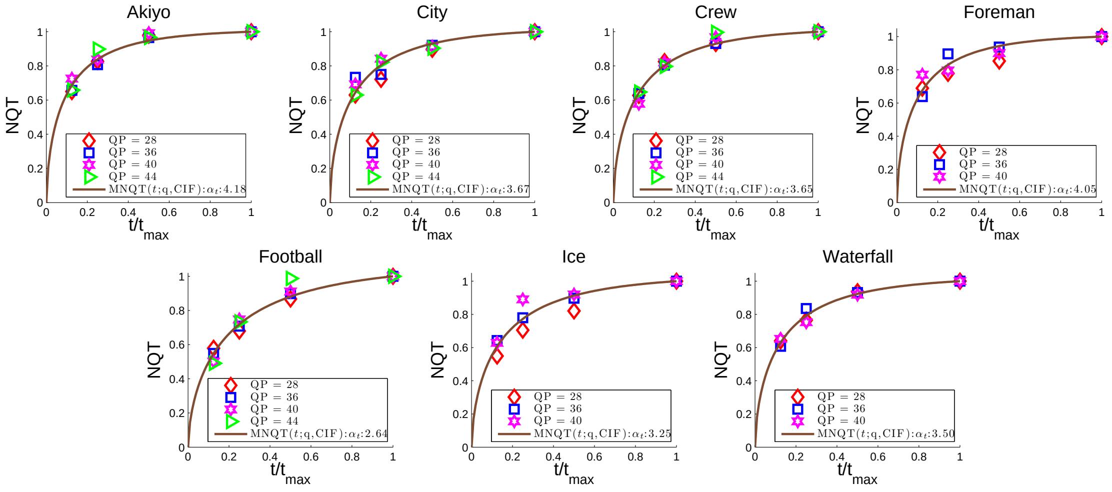


Figure 2.6: Predicted vs. measured MOS for `DataSet#1` against normalized FR by the metric proposed in (2.5). PCC = 0.968, RMES = 0.034.

2.2.1 Temporal Correction Factor

In a prior work [61], we have investigated the impact of the frame rate on the perceptual quality of uncompressed video, and found that the normalized quality can be modeled very accurately by an inverted exponential falling function. Here we adopt the same function:

$$\mathrm{TCF}(f) = \frac{1 - e^{-b\frac{f}{f_{\max}}}}{1 - e^{-b}}. \tag{2.2}$$

As can be seen in Figure 2.5, this function can predict the normalized MOS very well. For uncompressed video, normalized MOS is defined as,

$$\mathrm{NMOS}(f) = \frac{\mathrm{MOS}(f)}{\mathrm{MOS}(f_{\max})}, \tag{2.3}$$

and for compressed video at the same QP, it is defined as

$$\mathrm{NMOS}(\mathrm{QP}, f) = \frac{\mathrm{MOS}(\mathrm{QP}, f)}{\mathrm{MOS}(\mathrm{QP}, f_{\max})}. \tag{2.4}$$

We can see that the fitting is quite accurate for all sequences. Note that the parameter b characterizes how fast the quality drops as the frame rate reduces, with a smaller b indicating a faster drop rate. The b values for different sequences are provided in Figure 2.5. As expected, sequences with higher motion have faster drop rates (smaller b). To demonstrate the influence of the video content on the parameter, Figure 2.7 shows the TCF curves for different videos. We can clearly see that b is larger for slower motion sequences.

Since the development of the model in (2.2), as part of the research presented in Chapter 3, we have found that the TCF can also be modeled accurately using a generalized inverse exponential function of the form:

$$\mathrm{MNQT}(f) = \frac{1 - e^{-\alpha_t (\frac{f}{f_{\max}})^{\beta_t}}}{1 - e^{-\alpha_t}}. \tag{2.5}$$

where $f_{\max}$ is the maximum TR (here, $f_{\max}$ = 30Hz). where β_t = 0.63 is a constant. Figure 2.11 compares the predicted and measured TCF data with high PCC = 0.968, RMSE = 0.034.

The model in (2.2, 2.5) are chosen by comparing several one-parameter functions, including the exponential falling function in Eq. (2.2), the power function $(\frac{f}{f_{\max}})^b$, and the logarithmic function $\frac{\log(1+b\frac{f}{f_{\max}})}{\log(1+b)}$. During the model regression, we choose Root Mean Square Error (RMSE) and Pearson Correlation (PCC) to validate the goodness of prediction. Given a K data pairs (X_k, Y_k), the Pearson Correlation is defined as

$$\rho_{X,Y} = \frac{\sum_{k=1}^{K}(X_k - \bar{X})(Y_k - \bar{Y})}{\sqrt{\sum_{k=1}^{K}(X_k - \bar{X})^2}\sqrt{\sum_{k=1}^{K}(Y_k - \bar{Y})^2}}, \tag{2.6}$$

where X_k and Y_k are the paired data and they are usually referred to the subjective ratings and the predicted quality index, respectively, and $\bar{X}$ and $\bar{Y}$ are the means of the respective data sets. We will use this metric as the gauge to measure the fitting performance of the model in the current and following Chapters. Table 2.1 summarizes the PCC and RMSE obtained with different fitting functions on five data sets (`DataSet#1-#5` will be described in Section 2.3). It is shown that the inverted exponential function in Eq. (2.2) is the best.

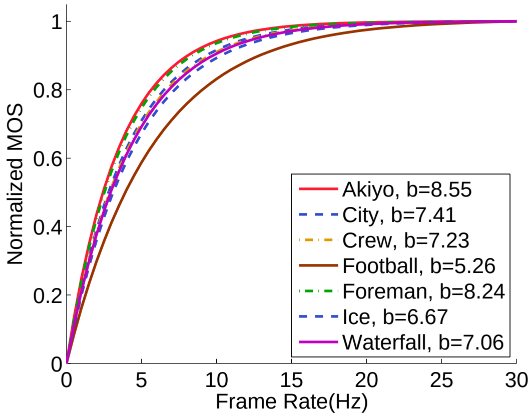


Figure 2.7: Temporal correction factor for different test sequences.

2.2.2 Spatial Quality Factor Using PSNR of Decoded Frames

In this subsection, we present the proposed model for the spatial quality, which is the perceptual quality of encoded frames without considering the frame rate effect. Signal-to-Noise Ratio (PSNR) is a commonly adopted metric for measuring quality of video with

Table 2.1: Goodness of Fitting by three functional forms of TCF

Quality Metrics	DataSet#1	#2	#3	#4	#5
Inverted exponential function in (2.5)					
RMSE	3.4%	2.0%	2.5%	7.2%	6.7%
PCC	0.968	0.978	0.982	0.965	0.950
Inverted exponential function in (2.2)					
RMSE	5.5%	2.3%	2.1%	3.8%	4.8%
PCC	0.95	0.98	0.98	0.99	0.97
Power function					
RMSE	4.2%	4.7%	4.4%	8.5%	7.8%
PCC	0.96	0.94	0.93	0.95	0.92
Logarithm function					
RMSE	4.7%	5.2%	3.6%	10.2%	9.5%
PCC	0.95	0.93	0.94	0.92	0.87

encoding distortion. From the test results shown in Figure 2.8 , we see that, in an intermediate range of PSNR, the perceived quality correlates quite linearly with PSNR. However, the human eyes tend to think video with very low PSNR as equally bad and those with very high PSNR as equally good. Taking into account of this saturation effect of the human vision, we propose to use a sigmoid function, following the model in [62],

$$\mathrm{SQF}(\mathsf{PSNR}) = \hat{Q}_{\max}(1 - \frac{1}{1 + e^{p(\mathsf{PSNR}-s)}}), \tag{2.7}$$

where $\hat{Q}_{\max}$ is the quality rating given for highest quality video (for uncompressed video with PSNR $= \infty$), and p, s are model parameters. Note that although the rating scale is $[0, 100]$ in our subjective test, viewers do not give a score of 100 even for videos with very high quality, as is commonly observed in subjective test. Because our subjective test does not include uncompressed video, we derive $\hat{Q}_{\max}$ for a video sequence from Q_o, the measured MOS for the same video sequence decoded at the lowest QP and highest frame rate. We have found that $\hat{Q}_{\max} = 1.04 {\cdot} Q_o$ yields a good result for all the sequences. Further we found for all sequences, $p = 0.34$ gives good result. Therefore we only vary s when fitting the model to the measured MOS data. Figure 2.8 compares the MOS obtained for sequences at 30Hz with those obtained using the model in (2.7). We can see that the model, with a single parameter s, is very accurate with a PCC of 0.996. In addition to PSNR, we examine the effectiveness of SSIM Index [63] to predict the measured MOS at the highest

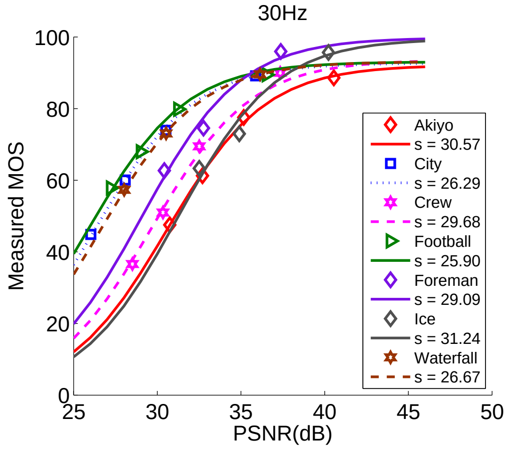


Figure 2.8: Measured and predicted MOS against PSNR for sequences coded at highest frame rate (30Hz). PCC$= 0.996$.

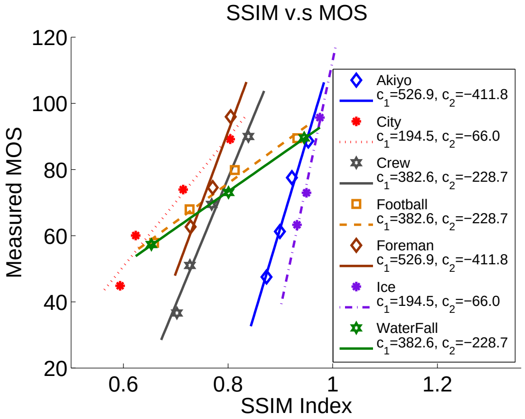


Figure 2.9: Measured and predicted MOS against SSIM index for sequences coded at highest frame rate (30Hz). PCC=0.99

frame rate, which has been shown to be more correlated than PSNR to perceptual spatial quality in other works. We compute SSIM for each decoded video frame and average SSIM over all frames. Figure 2.9 shows that MOS is quite linearly related to SSIM, i.e.,

$$Q(\mathsf{SSIM}) = c_1 \cdot \mathsf{SSIM} + c_2. \tag{2.8}$$

Although SSIM predict MOS with a high Pearson Correlation coefficient of 0.99, it requires two parameters, which vary significantly among sequences, while the SQF in (2.7) only needs one content-dependent parameter and also has a high PCC. Therefore, for our proposed model, we use the PSNR based function in (2.7) for predicting the spatial quality.

2.2.3 Spatial Quality Factor Using QS

The PSNR model (2.7) introduced above is a full reference method as described in Chapter 1. It requires both original and distorted video sequences in order to derive the PSNR value. Although the prediction accuracy is promising, this quality model will not be flexible and applicable if the retrievable information of the original video signals are limited, especially for those application more involving the real-time video transmission. Instead of using the PSNR of decoded frames, we also explored the relation between the normalized spatial quality factor (defined as MOS($q, f_{\max}$)/MOS($q_{\min}, f_{\max}$)) and the QS. Based on the data shown in Fig. 2.10, we directly model the relation with quantization stepsize q (QS) using an exponential function, i.e.,

$$Q_q(q) = e^c e^{-c\frac{q}{q_{\min}}}, \tag{2.9}$$

where c is the model parameter, $q_{\min} = 16$ as QP = 28 (QS is defined as $q = 2^{\frac{QP-4}{6}}$ in H.264/AVC standard). Figure 2.10 shows that the model captures the quantization-induced quality variation very well at $f_{\max}$.

In our later study described in Chapter 3, we found that the NQQ can also be modeled by an inverse exponential function of the inverse QS, i.e.,

$$\frac{1 - e^{-\alpha_q (\frac{q_{\min}}{q})^{\beta_q}}}{1 - e^{-\alpha_q}}, \tag{2.10}$$

where α_q is the model parameter and β_q=1 is a constant. Figure 2.11 shows the predicted curves with measured data and fitting is very accurate with PCC=0.995, RMSE=0.014.

2.2.4 Video Quality Metric Considering Temporal Resolution and Quantization(VQMTQ)

Combining Eqs. (2.1, 2.2, 2.7), we obtain the proposed video quality metric considering both temporal and quantization effect , when using PSNR to model the spatial quality

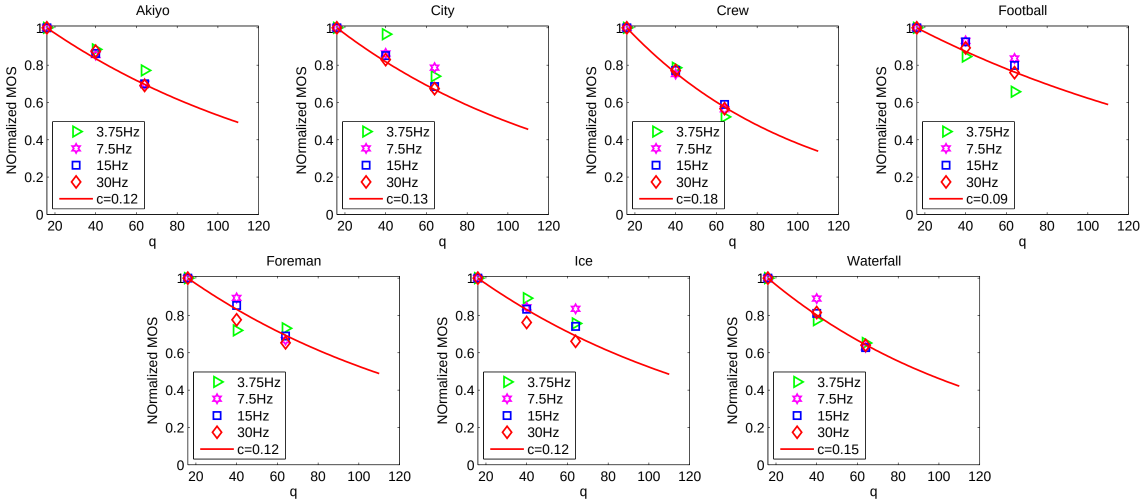


Figure 2.10: Normalized quality versus the quantization stepsize (NQQ) for different frame rates f. Points are measured data and curves are predicted quality for $t = 30$ Hz, using Eq. (3.6). PCC = 0.99

factor:

$$\mathrm{VQMTQ}_1(\mathsf{PSNR}, f) = \hat{Q}_{\max}\left(1 - \frac{1}{1 + e^{p(\mathsf{PSNR}-s)}}\right)\frac{1 - e^{-b\frac{f}{f_{\max}}}}{1 - e^{-b}}. \tag{2.11}$$

We plot predicted quality using this model together with measured MOS in Figure 2.12. We can see that predicted curves fit the measured MOS very well for most cases.

When using the QS to capture the spatial quality factor, the overall quality model can be written as

$$\mathrm{VQMTQ}_2(q, f) = Q_{\max} Q_q(q; f_{\max}) \mathrm{TCF}(f; q), \tag{2.12}$$

where $Q_{\max} = Q(q_{\min}, f_{\max})$ is the MOS for the video coded at $q_{\min}$ and $t_{\max}$; $Q_q(q; f_{\max}) = Q(q, f_{\max})/Q(q_{\min}, f_{\max})$ is the normalized quality versus quantization stepsize (NQQ) under the maximum frame rate $f_{\max}$

Combining Eqs. (2.2, 2.9 and 2.12), the overall video quality model can be expressed as

$$\mathrm{VQMTQ}_2(q, f) = Q_{\max}\frac{e^{-c\frac{q}{q_{\min}}}}{e^{-c}}\frac{1 - e^{-d\frac{f}{f_{\max}}}}{1 - e^{-d}}. \tag{2.13}$$

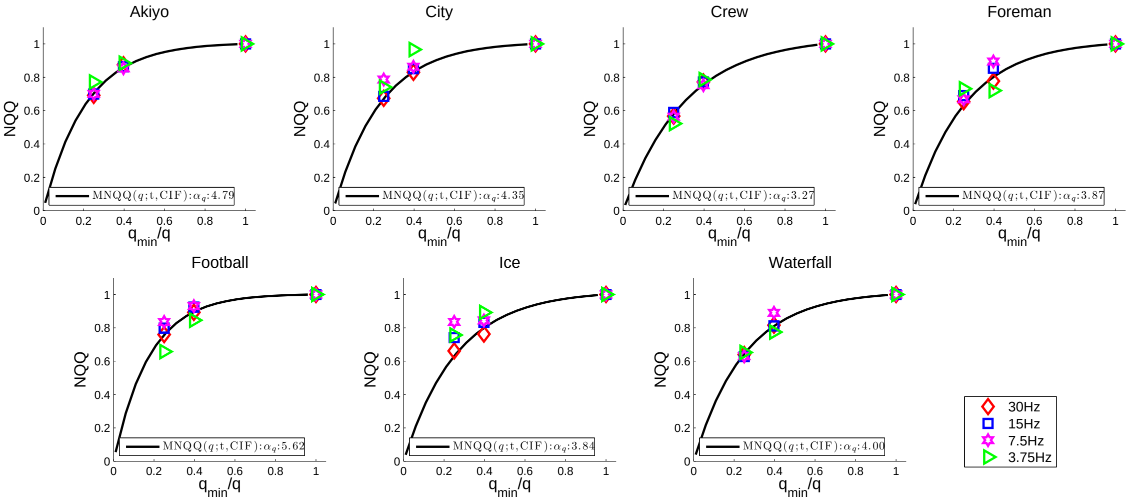


Figure 2.11: Predicted vs. measured MOS for `DataSet#1` against QS by the metric proposed in (2.10). PCC = 0.995, RMES = 0.014.

Recall that $Q_{\max}$ is the MOS given for the video at $q_{\min}$ and $f_{\max}$, which is set to 90 according to our subjective test data. The accuracy of this model is shown Fig. 2.13.

Combining eqs. (2.10, 2.5), we also yield an alternative model:

$$\mathrm{VQMTQ}_3(q,f) = Q_{\max}\frac{1-e^{-\alpha_q(\frac{q_{\min}}{q})^{\beta_q}}}{1-e^{-\alpha_q}}\frac{1-e^{-\alpha_t(\frac{f}{f_{\max}})^{\beta_t}}}{1-e^{-\alpha_t}}. \tag{2.14}$$

We plot the predicted quality with measured data in Fig. 2.14 using (2.14) with PCC = 0.974, RMSE = 0.038. Table 2.2 summarizes the parameters, PCC and RMSE for all VQMTQ_1, VQMTQ_2, and VQMTQ_3. Note that d is obtained by fitting the measured data from 3 QP's (28, 36, 40), while parameter b is obtained by fitting the measured data from 4 QP's (28, 36, 40, 44) for only Akiyo, City, Crew and Football.

2.3 Performance Comparison

In this section, we compare our proposed models in (2.2,2.5) with three metrics proposed in [1], [2] and [3]. We apply these models to a total of 7 data sets and compare their performance. Table 2.3 summarizes these data sets. `DataSet#2-#5` contain uncompressed video at different frame rates, and `DataSet#1`, `#6`, `#7` are compressed video obtained with different frame rates and QPs.

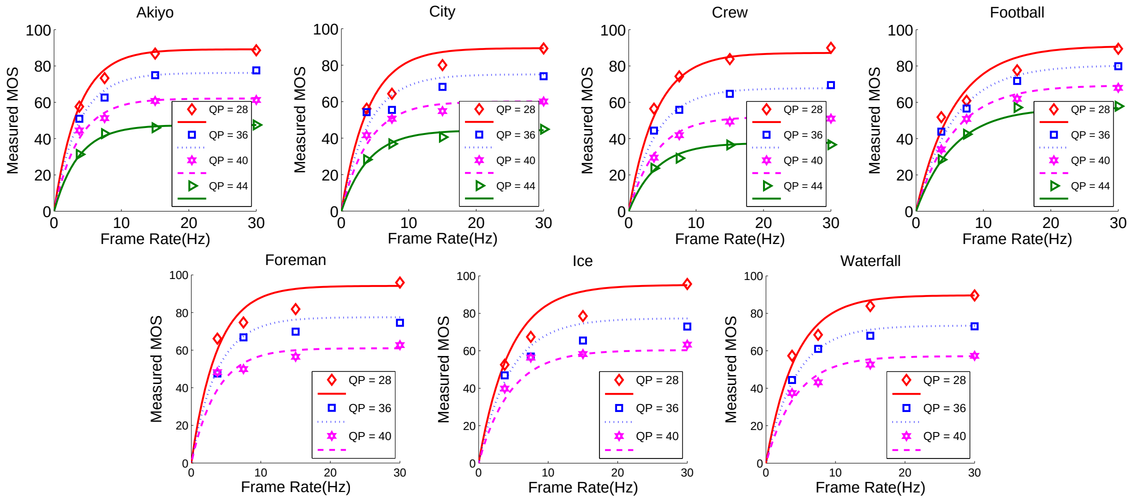


Figure 2.12: Predicted (in curve) (see Eq. 2.11) and measured (in points) MOS for videos coded at different QP and frame rate (`DataSet#1`). PCC=0.98.

The models in [1] and [2] only consider the effect of frame rate. The model in [1], called negative impact of frame-dropping on visual quality is given by,

$$\mathrm{NIFVQ}(f) = a_1 \cdot [\log(30) - \log(f)]^{a_2}, \tag{2.15}$$

with model parameters a_1 and a_2. In particular, they defined $\mathrm{NIFVQ} = 5 -$MOS as the degraded quality. It is noted that they assume the quality of all reference or highest frame-rate videos is 5, and the quality at lower frame rates are decreased according to (2.15).

The metric in [2] models the jerkiness of the video and is given by:

$$\mathrm{jerkiness}(f) = k_1 + \frac{k_2}{1 + e^{k_3 \cdot f + k_4}}. \tag{2.16}$$

In order to compare these two models and our proposed model, we apply all three models to the first five data sets in Table 2.3. To compare our model with the metrics in [1] and [2], which do not consider the quantization effect, we only evaluate the TCF portion of our model, and apply these three metrics to normalized MOS. For each data set, we normalized the MOS given for a test sequence at a particular frame rate (and quantization level) by the MOS for the same sequence at the highest frame rate (and at the same quantization level for `DataSet#1`). We apply all three models to the normalized MOS and determine the model parameters by least squares fitting. Figures 2.15 - 2.19 compare the predicted

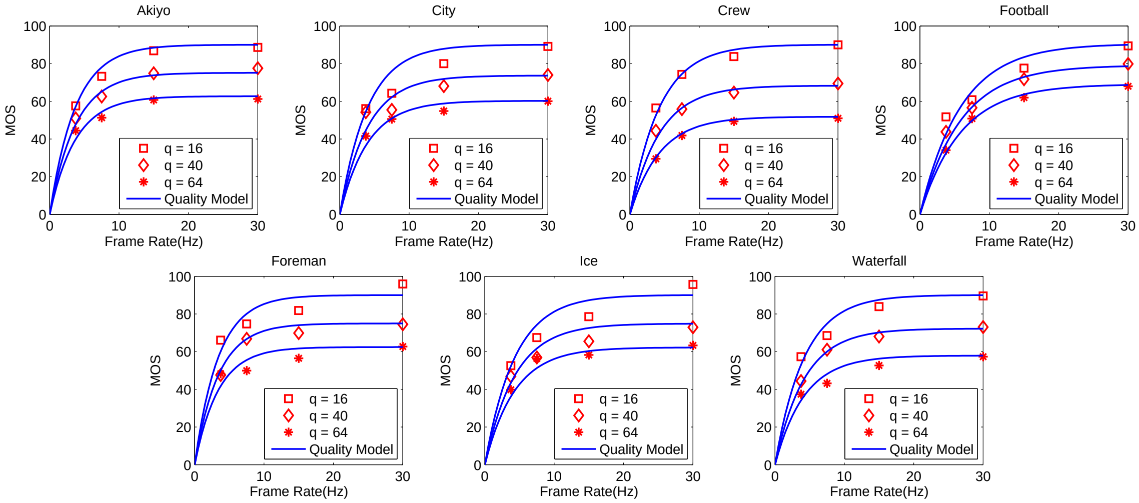


Figure 2.13: Quality vs. quantization stepsize and frame rate. Points are measured MOS data; curves are predicted quality using Eq. (2.13)

quality indices by these three models and the actual normalized MOS for the five data sets. Table 2.4 summarizes the Pearson correlation coefficients, and it is demonstrated that all three models can predict the normalized MOS very well with high correlation. Although the other two models have slightly higher correlation values for some datasets, our proposed model only uses one parameter to model the normalized MOS, instead of 2 and 4 parameters in the models proposed by [1] and [2], respectively. Note that the subjective ratings of `DataSet#5` in [2] show different trends for different sequences at the very low end of the frame rates. The model proposed in [2] was able to follow the subjective ratings accurately because it has four parameters. However, it is not clear whether these inconsistent trends are due to viewer inconsistencies at very low frame rates. We should note that the work in [2] actually applied the model (2.16) to the average subjective ratings over all test sequences. The average quality actually decreased with the frame rate in the same trend indicated by the model given in (2.2) and (2.15).

The model in [3] considers both frame rate and quantization effect and is given by,

$$\overline{\mathrm{QM}}(\overline{\mathsf{PSNR}}, f) = \beta_1 \cdot \overline{\mathsf{PSNR}} + M_n^{\beta_2} \cdot (30 - f), \tag{2.17}$$

where β_1 and β_2 are model parameters. Here M_n represents normalized motion vector magnitude, which is defined as the average of the motion vector magnitudes at the top 25%

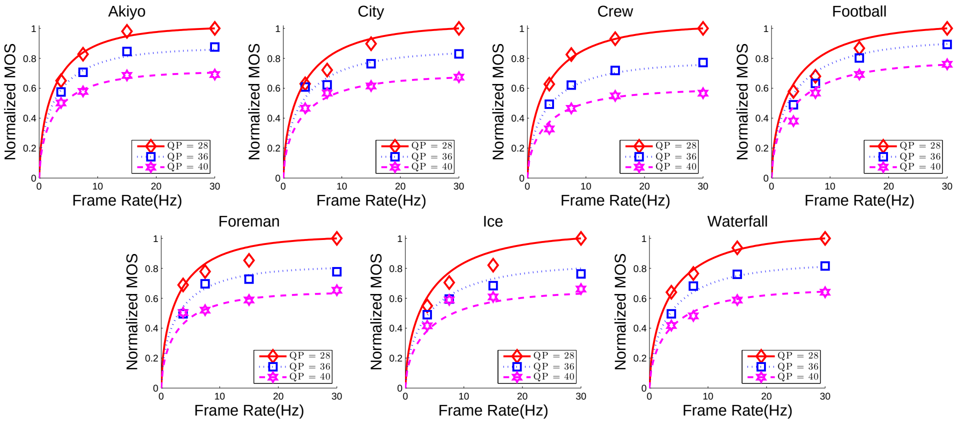


Figure 2.14: Predicted (in curve) (see Eq. 2.11) and measured (in points) MOS for videos coded at different QP and frame rate (DataSet#1). PCC=0.98.

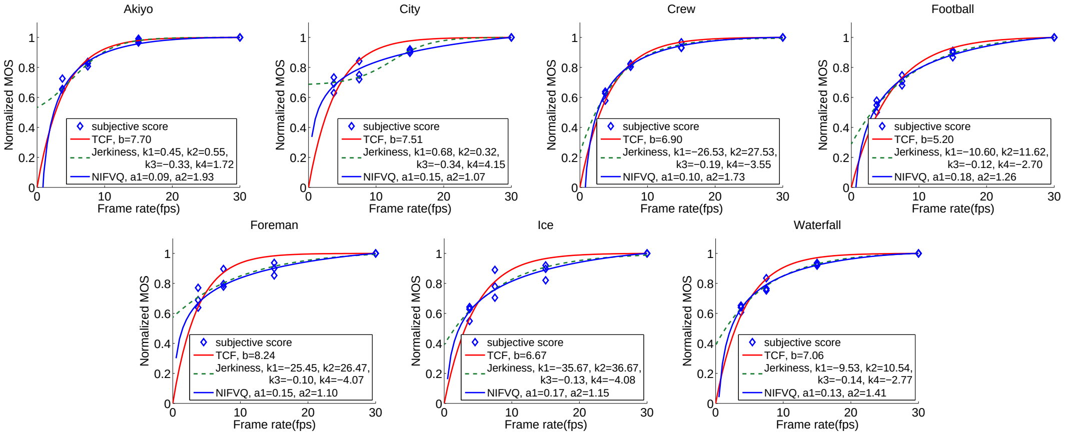


Figure 2.15: Predicted v.s measured normalized MOS for DataSet#1 by three metrics

of all MV magnitudes normalized by the width of the display frame. Note that in [3], a low frame-rate video is interpolated to the full frame-rate by using frame repetition. The $\overline{\mathsf{PSNR}}$ in (2.17) is the average PSNR of all frames, including interpolated frames. This $\overline{\mathsf{PSNR}}$ depends on the frame rate, and is significantly lower than the average PSNR computed from non-interpolated frames. To compare this model with our VQMTQ model, we apply them to our dataset (DataSet#1), as well as DataSet#6 and #7, all containing compressed video with different frame rates and quantization levels. Figure 2.20 shows predicted MOS vs. measured MOS for DataSet#1 by the $\overline{\mathrm{QM}}$ model. We see that the fit is not very good, significantly worse than the fit using the VQMTQ model given in Figure 2.12 earlier.

Table 2.2: Optimal parameters and model accuracy for VQMTQ

	akiyo	city	crew	football	foreman	ice	waterfall	Ave
	obtained by least square fitting using $\mathrm{VQMTQ}_1(\mathsf{PSNR}, f)$							
s	30.57	26.3	29.68	25.9	29.09	31.24	26.67	-
b	8.55	7.41	7.23	5.25	8.24	6.67	7.06	-
RMSE	2.47%	5.39%	2.23%	3.90%	4.0%	4.87%	7.12%	4.29%
PCC	0.99	0.97	0.99	0.98	0.98	0.97	0.95	0.98
	obtained by least square fitting using $\mathrm{VQMTQ}_2(q, f)$							
c	0.12	0.13	0.18	0.09	0.12	0.12	0.15	
d	7.70	7.51	6.90	5.20	8.24	6.67	7.06	
RMSE	3.06%	6.41%	2.50%	4.54%	5.49%	5.38%	3.65%	4.40%
PCC	0.98	0.94	0.99	0.98	0.94	0.95	0.98	0.96
	obtained by least square fitting using $\mathrm{VQMTQ}_3(q, f)$							
α_q	4.79	4.34	3.27	5.62	3.86	3.83	4.00	
α_t	4.18	3.66	3.64	3.46	4.04	3.24	3.49	
RMSE	2.09%	3.83%	1.79%	5.66%	4.25%	5.31%	1.81%	3.80%
PCC	0.99	0.97	0.99	0.98	0.97	0.96	0.99	0.97

We further show the scatter plots of predicted MOS vs. measured MOS by the two methods in Figure 2.21. It can be seen that the VQMTQ model is more linearly correlated with the measured MOS.

We next compare these two models using `DataSet#6`. Because we do not have access to the actual video clips used in `DataSet#6`, we are not able to compute the PSNR of decoded frames and hence are not able to apply our VQMTQ model to the entire dataset. Instead we only apply the TCF model to the normalized MOS for a subset of clips that are coded with the same QP (QP=6) at different frame rates, using the MOS for the clip coded at the highest frame rate as the normalizing factor. Figure 2.22 shows the predicted MOS values by the TCF and $\overline{\mathrm{QM}}$ models vs. the measured MOS values for this dataset (`DataSet#6`). Note that in the plot the predicted curve by TCF is obtained after we multiply the predicted NMOS value (by TCF) using the MOS given for the highest frame rate clip.

Finally we compare the $\overline{\mathrm{QM}}$ and VQMTQ model using the dataset reported in [5] (`DataSet#7`). Figure 2.24(a,b) show the scatter plots of predicted MOS vs. measured MOS using $\overline{\mathrm{QM}}$ and VQMTQ models. In the results shown for all other data sets, the

Table 2.3: Data Set Description

Data Sets	Source Definition
DataSet#1	7 CIF sequences used in this chapter each with 4 frame rates (30, 15, 7.5, 3.75 Hz) and four quantization levels. Normalized MOS is obtained by Eq. 2.4
DataSet#2	6 uncompressed CIF sequences used in [61], each with 5 frame rates (30, 15, 10, 7.5, 6 Hz). Normalized MOS is obtained by Eq. 2.3
DataSet#3	6 uncompressed QCIF sequences used in [61], each with 5 frame rates (30, 15, 10, 7.5, 6 Hz). Normalized MOS is obtained by Eq. 2.3
DataSet#4	4 uncompressed CIF sequences used in [1], each with 7 frame rates (30, 15, 10, 7.5, 6, 5, 3Hz). Normalized MOS is obtained by Eq. 2.3.
DataSet#5	7 uncompressed CIF sequences used in [2], each with 6 frame rates (25, 12.5, 8.33, 6.25, 5, 2.5Hz). Normalized MOS is obtained by Eq. 2.3.
DataSet#6	The subset of 5 CIF sequences used in [3], obtained with 3 frame rates (30, 15, 7.5 Hz) at the same QP (QP=6)
DataSet#7	5 CIF sequences used in [5], each with 3 frame rates (30, 15, 7.5 Hz) and 4 bit rate levels

parameter p in our VQMTQ model in Eq. (2.11) was fixed at 0.34. But for this data set, we found that p was sequences dependent. Therefore, we determine all three parameters p, s, and b through least square fitting. As can be seen, the VQMTQ model correlates with the measured MOS much better than the $\overline{\mathrm{QM}}$ model. We note however that the VQMTQ model in this case uses 3 parameters, whereas the $\overline{\mathrm{QM}}$ model uses 2 parameters. Figure 2.23 shows the measured and predicted MOS by VQMTQ vs. the bit rate. It is encouraging to see that

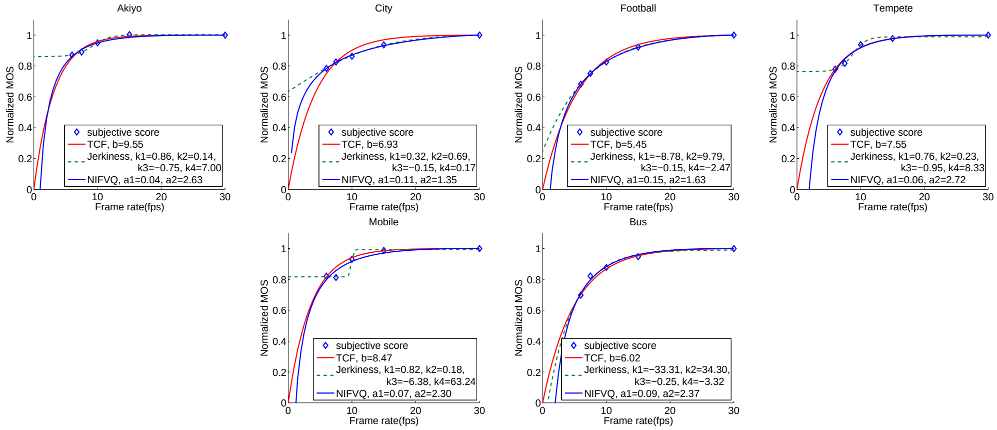


Figure 2.16: Predicted v.s measured normalized MOS for `DataSet#2` by three metrics

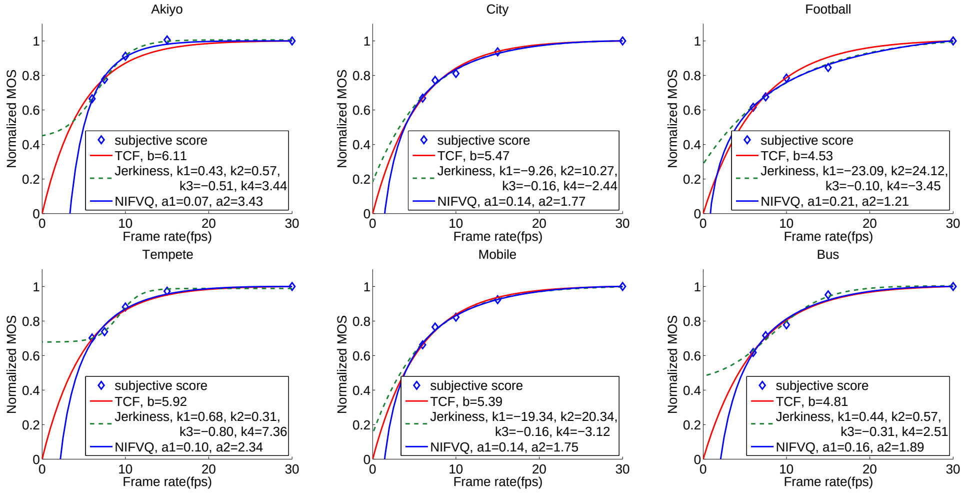


Figure 2.17: Predicted v.s measured normalized MOS for `DataSet#3` by three metrics

for the entire bit rate range, the VQMTQ method was able to correctly predict the frame rate that leads to the highest perceptual quality at a given bit rate, even though the actual predicted MOS do not fit the measured MOS perfectly.

We further apply the QSTAR model introduced in Chapter 3 on `DataSet#1`. Table 2.4 summarizes the PCC and RMSE and shows that the fitting it quite well. This indicated that the proposed models in Chapter 3 are improved and this function form (e.g., inverse exponential) can be utilized for predicting the quality degradation on both laptop (larger screen size) and mobile (smaller screen size) devices. However, the function form does not show much improvement on other datasets.

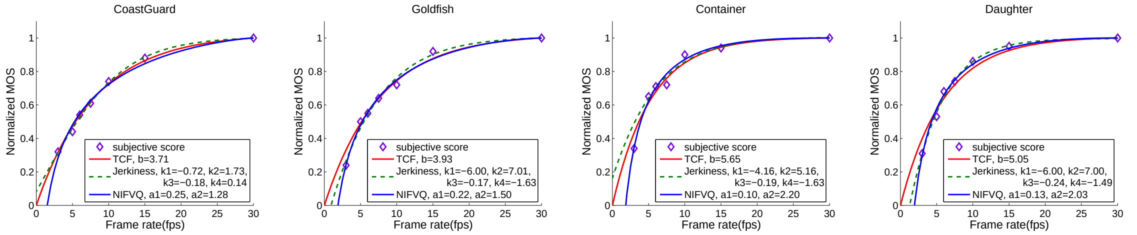


Figure 2.18: Predicted v.s measured normalized MOS for `DataSet#4` [1] by three metrics

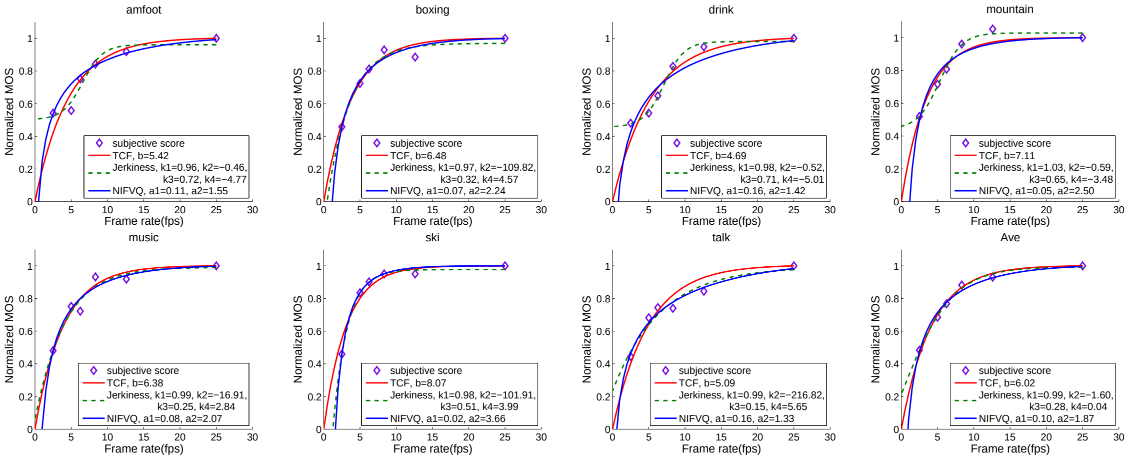


Figure 2.19: Predicted v.s measured normalized MOS for `DataSet#5` [2] by three metrics

2.4 Summary

This work is concerned with the impact of quantization and frame rate on the perceptual quality of a video. We demonstrate that the degradation of the perceptual quality due to quantization and frame-rate reduction can be accurately captured by two functions separately (a sigmoid function of the average PSNR of decoded frames and an inverted falling exponential function of the frame rate). Besides the PSNR model, we further propose to use an exponential function or inverse exponential function of the inverse QS to model the quality with quantization stepsize. Each function has a single parameter that is video-content dependent. all the proposed models are shown to be highly accurate, compared to the subjective ratings from our own subjective tests as well as test results reported in several other papers. Even though the overall VQMTQ model is validated for CIF video only, we expect the model to be applicable to videos at other resolutions as well. In fact,

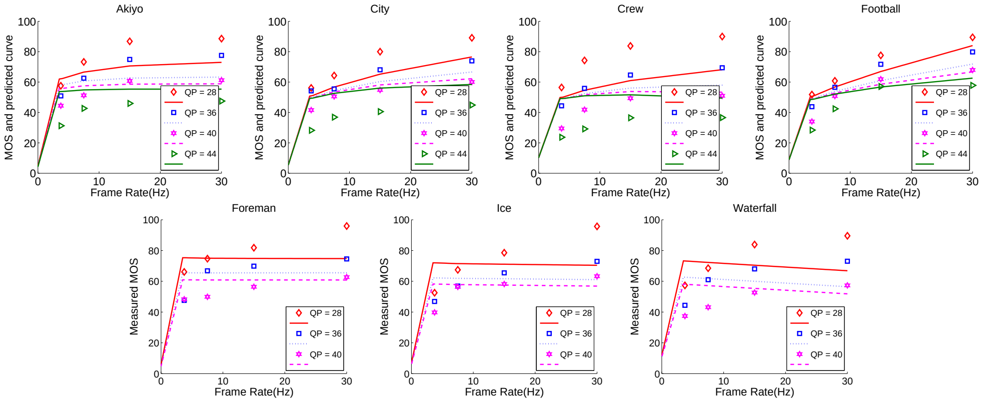
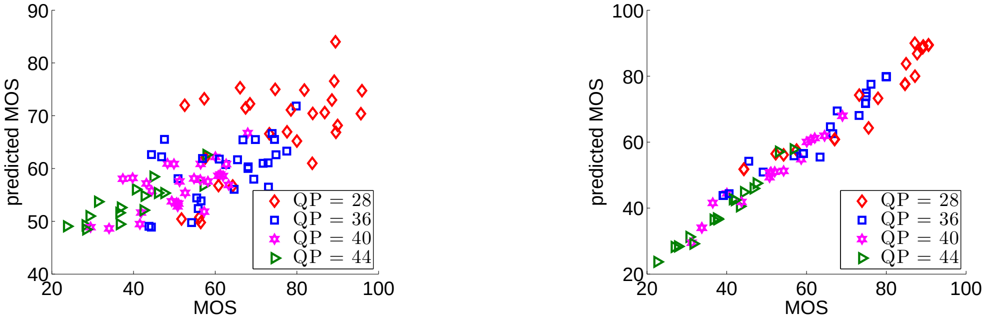


Figure 2.20: Predicted (in curves) vs. measured (in points) MOS for `DataSet#1` by the metric $\overline{\mathrm{QM}}$



(a)

(b)

Figure 2.21: Predicted against measured MOS for `DataSet#1` using (a) $\overline{\mathrm{QM}}$ proposed in [3], (b) VQMTQ.

the TCF part of our model has been shown to be accurate for both CIF and QCIF video (`DataSet#3`). Regarding the spatial quality model using PSNR or QS, the overall quality model can predict better quality when using PSNR than using QS. However, the model accuracy for VQMTQ_3 are still promising with PCC = 0.97 comparing with VQMTQ_2.

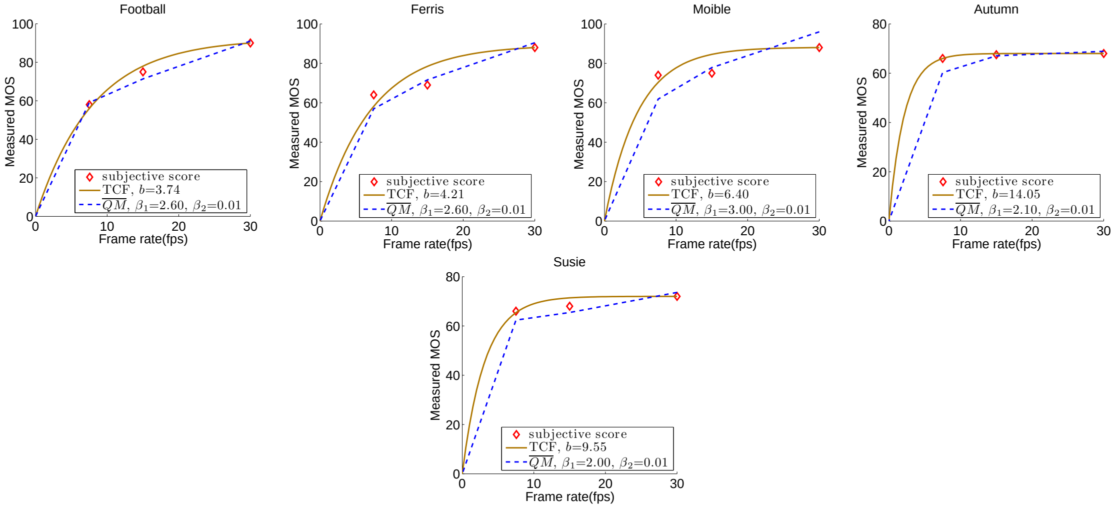


Figure 2.22: Predicted (in curves) vs. measured (in points) MOS for `DataSet#6` against frame rate at $QP = 6$ by the metric $\overline{\mathrm{QM}}$ and TCF

Table 2.4: Pearson Correlation Coefficients of different models

Quality Metrics	DataSet#1	#2	#3	#4	#5	#6	#7
Modeling of the normalized MOS							
Jerkiness [2]	0.97	1	1	0.99	0.99	–	–
NIFVQ [1]	0.97	0.99	0.99	0.99	0.97	–	–
TCF	0.95	0.98	0.98	0.99	0.97	–	–
MNQT	0.97	0.98	0.98	0.97	0.95	–	–
Modeling of MOS, compressed video							
VQMTQ_1	0.98	–	–	–	–	0.92	0.96
VQMTQ_2	0.96	–	–	–	–	–	–
VQMTQ_3	0.97	–	–	–	–	–	–
$\overline{\mathrm{QM}}$ [3]	0.75	–	–	–	–	0.92	0.65

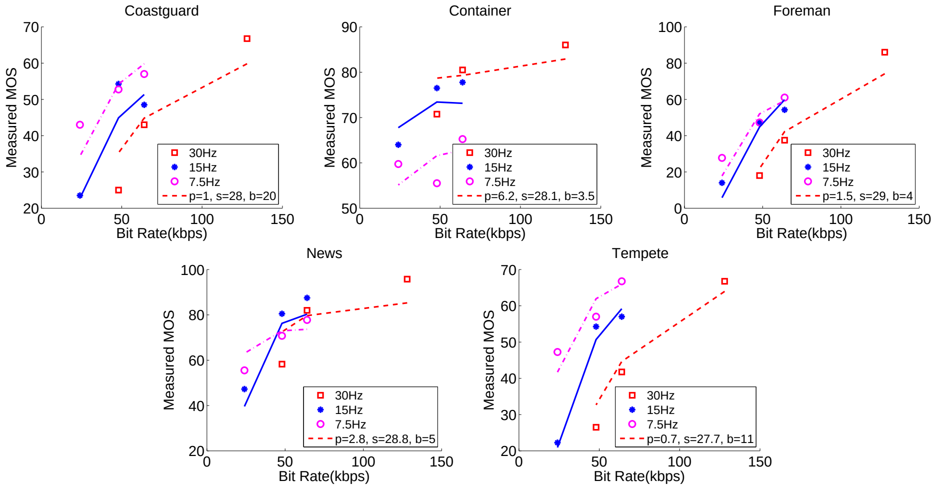
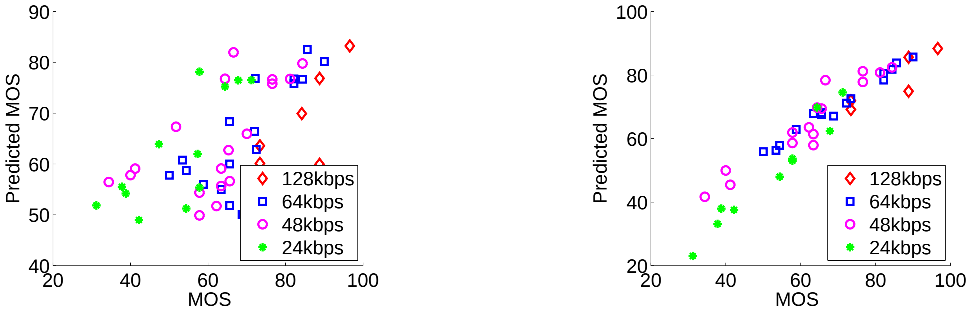


Figure 2.23: Predicted vs. measured MOS for `DataSet#7` against bit rate by the metric VQMTQ (using parameter s, p, and b). The points are measured MOS at different frame rates and the curves are the corresponding predicted MOS



(a)

(b)

Figure 2.24: Predicted against measured MOS for `DataSet#7` using (a) $\overline{\mathrm{QM}}$, and (b) VQMTQ.

Chapter 3

Perceptual Quality Modeling for Mobile Platforms Considering Impact of Spatial, Temporal, and Amplitude Resolutions

In this chapter, we investigate the impact of spatial, temporal and amplitude resolution (STAR) on the perceptual quality of a compressed video. Subjective quality tests were carried out on the TI Zoom2 mobile development platform (MDP). Seven source sequences are included in the tests and for each source sequence we have 32 test configurations generated by JSVM encoder (4 QP levels, 5 spatial resolutions, and 3 temporal resolutions), resulting a total of 224 processed video sequences (PVSs). Videos coded at different spatial resolutions are displayed at the full screen size of the mobile platform. Subjective data reveal that the impact of spatial resolution (SR), temporal resolution (TR) and quantization stepsize (QS) can each be captured by a function with a single content-dependent parameter. The joint impact of SR, TR and QS can be modeled by the product of these three functions with only three parameters. The complete model correlates well with the subjective ratings with a Pearson Correlation Coefficient (PCC) of 0.992. We further found that the TR affects the quality independently of SR and QS, but there is significant interaction between SR and QS. We also investigate the relation between quantization-induced quality impact with respect to bit rate and PSNR. Each of these two quality metrics only requires one content-dependent parameters as well.

The remainder of this chapter is organized as follows: Section 3.1 introduces the quality assessment environment, test methodology and data post-processing. Section 3.2 analyzes the results of subjective tests and present our proposed model. Section 3.3 presents quality modeling of NQQ using PSNR and bit rate. We apply our proposed model on several datasets reported by others in Sec. 3.4 and conclude our work in Sec. 3.5.

3.1 Testing Platform and Methodology

3.1.1 Testing Platform

Targeting for wireless mobile applications, we choose TI's Zoom2 mobile development platform (MDP) [64] as our test platform. This MDP runs on powerful TI OMAP34x processor with a 4.1-inch WVGA (854×480) resolution capacitive multi-touch screen. Google's Android [65] mobile operating system (OS) version 2.1 (Eclair) is used for our test interface development. Our approach for constructing the interface is using Java and XML code to control the high-level program flow, with the help of Android's SDK library to operate low-level video decoding process.

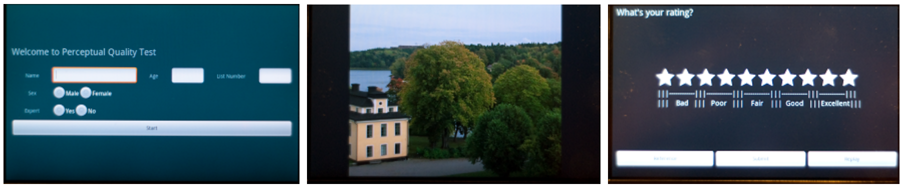


Figure 3.1: Screenshots of the subjective rating interface on TI Zoom2 MDP.

Figure 3.1 illustrates subjective rating interface on our Zoom2 MDP. A *welcome screen* is shown to each viewer at the beginning of each test to record his/her basic information (name, age and gender) and then this is followed by a *playback screen* on which a random 8-second processed video sequence (PVS) is played. Each viewer will be asked to give a score on a *rating screen* after a PVS is played completely. In all tests we allow each subject to replay the current PVS if he/she doesn't feel confident to give a proper judgement, so as to assure more reliable subjective ratings. We adopted a 10-level rating

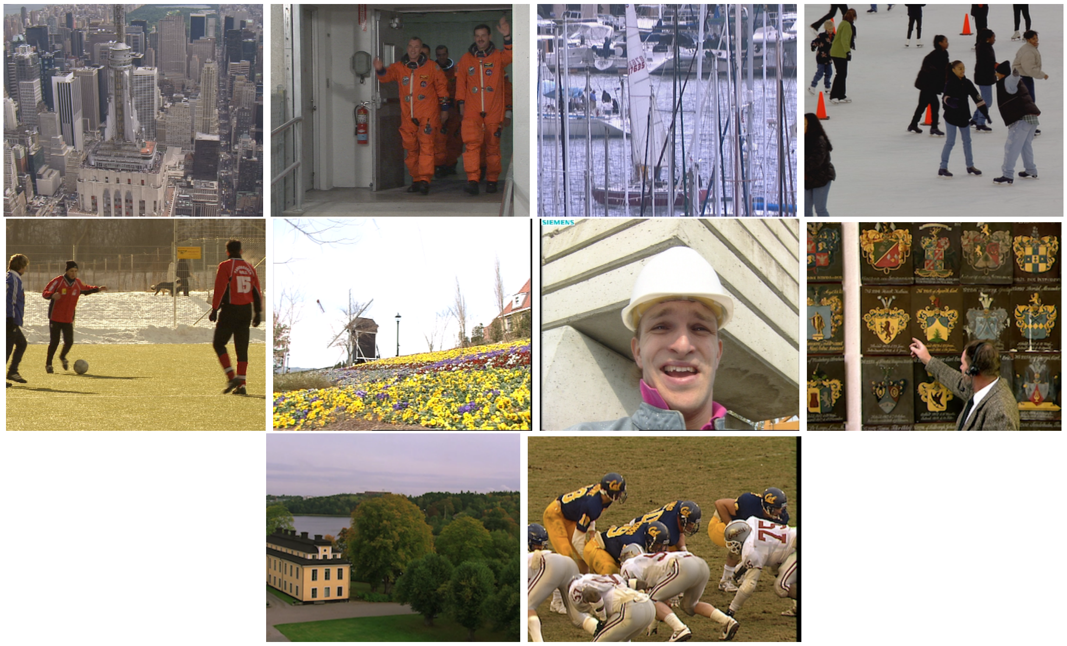


Figure 3.2: Test video pool for subjective tests.

scale as shown in Fig. 3.1 (c). We did not put a level below the scale “1”, which would correspond to a “totally useless video”, since a viewer can still understand the video scene content even from the video at the lowest STAR in our test video pool. So it is reasonable to interpret the effective rating scale as being 11 levels, as recommended by ITU P.910 [17].

3.1.2 Test Video Pool

Seven different videos, i.e., *City*, *Crew*, *Harbour*, *Ice*, *Soccer*, *FlowerGarden* and *Foreman*, five at 4CIF (704×576) and two at VGA (640×480) resolution, are included in our subjective tests. Three additional sequences, i.e., *InToTree*, *Shields* and *Football*, are used as training sequences. The first two are cropped from original 720p high-definition (HD) source to match our Zoom2 MDP display screen size and *Football* is in VGA. These videos are selected from the standard video pool to include various content activities. All the test and training sequences are shown in Fig. 3.2. We plot the spatial information (SI) and temporal information (TI) indices [17] of all source sequences in Fig. 3.3. It demonstrates that the test sequence pool covers a wide range of video contents in terms of motion and spatial details. For the testing consistency through all the PVSs, those VGA videos are cropped and interpolated to 4CIF. According to our pretest [55] performed on the same test platform, it suggests that VGA derived 4CIF versions and original 4CIF versions of the same videos acquire very similar viewer ratings. Low-resolution (i.e., CIF, QCIF) source videos are obtained by downsampling using the Sine-waved Sinc function [66] recommended in the SVC reference software JSVM [60]. Each source video is encoded by JSVM918 [60] using combined spatial and temporal scalabilities, with 3 spatial layers (4CIF, CIF, QCIF) and 3 temporal layers (30, 15, 7.5Hz). Videos corresponding to different QS's are obtained by coding at different QP's. A PVS at a particular STAR is obtained by decoding the spatio-temporal scalable bitstream coded using the desired QP, at the desired SR and TR. Each PVS is about 8 seconds.

For display, each PVS under 4CIF resolution is interpolated to 4CIF using the AVC 6-tap half-pel with bilinear quarter-pel interpolation filter [67]. The test interface will then automatically resize these 4CIF sequences to a spatial resolution with 480 rows. Each PVS is played back in its native frame rate without temporal interpolation. Note that although the screen size of Zoom2 MDP is WVGA, we set our maximum SR to 4CIF, which is the most resolution size over all source sequences.

3.1.3 Test Protocol

Three separate experiments were carried out. Test 1 focuses on the perceptual impact of SR; Test 2 focuses on joint impact of SR and QP; Test 3 focuses on joint effects of STAR. In order to combine subjective scores from these three tests, we include several common sequences between three tests. Common sequences are selected such that they represent a broad quality range in order to facilitate a valid and robust mapping between the tests when combining the datasets. Table 3.1 lists the testing configurations for the three tests. Table 3.2 lists all the common sequences.

Single Stimulus, as recommended by [17] is used for all tests. Before the testing session, a training session, which allows viewers to get familiar with the test, is employed. In Test 2 and 3, we design several subsessions with overlapping sequences, to reduce the viewing time of each subject. Each viewer can participate in one or more subsessions. On

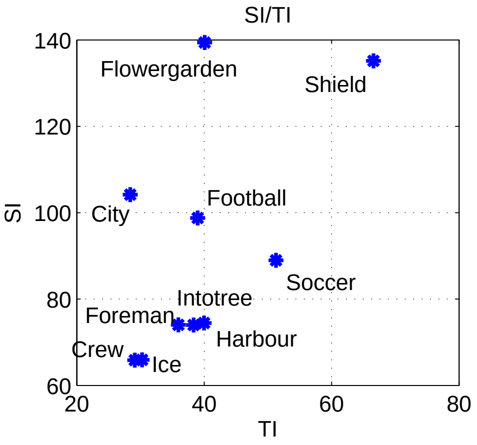


Figure 3.3: The spatial and temporal information indices of the test sequences

Table 3.1: STAR parameters used in tests

	SR	QP	TR
Test 1	176x144 (QCIF), 256x208, 352x288 (CIF), 528x432, 704x576 (4CIF)	22	30
Test 2	QCIF, CIF, 4CIF	22, 28, 36, 44	30
Test 3	QCIF, CIF, 4CIF	22, 28, 36, 44	30, 15, 7.5

Table 3.2: Common Sequences

Test 1&2	*City*@QP22/4CIF/30Hz, *City*@QP22/QCIF/30Hz, *Crew*@QP22/CIF/30Hz, *Harbour*@QP22/QCIF/30Hz, *Ice*@QP22/CIF/30Hz, *Soccer*@QP22/4CIF/30Hz, *Fg*@QP22/QCIF/30Hz, *Foreman*@QP22/CIF/30Hz.
Test 2&3	For all video contents, QP22/4CIF, QP28/4CIF&QCIF, QP36/CIF, QP44/4CIF&QCIF, all at 30Hz.

average, each viewer spends about 18-20 minutes in one viewing session.

3.1.4 Data Processing

Data Collection

We have around 60 evenly distributed male and female viewers participating the tests. Each PVS is rated by 18-20 different viewers. All viewers have normal visual (or after correction) and color perception. About 80% of viewers are non-expert with no related background in video processing. The raw ratings are converted to Z-scores [68] based on the mean and standard deviation of all the scores of each viewer, given by

$$Z_{mij} = \frac{X_{mij} - \mathrm{MEAN(X_i)}}{\mathrm{STD(X_i)}}. \tag{3.1}$$

Here, X_{mij} and Z_{mij} denote the raw rating and Z-score of m^{th} sequence at j^{th} STAR combination, from i^{th} viewer, respectively. X_i denotes all ratings from i^{th} viewer. $\mathrm{MEAN}(\cdot)$ and $\mathrm{STD}(\cdot)$ represent the operator for taking the mean and the standard deviation of a given set, respectively.

Post Screening

Two post screening methods are used in concatenation. We first perform BT.500-11 post screening method [16] in Z-score domain to remove all ratings by certain viewers because their ratings are outside the range of the majority of the viewers. On average, one viewer is eliminated for each PVS. We then conduct the second step to the remaining ratings in the raw score domain, a ratio/averaging method is adopted. We make use of the fact that a video coded at a lower SR or TR but higher QP would not have a rating higher than a video coded at a higher SR or TR but lower QP, if the viewer's judgement is consistent. Therefore, we calculate the ratio of ratings by the same viewer for each pair of PVS's with adjacent SR, QP or TR, respectively. For each source video and each viewer, we count the number of times that the ratio is greater than a threshold ($= 1.1$) for all possible pairs in each dimension, and then we remove all the ratings by a viewer for the same source video if the outlier counter in any dimension is larger than 2. For the remaining pairs of ratings by each viewer, if the ratio is larger than 1, we replace both ratings by their average. After this step, approximately 16-18 ratings remain for each PVS.

Datasets Combining

After the post-processing, we map all the Z-scores from Test 1 and Test 3 to Test 2 using the method recommended in [22]. We map all other tests to Test 2 based on the consideration that only Test 2 has a sufficient number of common sequences with both Test 1 and Test 3.

To map Test 1 data to Test 2, we use a single linear mapping function for all test sequences, because we only have one common PVS for each source sequence. To map Test 3 data to Test 2, since we have many common PVS's for each source video, we form a different linear mapping function for each video.

After combining, we scale the mapped Z-scores back to [0 10] scale, using:

$$\begin{aligned} X_{mij,\mathrm{scl}} = {} & (\mathrm{MEDIAN}(\mathbb{X}^{\mathbb{I}}_{\mathrm{max}}) - \mathrm{MEDIAN}(\mathbb{X}^{\mathbb{I}}_{\mathrm{min}})) \\ & \times \frac{Z_{mij} - Z_{i,\mathrm{min}}}{Z_{i,\mathrm{max}} - Z_{i,\mathrm{min}}} + \mathrm{MEDIAN}(\mathbb{X}^{\mathbb{I}}_{\mathrm{min}})), \end{aligned} \tag{3.2}$$

where $\mathrm{MEDIAN}(\cdot)$ represents the median operator. $\mathbb{X}^{\mathbb{I}}_{\mathrm{max}}$ and $\mathbb{X}^{\mathbb{I}}_{\mathrm{min}}$ are the set of all viewers' maximum and minimum ratings, respectively. $Z_{i,\mathrm{max}}$ and $Z_{i,\mathrm{min}}$ denote the maximum and minimum Z-scores of viewer i. With this scaling, the ratings from all viewers have a common range of $\mathrm{MEDIAN}(\mathbb{X}^{\mathbb{I}}_{\mathrm{min}})$ to $\mathrm{MEDIAN}(\mathbb{X}^{\mathbb{I}}_{\mathrm{max}})$. In our subjective test data, $\mathrm{MEDIAN}(\mathbb{X}^{\mathbb{I}}_{\mathrm{min}})$ =1, and $\mathrm{MEDIAN}(\mathbb{X}^{\mathbb{I}}_{\mathrm{max}})$ =10.

Finally, we average the scaled Z-scores from all viewers for each PVS to obtain its mean opinion score (MOS). The MOS for a sequence with a particular STAR combination, denoted by s, t, q, is indicated by $\mathrm{MOS}(s, t, q)$.

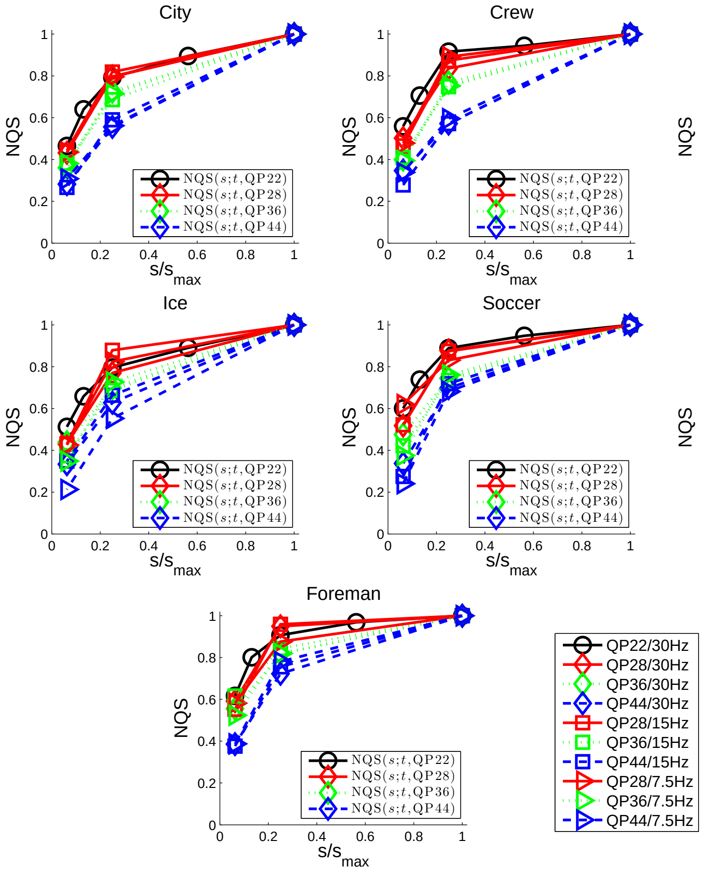


Figure 3.4: Measured NQS under different QS's and TR's. Note that lines with the same color correspond to NQS data at different TR's but the same QP.

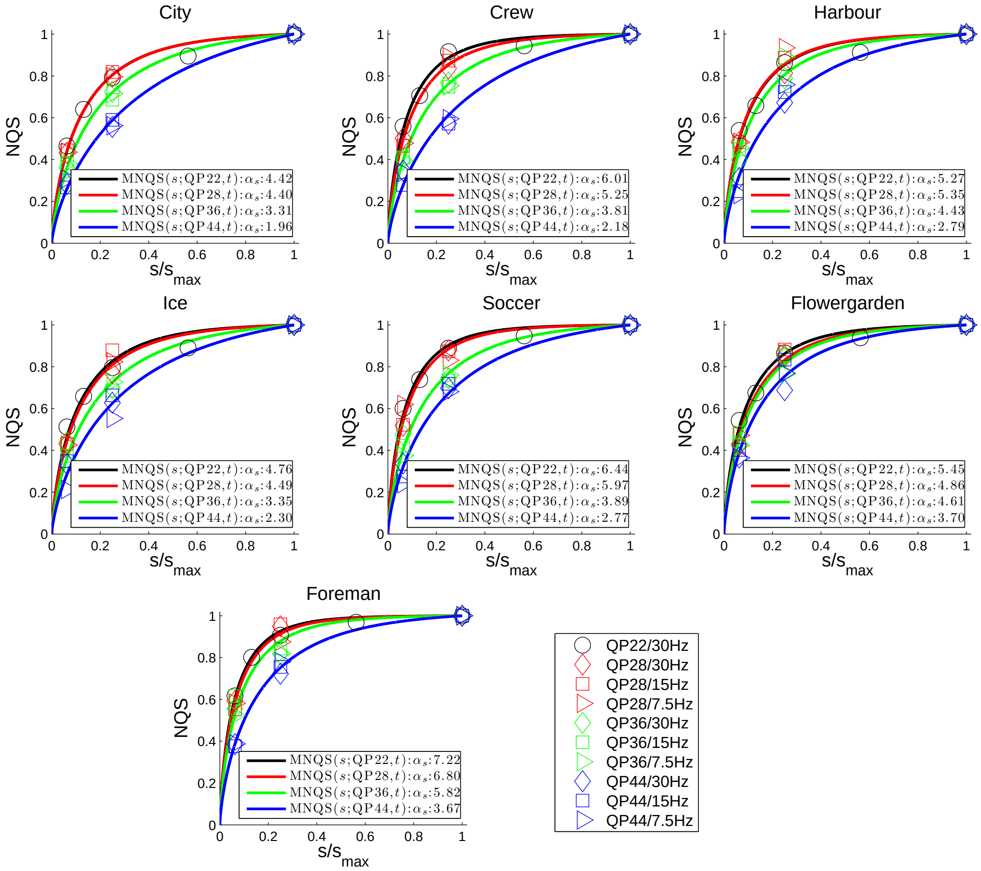


Figure 3.5: Normalized quality v.s. normalized SR. Points are measured data under different QS's and TR's. Curves are derived from the model given in (3.4) with β_s=0.74. The parameter α_s for each sequence and QS is determined by least square fitting of data points at all TR's. PCC=0.992, RMSE=0.03.

3.2 Subjective Test Results and Proposed Quality Model

In order to analyze the test results and derive a quality model reflecting the quality impact of SR, TR, and QS, we first explore how SR, TR or QS individually affects the quality ratings. In each of the following three subsections, we show how MOS varies with one variable (e.g., SR), while holding the other two variables fixed (e.g. TR and QS). Based on the trend observed from the data, we propose a mathematical model that characterizes the degradation of the quality with this variable (e.g. SR). We further examine the interactions of different variables through the two-way Analysis of Variance (ANOVA) [69]. Finally in the last subsection, we propose an overall quality model by taking the product of the three model functions of individual variables, and validate its accuracy.

3.2.1 Modeling Normalized Quality v.s. Spatial Resolution

In this subsection, we examine how SR affects the perceived quality, when TR and QS are fixed. Towards this goal, we plot the normalized quality (NQS) v.s. Normalized SR $s/s_{\max}$ (here, $s_{\max}$ = 4CIF) at the same TR and QS in Fig. 3.4 for each source sequence. The NQS function is defined as

$$\mathrm{NQS}(s;t,q) = \frac{\mathrm{MOS}(s,t,q)}{\mathrm{MOS}(s_{\max},t,q)}. \tag{3.3}$$

From Fig. 3.4, we can observe that the dropping curves of different TR's but same QS tend to cluster together. To examine the dependency of the NQS on TR and QS, respectively, we conduct the three-way ANOVA test for STAR on MOS data. As shown in Tab. 3.3, there are significant differences for each variable of STAR as well as their two way interaction, but no statistical significance for three way interaction. This test only examines the quality differences on MOS data but does not reflect the interaction between the quality dropping trend (e.g., NQS) and the other variable (e.g., TR or QS). Therefore, we further conduct the three-way ANOVA test for the NQS data. By performing the two way interaction between SR and TR/QS, we compute the probability (p-value, which is derived from the cumulative distribution function of F based on the F-value) of the null hypothesis that the differences in NQS as TR (or QS) changing is due to chance. If this probability is low (i.e. p-value $<$ 0.05), we consider TR (or QS) as having statistically significant influence on NQS. If the p-value is much larger, then we say that TR (or QS) has statistically insignificant influence on NQS. The analysis for NQT, NQQ follow the same procedure. As shown in Tab. 3.8, the interaction of SR and TR has a p-value of 0.73 ($>$ 0.05) and a p-value of 0 ($<$ 0.05) for the interaction of SR and QS. This implies that NQS depends on QS but not TR. Therefore, we can approximate the NQS data at the same QS but different TR's with the same model function. By examining the general trend of how NQS changes with normalized SR, we propose the following model for NQS data, called MNQS, i.e.,

$$\mathrm{MNQS}(s;q) = \frac{1 - e^{-\alpha_s(q)(\frac{s}{s_{\max}})^{\beta_s}}}{1 - e^{-\alpha_s(q)}}. \tag{3.4}$$

where $\alpha_s(q)$ is the model parameter, which depends on q but not t. This parameter characterizes the quality decay rate as s decreases, with a smaller value corresponding to a faster dropping rate. We further found that for all 7 source sequences a constant value 0.74 can be used for β_s, so that only a single parameter α_s is content-dependent and QS-dependent. Figure 3.5 shows that this model fits the NQS data at different QP very well. To quantify the accuracy of the fitting, we measure the Pearson Correlation Coefficient (PCC) and root mean square error (RMSE) between the measured and predicted data. For the data presented in Fig. 3.5, PCC= 0.992, RMSE=0.03. The parameter α_s for each QP is obtained by least squares fitting to NQS data at this QP but all different TR's. In addition to ANOVA test, we further examine the model performance when we allow α_s to vary with TR. Table 3.10 shows that this does not lead to significant improvement in PCC and RMSE. This further confirms that by assuming α_s to be independent of TR, we can reduce the model complexity without sacrificing the model accuracy.

In Fig. 3.4 each subplot contains the MNQS curves corresponding to different QS's, for the same video content. We can see that the quality drops faster at larger QS. This is because that larger QS introduces more blurring artifacts compared with smaller QS given the same SR. Note that for most sequences, measured NQS (points) and the MNQS curves at QP=22 and those at QP=28 are indistinguishable, implying that the quantization artifact

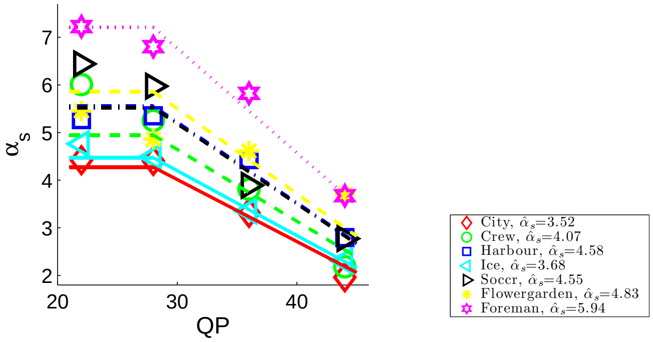


Figure 3.6: The Predicted (in curves) and fitted (in points) α_s v.s. QP for each sequence.

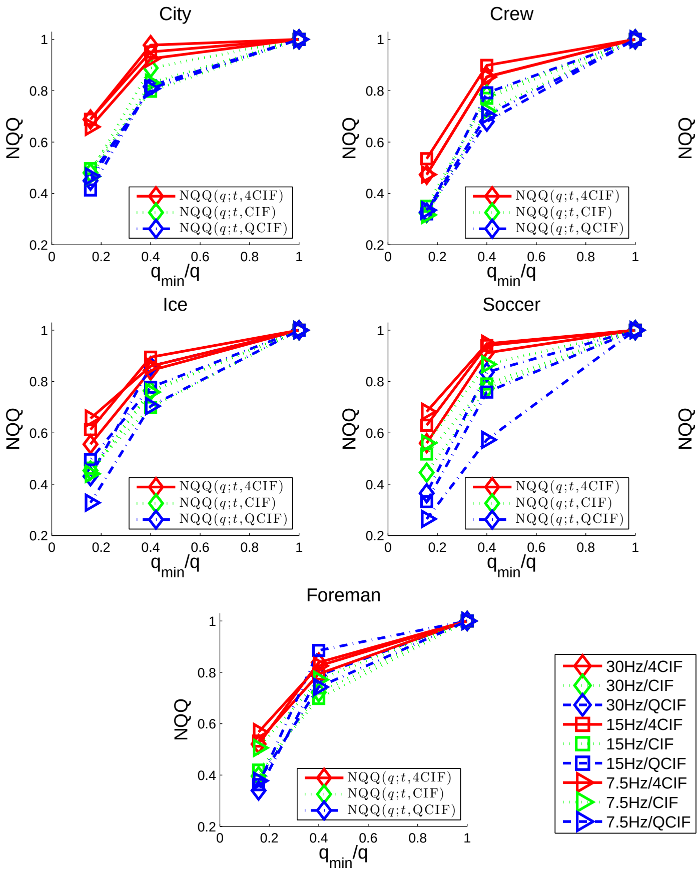


Figure 3.7: Measured NQQ under different SR's and TR's. Note that lines with the same color correspond to NQQ data at different TR's but the same SR.

Table 3.3: Three way ANOVA for STAR.

Factors	F-value	p-value
QS	362.62	0
SR	698.21	0
TR	84.49	0
QS ·SR	5.5	0.0004
QS ·TR	3.68	0.0068
SR ·TR	3.77	0.0059
QS·TR·SR	0.4	0.921

Table 3.4: ANOVA test for statistical significance of the interactions among SR, TR, QS.

	Factors	F-value	p-value
NQS	SR·TR	0.5	0.73
	SR·QS	21.17	0
NQT	TR·SR	1.34	0.25
	TR·QS	1.08	0.37
NQQ	QS·SR	0.21	0
	QS·TR	0.34	0.85

induced by the H.264 encoder becomes almost invisible at QP=28, and smaller QP does not lead to noticeable improvement. With other encoders, the saturation point may differ.

To further simplify the model, we investigate the relationship between α_s and q. Figure 3.6 shows that α_s has an approximately linear relationship with QP, for QP >=28, and the α_s for QP=22 is very close to that for QP=28. Therefore, we propose to model the dependency of α_s on q (and equivalently on QP) by

$$\alpha_s(q) = \hat{\alpha}_s L(\mathrm{QP}(q)),$$
$$\text{with } L(\mathrm{QP}) = \begin{cases} \upsilon_1\mathrm{QP} + \upsilon_2, & \text{if QP} >= 28 \\ 28\upsilon_1 + \upsilon_2, & \text{if QP} < 28, \end{cases} \tag{3.5}$$

where QP is related to q with $\mathrm{QP}(q) = 4{+}6\log_2 q$, as defined by the H.264/SVC codec [70]. We derive the constants υ_1, υ_2, and β_s (which are sequence independent) together with the model parameter $\hat{\alpha}_s$ (sequence dependent) by minimizing the mean squares error between the measured NQS data at all STAR combinations and the predicted NQS using (3.4) and (3.5). The best fitting constants are $\upsilon_1 = -0.037$, $\upsilon_2 = 2.25$, and $\beta_s = 0.74$. Figure 3.6 shows that the α_s determined using (3.5) are quite close to the original α_s, except for a few cases (e.g. *Flowergarden* and *Soccer*). Even in those cases, the differences in α_s values do not have a significant impact on the resulting MNQS curves. The MNQS curves obtained using (3.4) and (3.5) with only a single content-dependent parameter $\hat{\alpha}_s$ are very similar to those shown previously in Fig. 3.5, and hence are not included to save the space. The predicted NQS has PCC=0.989, RMSE=0.0352, only slightly worse than those using (3.4) with independently determined parameter α_s for each QP. Therefore, we propose to use (3.5) together with (3.4) to model NQS, which needs only one parameter $\hat{\alpha}_s$ across different QP levels.

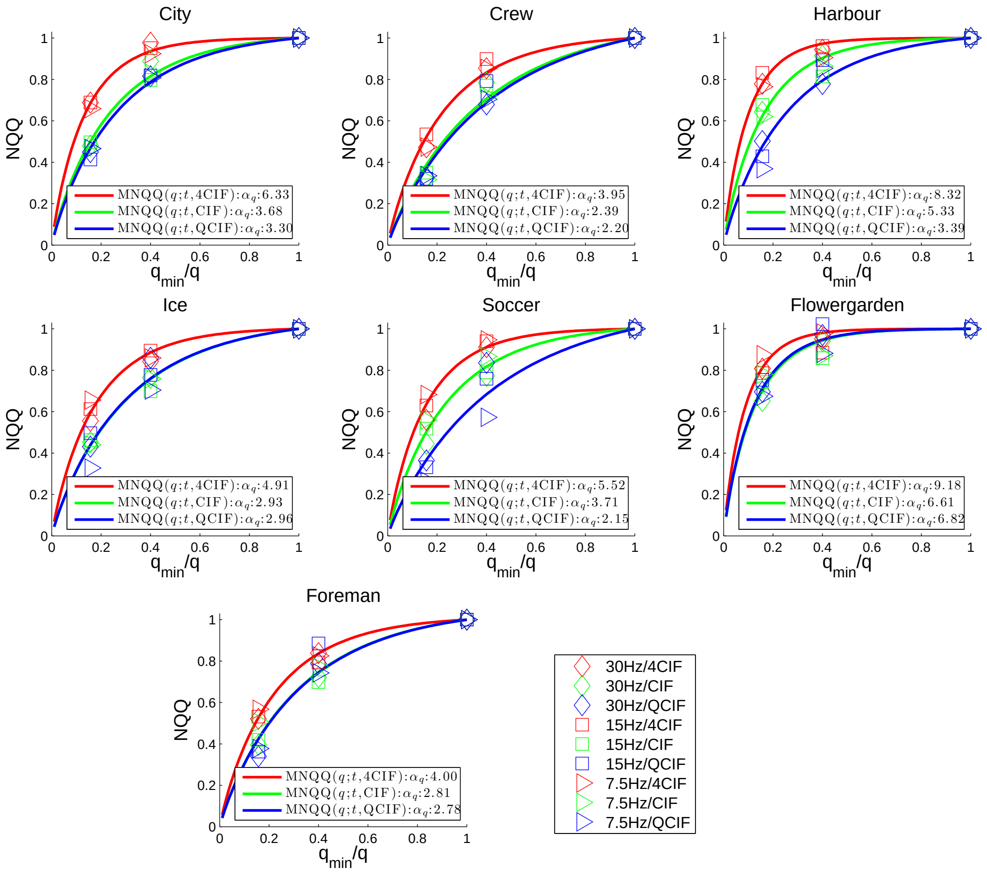


Figure 3.8: Normalized quality v.s. normalized QS. Points are measured data under different SR's and TR's. Curves are derived from the model in (3.7) for all t at a given s. The parameter α_q for each sequences and SR is determined by least squares fitting of data at all TR's. PCC=0.982, RMSE=0.041.

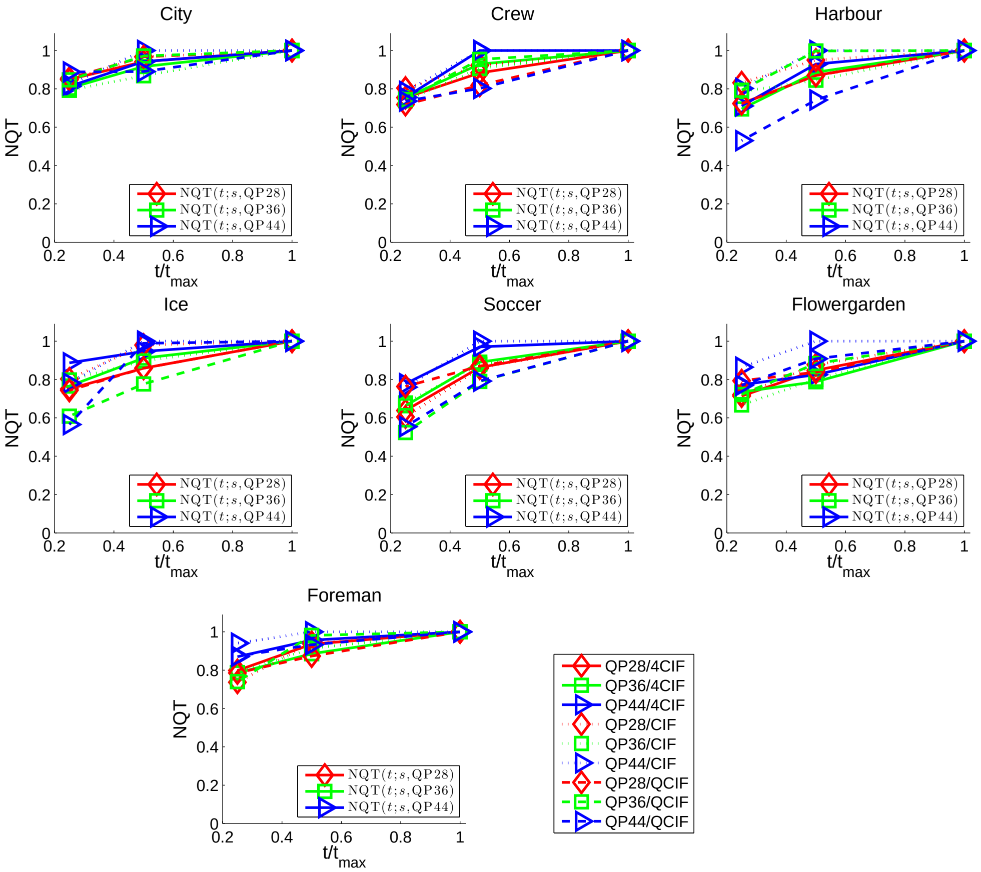


Figure 3.9: Measured NQT under different QS's and SR's. Note that lines with the same color correspond to NQT data for test sequences at different SR's (4CIF, CIF, QCIF), but the same QP.

3.2.2 Modeling Normalized Quality v.s. Quantizations

In this subsection, we explore how QS affects the perceived quality when SR and TR are fixed. Towards this goal, we plot the normalized quality v.s. inverse normalized QS $q_{\min}/q$ (NQQ) at same SR and TR in Fig. 3.7. Note that $q_{\min}/q$ can be considered normalized amplitude resolution. The NQQ is defined as

$$\mathrm{NQQ}(q; s, t) = \frac{\mathrm{MOS}(s, t, q)}{\mathrm{MOS}(s, t, q_{\min})}, \tag{3.6}$$

where $q_{\min}$ is the minimum QS ($q_{\min}$ = 16 in our study). In Fig. 3.7, we can observe that the dropping trends of NQQ for different TR's but the same SR tend to cluster together. We performed two-way ANOVA to examine the statistical significance of the interaction between QS and SR and that between QS and TR. As shown in Tab. 3.8, the interaction of QS and TR has a p-value of 0.85 ($>$ 0.05) and the interaction of QS and SR has a p-value of 0 ($<$ 0.05). This reveals that the NQQ depends on SR but not TR. By examining the general trend of how NQQ changes with normalized QS at the same SR, we propose a model for NQQ data, called MNQQ, with a function form of,

$$\mathrm{MNQQ}(q; s) = \frac{1 - e^{-\alpha_q(s)(\frac{q_{\min}}{q})^{\beta_q}}}{1 - e^{-\alpha_q(s)}}, \tag{3.7}$$

where α_q is the model parameter. This parameter characterizes the quality decay rate as q increases, with a smaller value corresponding to a slower dropping rate. Based on the previous analysis, we assume α_q depends on s but not t. We derive α_q for each SR for a test sequence by least squares fitting using measured NQQ data for that SR, at all TR's. Similar to β_s in Eq. 3.4, we found that for all 7 source sequences a constant value 1 can be used for β_q, so that only a single parameter α_q is content-dependent and QS-dependent. Figure 3.8 shows that the MNQQ model is very accurate. To further validate the assumption that NQQ is independent of TR, we also evaluate the model accuracy when the parameter α_q is allowed to vary with TR. Table 3.10 shows that allowing α_q to vary with TR does not improve the model accuracy significantly. Both this comparison and ANOVA study suggest that we can use the same model parameter for different TR's in MNQQ to reduce model complexity.

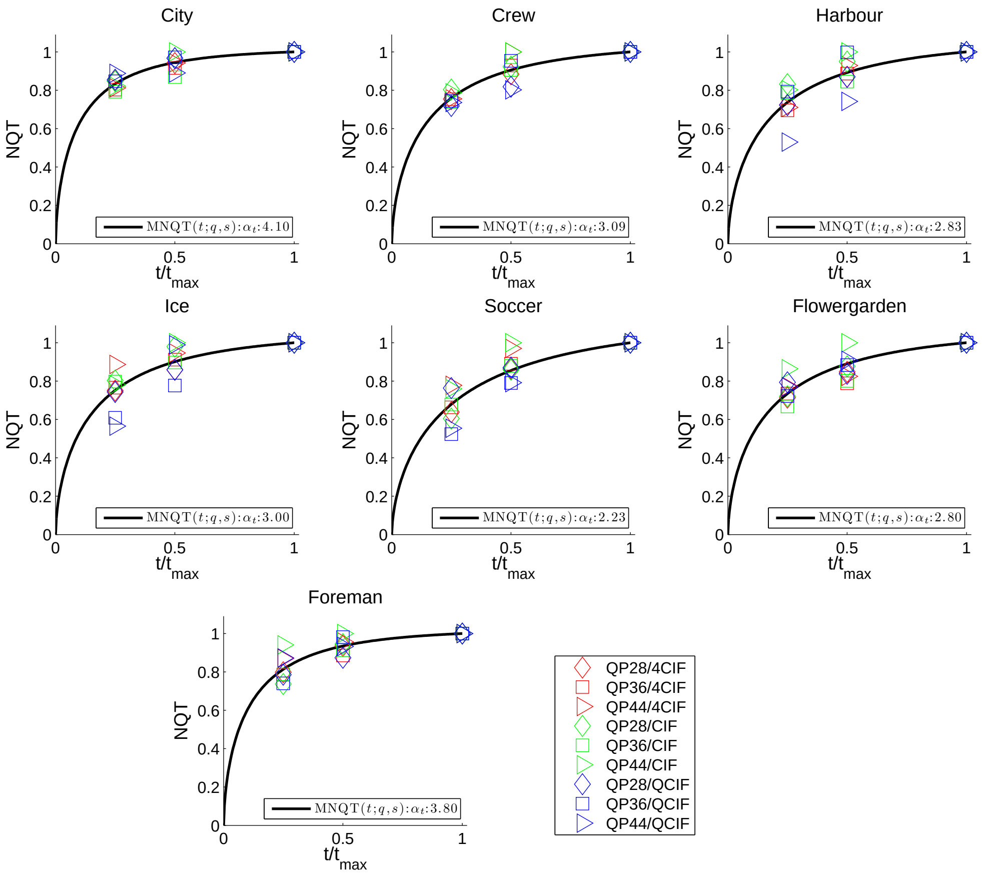


Figure 3.10: Normalized quality v.s. normalized TR. Points are measured data under different QS's and SR's. Curves are derived from the model given in (3.9) with β_t=0.63. The model parameter α_t is determined by least squares fitting of data at all SR's and QS's. PCC=0.891, RMSE=0.052.

Table 3.5: Model accuracy under different assumptions.

Model	Assumptions	PCC	RMSE
MNQS	α_s depends on TR	0.995	0.025
	α_s independent of TR	0.992	0.030
MNQT	α_t depends on SR and QS	0.972	0.026
	α_t independent of SR and QS	0.891	0.052
MNQQ	α_q depends on TR	0.995	0.020
	α_q independent of TR	0.982	0.041

3.2.3 Modeling Normalized Quality v.s. Temporal Resolution

In this subsection, we explore how TR affects perceived quality when SR and QS are fixed. Towards this goal, we plot the normalized quality v.s. normalized TR $t/t_{\max}$ (NQT) at same SR and QS in Fig. 3.9. The NQT is defined as

$$\mathrm{NQT}(t; s, q) = \frac{\text{MOS}(s, t, q)}{\text{MOS}(s, t_{\max}, q)}, \tag{3.8}$$

where $t_{\max}$ is the maximum TR (here, $t_{\max}$ = 30Hz). From Fig. 3.9, we can observe that the dropping trends of NQT for different SR's and QS's tend to cluster together. To quantify the statistical significance of the dependency of NQT with SR and QS, we perform two-way ANOVA between TR and SR, and between TR and QS. As shown in Tab. 3.8, the interaction of TR and SR has p-value of 0.25 ($>$ 0.05) and the interaction of TR and QS has p-value of 0.37 ($>$ 0.05). This shows that the NQT neither depends on SR nor QS. By examining the general trend of how NQT changes with normalized TR, we propose a model for NQT data, called MNQT, with a function form of,

$$\mathrm{MNQT}(t) = \frac{1 - e^{-\alpha_t (\frac{t}{t_{\max}})^{\beta_t}}}{1 - e^{-\alpha_t}}. \tag{3.9}$$

The parameter α_t controls how fast the NQT drops as t decreases, with a smaller value corresponding to a faster dropping rate. Based on the previous analysis, we assume α_t is independent of both SR and QS, and derive its value for each test sequence by least squares fitting using measured NQT data at all SR's and QS's. Similar to β_s in (3.4), β_t is a constant of 0.63 for all seven sequences, which is found by least square fitting for all NQT data. Figure 3.10 shows that the model curves can capture the data trends well with PCC=0.891, RMSE=0.052. We also compute the PCC and RMSE of the model when using a best fitting α_t for each different pair of SR and QS. Table 3.10 (middle two rows) shows that this brings slight improvement in terms of PCC and RMSE. However, considering that we already achieve high PCC and low RMSE with a parameter that is independent of both SR and QS, we choose to use this option to reduce the model complexity.

3.2.4 The Overall Q-STAR Model

To derive the overall quality model as a function of s, t, q, we recognize that the normalized MOS can be decomposed in any of the following ways:

$$\begin{aligned}
&\frac{\mathrm{MOS}(s,t,q)}{\mathrm{MOS}(s_{\max},t_{\max},q_{\min})} \\
&= \mathrm{NQS}(s;t_{\max},q_{\min})\mathrm{NQT}(t;s,q_{\min})\mathrm{NQQ}(q;s,t) && \text{(3.10a)}\\
&= \mathrm{NQS}(s;t_{\max},q_{\min})\mathrm{NQQ}(q;s,t_{\max})\mathrm{NQT}(t;s,q) && \text{(3.10b)}\\
&= \mathrm{NQT}(t;s_{\max},q_{\min})\mathrm{NQS}(s;t,q_{\min})\mathrm{NQQ}(q;s,t) && \text{(3.10c)}\\
&= \mathrm{NQT}(t;s_{\max},q_{\min})\mathrm{NQQ}(q;s_{\max},t)\mathrm{NQS}(s;t,q) && \text{(3.10d)}\\
&= \mathrm{NQQ}(q;s_{\max},t_{\max})\mathrm{NQS}(s;t_{\max},q)\mathrm{NQT}(t;s,q) && \text{(3.10e)}\\
&= \mathrm{NQQ}(q;s_{\max},t_{\max})\mathrm{NQT}(t;s_{\max},q)\mathrm{NQS}(s;t,q). && \text{(3.10f)}
\end{aligned}$$

Among these decomposition orders, we choose the one that will require the least number of model parameters while maintaining high accuracy. Because NQT term is independent of both SR and QS, and the NQS and NQQ terms are both independent of TR, we could put NQT at any place, and it will only require a single parameter α_t, and it will not affect the number of parameters needed for NQS and NQQ. Between NQS and NQQ, if we choose to put NQQ term after the NQS term, we would need to estimate α_q for each s. This is because the NQQ parameter α_q depends on s, and we don't have a good model that relates α_q with s. On the other hand, if we put the NQS term after the NQQ term, we only need to estimate α_q for s=$s_{\max}$, and because of (3.5), we only need to estimate $\hat{\alpha}_s$ to obtain α_s for all q. Based on these considerations, we could use (3.10d), (3.10e) or (3.10f) to reduce the model parameters while maintain high model performance.

By replacing NQS, NQQ, NQT in (3.10e) with their models described in (3.4), (3.7) and (3.9), respectively, the proposed overall quality model as a function of s, t, q, to be

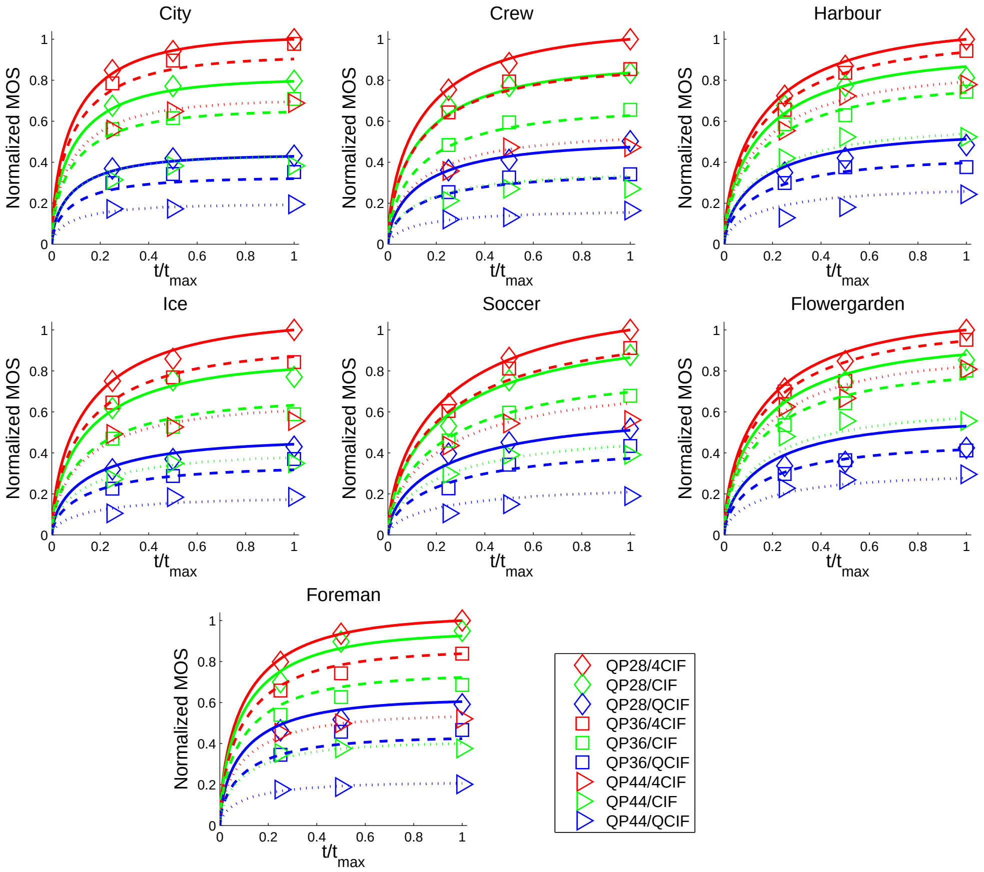


Figure 3.11: Predicted normalized quality (in curves) and measured normalized MOS (in points) v.s. $t/t_{\max}$ under different SR's and QS's. The model parameters α_q, $\hat{\alpha}_s$ and α_t, are obtained by least square fitting, given in Table 5.5.

called QSTAR, can be written as,

$$\begin{aligned} \mathrm{QSTAR}(s,t,q) &= \mathrm{MNQQ}(q;s_{\max})\mathrm{MNQS}(s;q)\mathrm{MNQT}(t) \\ &= \frac{1-e^{-\alpha_q(\frac{q_{\min}}{q})}}{1-e^{-\alpha_q}} \frac{1-e^{-\hat{\alpha}_s L((\mathrm{QP}(q))(\frac{s}{s_{\max}})^{\beta_s}}}{1-e^{-\hat{\alpha}_s L(\mathrm{QP}(q))}} \frac{1-e^{-\alpha_t(\frac{t}{t_{\max}})^{\beta_t}}}{1-e^{-\alpha_t}}, \end{aligned} \tag{3.11}$$

where $\beta_s = 0.74$, $\beta_t = 0.63$ and $L((\mathrm{QP}(q))$ is defined in (3.5), with υ_1=−0.037, υ_2=2.25. The model has three content-dependent parameters α_q, $\hat{\alpha}_s$ and α_t. We compare the predicted quality using this model with measured MOS in Fig. 3.11, where the model parameters are obtained by least squares fitting into the data of NQQ, NQS, and NQT, respectively. As can be seen that the model matches very well with measured MOS points for most cases. Table 5.5 (upper portion) summarizes the model parameters, RMSE and PCC of the proposed QSTAR model. In this table we also list the average 95% confidence interval (CI) of user ratings (normalized by maximum possible rating of each source sequence) for each source sequence. We see that the RMSE of the prediction error is much lower than the CI for all sequences. The correlation scatter plot between predicted and measured quality is presented in Fig. 3.24 (left part). Note that we also investigated using other functional forms, such as power law in the form of $(\frac{s}{s_{\max}})^{\alpha}$ and logarithm functions with a form of $(\frac{\log(1+\alpha(\frac{s}{s_{\max}}))}{\log(1+\alpha)})$, to model NQS and NQT data. We have found that our proposed models in (3.4) and (3.9) can more accurately capture the impact of SR and TR on the perceived quality, at least for our dataset.

3.2.5 Comparison of Subjective Test Results obtained using Laptop and Mobil Platforms

We also applied the NQQ model in (3.7) and the NQT mode in (3.9) to the subjective test results for laptop screens presented in Chapter 2. We used the same β_q and β_t derived here and summarize the dropping rate of NQQ and NQT data for both laptop and mobile devices in Tab. 3.6. We would like to observe whether there is any quality impact of display screen size. Table 3.6 demonstrates the parameter values on four common source sequences tested on both experiments. We can see that in most cases, they are similar, although somehow α_q is slightly smaller (i.e. a faster dropping rate) for smaller screen size. This is intuitively reasonable; because viewers are more sensitive to blur effects caused by quantization artifacts when display screen size of the devices is smaller. Or because when both are in CIF resolution, it blows up to 4CIF resolution for mobile devices and viewers are more sensitive to blur effect of interpolation filter as QS increases. On the other hand, the differences in α_t are inconsistent, indicating no consistent trend as to which derives make the viewer more sensitive to temporal resolution.

Table 3.6: The parameters comparison between datasets

dataset		city	crew	foreman	ice
DataSet#9 (laptop)	α_q (CIF)	5.09	3.23	4.32	5.04
	α_t	3.67	3.65	4.05	3.25
DataSet#1 (mobile)	α_q (4CIF)	7.25	4.51	4.57	5.62
	α_q (CIF)	4.21	2.79	3.25	3.38
	α_t	4.1	3.09	3.8	3

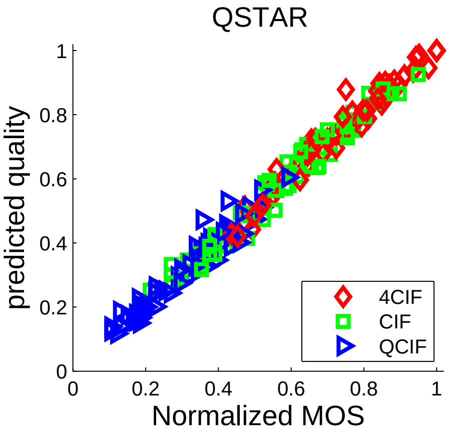


Figure 3.12: Predicted quality using QSTAR model against measured MOS, where the model parameters obtained by least square fitting

3.3 Alternative QSTAR Model using PSNR and Bit Rate

As proposed from our previous Sec. 3.2, we derive a quality model reflecting the quality impact of SR, TR and QS. We first individually explore how SR, TR, or QS with

Table 3.7: The parameters and performance of QSTAR model.

	city	crew	harbour	ice	soccer	fg	foreman	Avg.
	Parameters obtained by least square fitting with MOS data							
α_q	7.25	4.51	9.65	5.61	6.31	10.68	4.57	
$\hat{\alpha_s}$	3.52	4.07	4.58	3.68	4.55	4.83	5.94	
α_t	4.10	3.09	2.83	3.00	2.23	2.80	3.80	
RMSE	0.018	0.025	0.038	0.033	0.032	0.058	0.038	0.035
PCC	0.998	0.996	0.992	0.993	0.992	0.979	0.991	0.991
avg. CI	0.048	0.049	0.050	0.050	0.050	0.051	0.049	0.050

Table 3.8: ANOVA test for statistical significance of the interactions among SR, TR, QS.

	Factors	F-value	p-value
NQS	SR·TR	0.5	0.73
	SR·QS	21.17	0
NQT	TR·SR	1.34	0.25
	TR·QS	1.08	0.37
NQQ	QS·SR	0.21	0
	QS·TR	0.34	0.85

the quality ratings. Then, the overall quality model is presented as a function of SR, TR and QS, which is the product of three one-parameter models, and each of which characterizes the degradation of quality with these variables (e.g., SR). However, this quality model will not be flexible and applicable if the retrievable information of the target video sequences are limited. For example, without acknowledgement of the QS (or QP), we only can access the PSNR or the bit rate of each PVS, our QSTAR model in Sec. 3.2 cannot be applied properly. Therefore, it would be more robust to replace the MNQQ model with QS to the one with PSNR or bit rate, if one of them is accessible. In the following subsection, we will introduce the variant of the MNQQ model form as a function of PSNR and bit rate and investigate the performance of individual model as well as the joint form in QSTAR.

3.3.1 Modeling Normalized Quality v.s. PSNR

In this subsection, we explore how PSNR affects the perceived quality when SR and TR are fixed. Towards this goal, we plot the normalized quality v.s. normalized PSNR

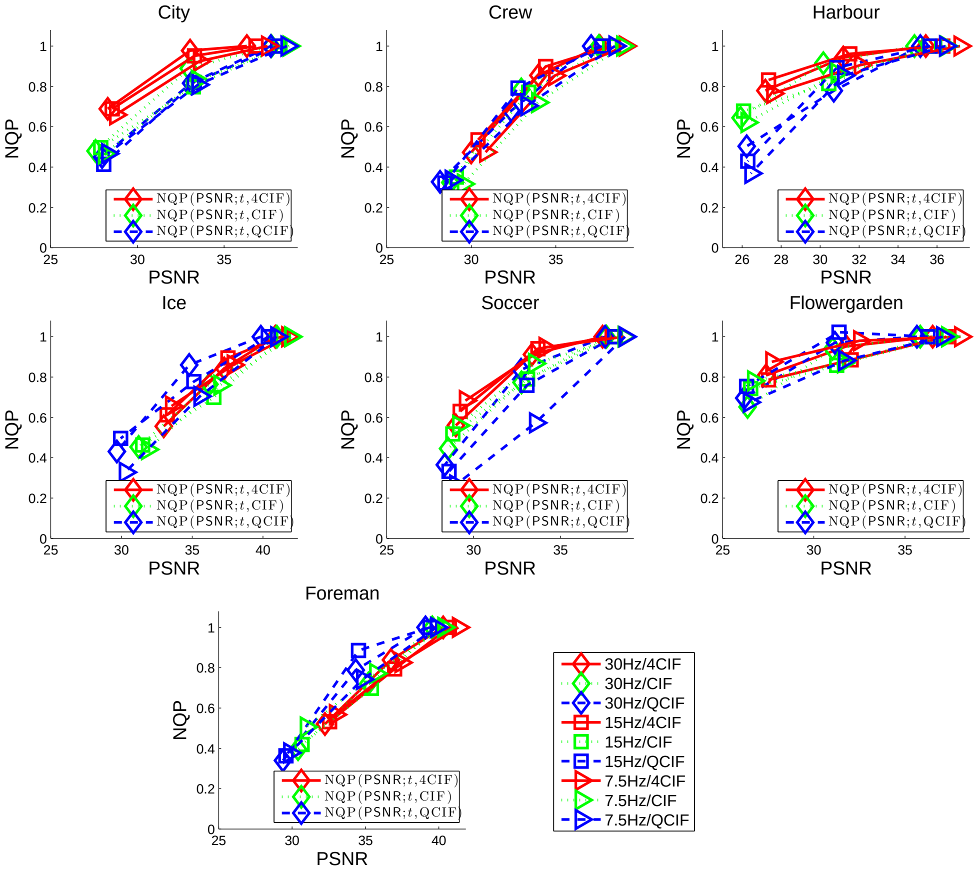


Figure 3.13: Measured NQP under different SR's and TR's. Note that lines with the same color correspond to NQP data at different TR's but the same SR.

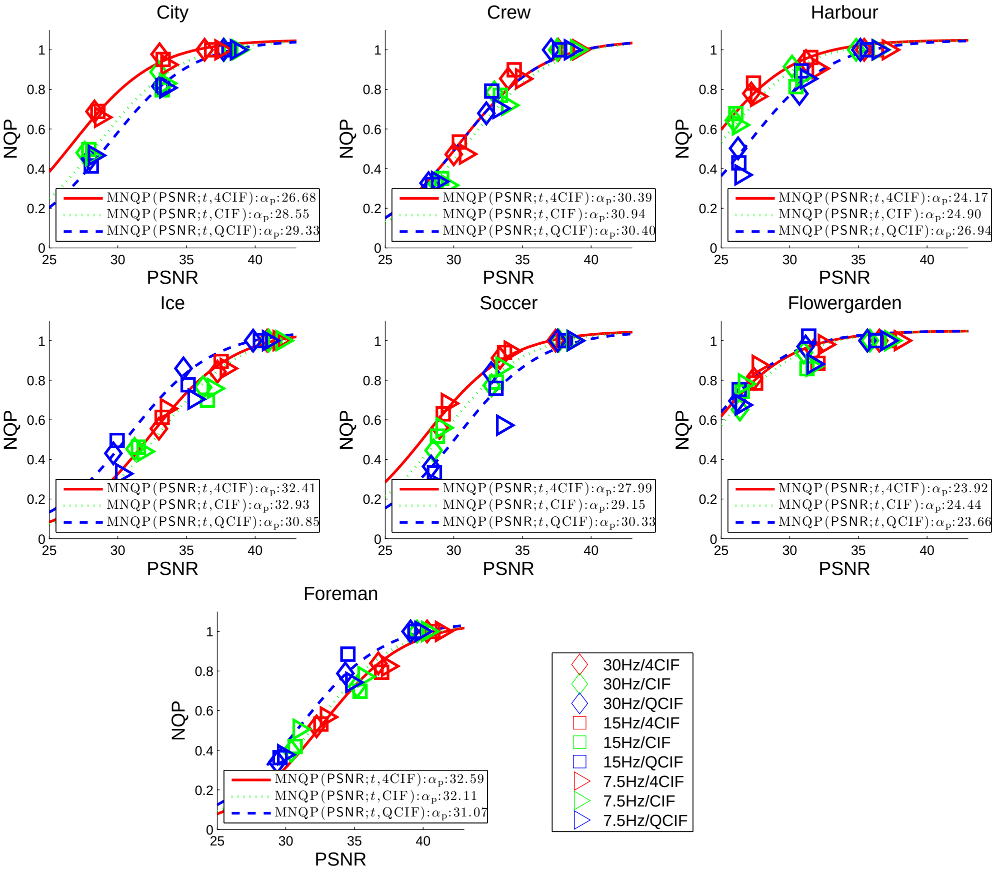


Figure 3.14: Normalized quality v.s. PSNR. Points are measured data under different SR's and TR's. Curves are derived from the model in (2.7) for all TR's at a given SR. The parameter α_q for each sequences and SR is determined by least squares fitting of data at all TR's. PCC=0.985, RMSE=0.037.

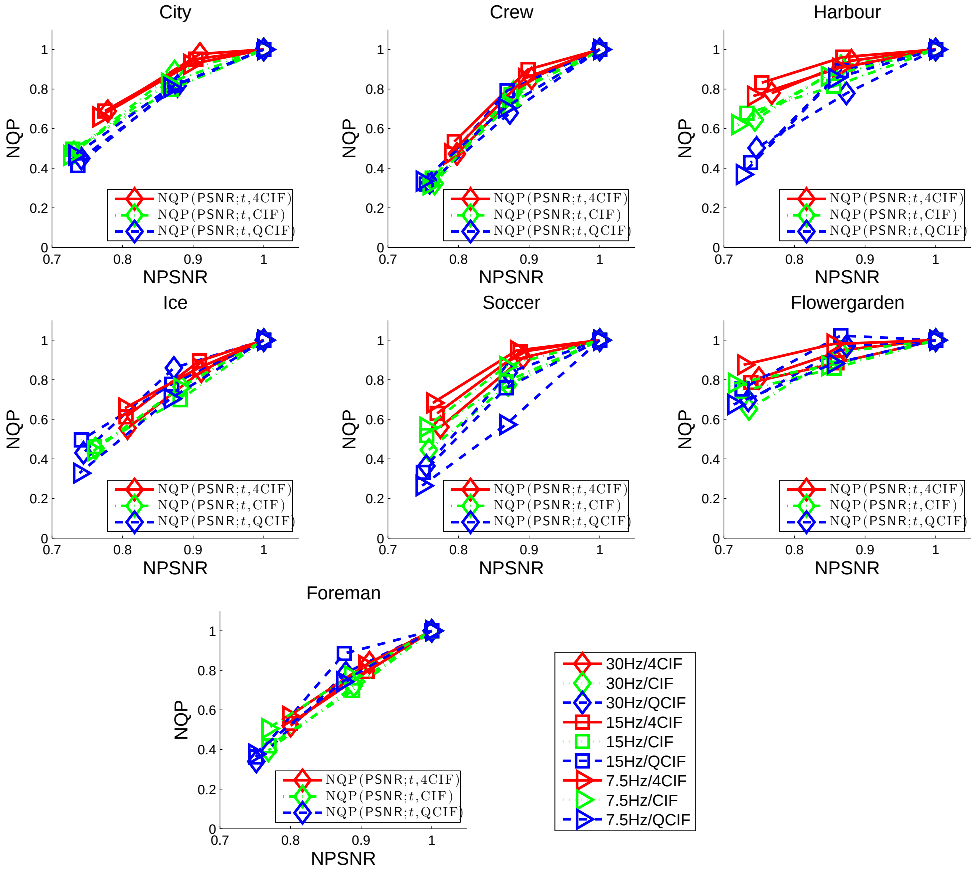


Figure 3.15: Measured NQP v.s. normalized PSNR under different SR's and TR's. Note that lines with the same color correspond to NQP data at different TR's but the same SR.

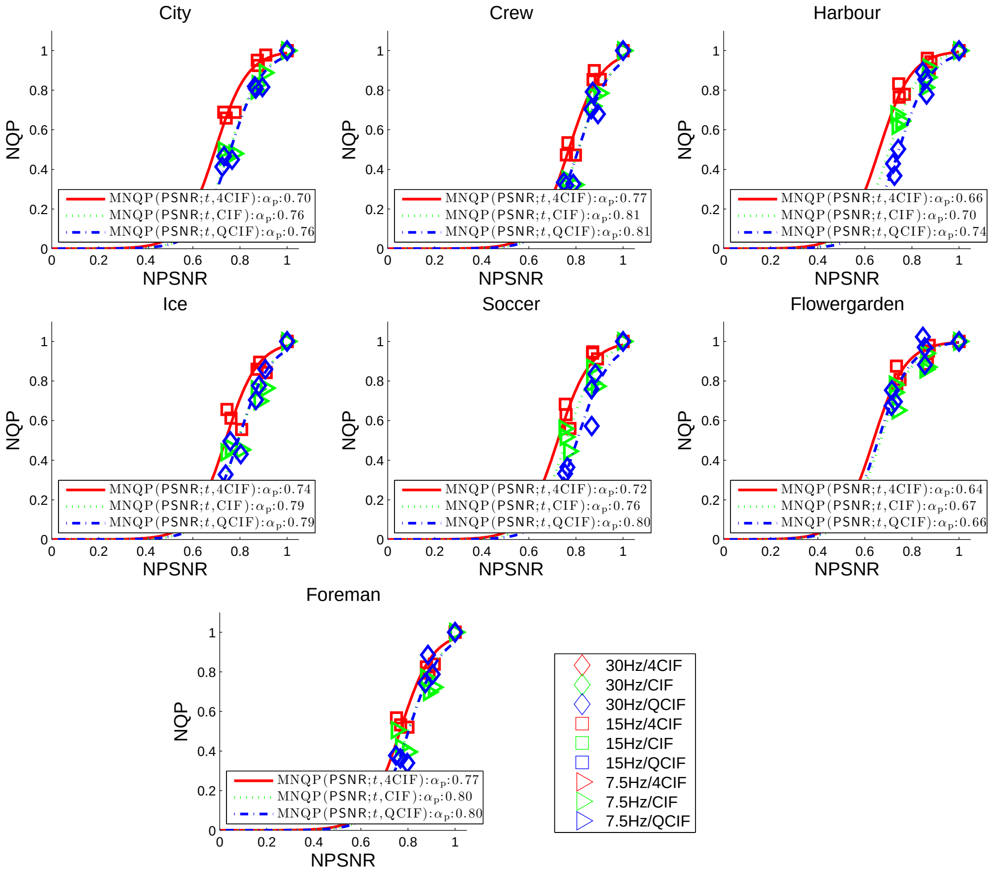


Figure 3.16: Normalized quality v.s. NPSNR. Points are measured data under different SR's and TR's. Curves are derived from the model in (3.13) for all t at a given s. The β_p : $(1, 14.6)$ parameter α_q for each sequences and SR is determined by least squares fitting of data at all TR's. PCC=0.964, RMSE=0.058.

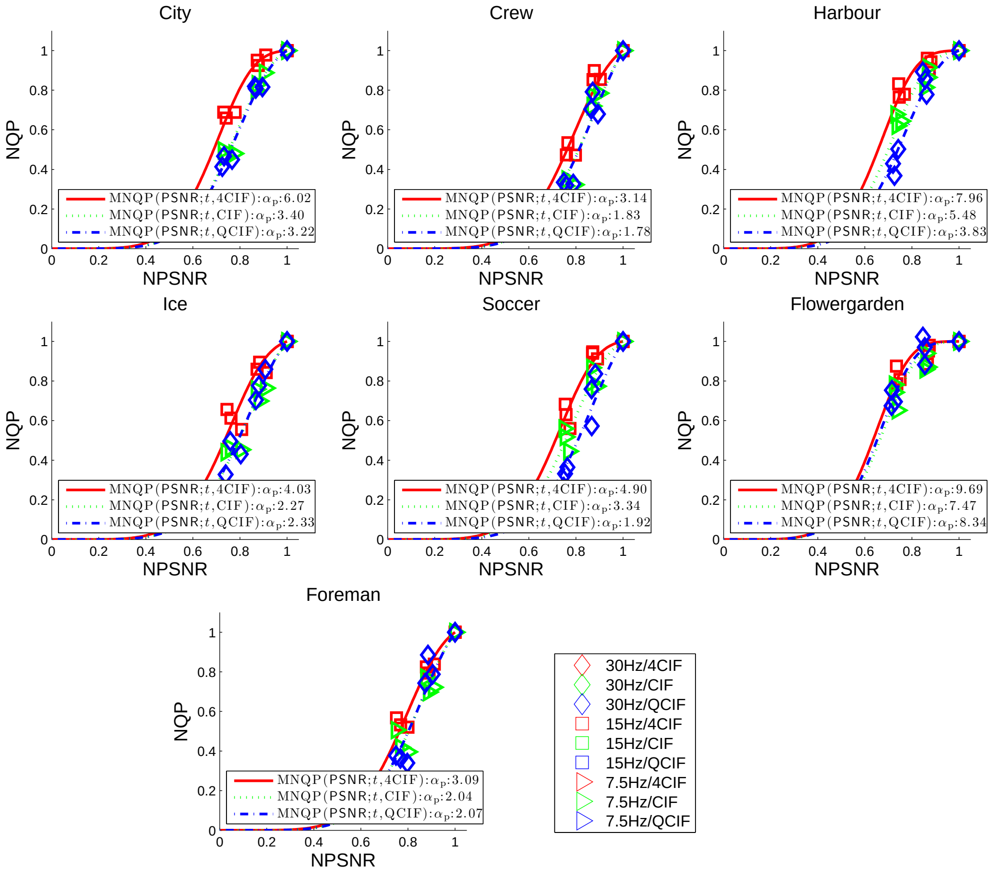


Figure 3.17: Normalized quality v.s. normalized PSNR. Points are measured data under different SR's and TR's. Curves are derived from the model in (3.14) for all t at a given s. The parameter α_p and $\beta_p = 5.8$ for each sequences and SR is determined by least squares fitting of data at all TR's. PCC=0.970, RMSE=0.053.

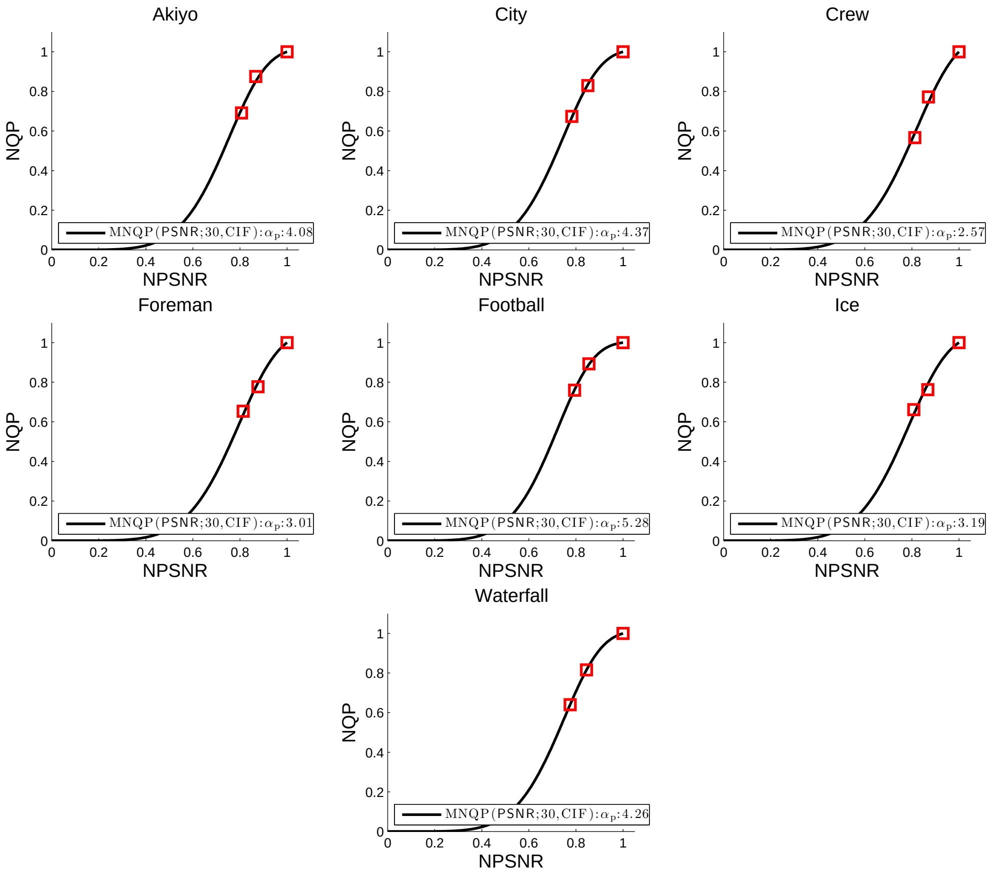


Figure 3.18: Normalized quality v.s. NPSNR. Points are measured data in CIF at 30Hz (CSVT data). Curves are derived from the model in (3.14) for all t at CIF. The parameter α_p for each sequences and SR is determined by least squares fitting of data at all TR's. PCC=0.995, RMSE=0.014.

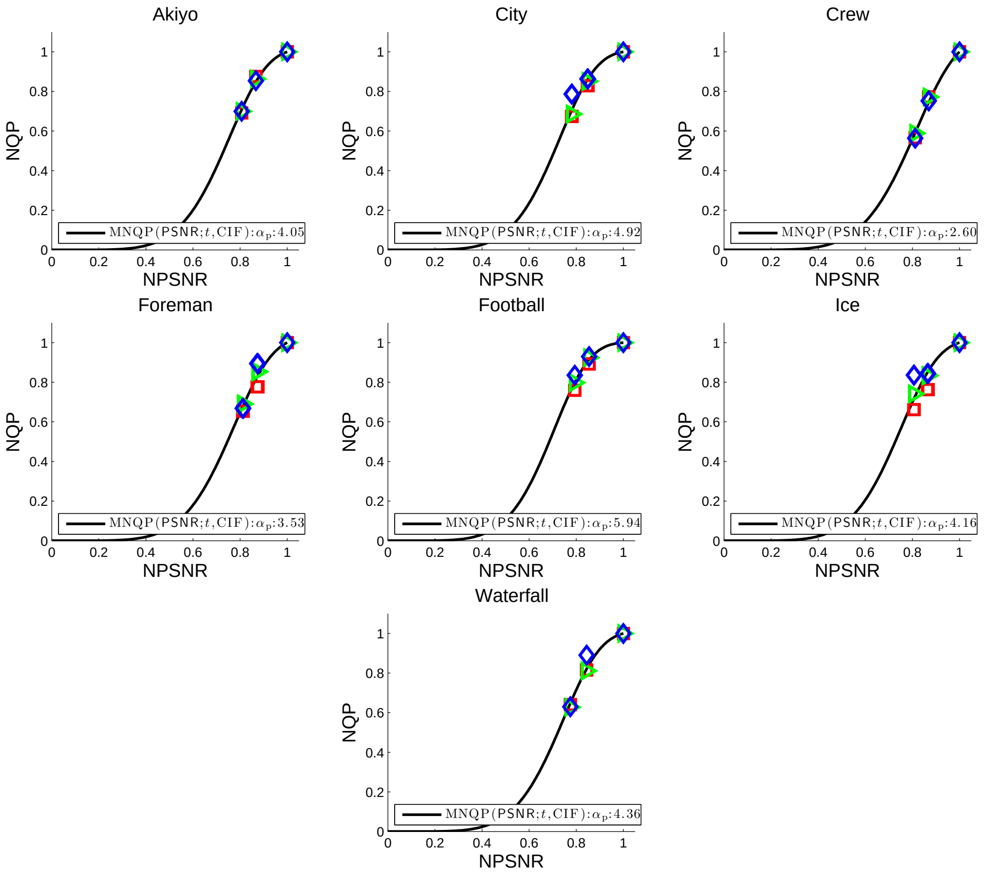


Figure 3.19: Normalized quality v.s. NPSNR. Points are measured data (CSVT data). Curves are derived from the model in (3.14) for all t at CIF. The parameter α_p for each sequences and SR is determined by least squares fitting of data at all TR's. PCC=0.973, RMSE=0.031.

$\mathsf{PSNR}/\mathsf{PSNR}_{\max}$ (NQP) at same SR and TR in Fig. 3.15. The NQP is defined as

$$\mathrm{NQP}(\overline{\mathsf{PSNR}}; s, t) = \frac{\mathrm{MOS}(s, t, \mathsf{PSNR})}{\mathrm{MOS}(s, t, \mathsf{PSNR}_{\max})}, \tag{3.12}$$

Let $\overline{\mathsf{PSNR}} = \mathsf{PSNR}/\mathsf{PSNR}_{\max}$ denote the normalized PSNR and $\mathsf{PSNR}_{\max}$ is the maximum PSNR of each source sequence. It is noted that the PSNR is the average PSNR of frames included in the decoded video (not including interpolated frames). In Fig. 3.15, we can observe that the dropping trends of NQP for different TR's but the same SR tend to cluster together. This reveals that the NQP depends on SR but not TR. By examining the general trend of how NQP changes with PSNR at the same SR, we propose a model for NQP data, called MNQP, with a function form of,

$$\mathrm{MNQP}_1(\overline{\mathsf{PSNR}}; s) = \beta_{\mathrm{p}_1} - \frac{\beta_{\mathrm{p}1}}{1 + e^{\beta_{\mathrm{p}2}(\overline{\mathsf{PSNR}} - \alpha_{\mathrm{p}})}}, \text{or} \tag{3.13}$$

Where α_{p} is the model parameter. Based on the previous analysis, we assume α_{p} depends on s but not t. We derive α_{p} for each SR for a test sequence by least squares fitting using measured NQP data for that SR, at all TR's. We further found that for all seven source sequences, β_{p_1} and β_{p_2} are two constants of 1.05 and 0.33, respectively. Figure 3.14 shows that the MNQP model is very accurate with PCC=0.986, RMSE=0.026. The similar trend of quality degradation with PSNR is also observed in [53], in which we conducted the subjective quality assessment with quantization and frame rate artifacts at fixed CIF resolution on laptop monitor. It shows that predicted quality with PSNR in both tests can be approximated using this sigmoid function accurately. To further validate the assumption that NQP is independent of TR, we also evaluate the model accuracy when the parameter α_{p} is allowed to vary with TR. Table 3.10 shows that allowing α_{p} to vary with TR does not improve the model accuracy significantly. This suggests that we can use the same model parameter for different TR's in MNQP to reduce model complexity.

We further propose a new function to model the quality degradation with $\overline{\mathsf{PSNR}}$ using inverted exponential function, i.e.,

$$\mathrm{MNQP}_2(\overline{\mathsf{PSNR}}; s) = \frac{1 - e^{-\alpha_{\mathrm{p}} \cdot (\overline{\mathsf{PSNR}})^{\beta_{\mathrm{p}}}}}{1 - e^{-\alpha_{\mathrm{p}}}}, \tag{3.14}$$

where α_{p} is the model parameter and β_{p} is a constant value of 5.8. Figure 3.17 illustrates the fitting curves with measured data very well with PCC = 0.985.

Table 3.9: The model performance of NQP

Function	Assumption	PCC	RMSE
$\beta_{\mathrm{p}_1} - \frac{\beta_{\mathrm{p}_1}}{1+e^{\beta_{\mathrm{p}_2}(pp-\alpha_{\mathrm{p}})}}$	$pp = \mathsf{PSNR}$	0.985	0.037
	$pp = \overline{\mathsf{PSNR}}$	0.964	0.058
$\frac{1-e^{-\alpha_{\mathrm{p}}\cdot(\overline{\mathsf{PSNR}})^{\beta_{\mathrm{p}}}}}{1-e^{-\alpha_{\mathrm{p}}}}$		0.970	0.053

Table 3.10: Model accuracy under different assumptions.

Model	Assumptions	PCC	RMSE
MNQP (3.14)	α_{p} depends on TR	0.988	0.33
	α_{p} independent of TR	0.970	0.053
MNQR	α_r depends on TR	0.986	0.036
	α_r independent of TR	0.985	0.039

3.3.2 Modeling Normalized Quality v.s. Normalized Bit Rate

In this subsection, we explore how bit rate affects perceived quality when TR and SR are fixed. Towards this goal, let $r_{\max}$ denote the maximum bit rate for each source sequences at each SR and TR combination, we plot the normalized quality v.s. normalized bit rate $\tilde{r} = r/r_{\max}$ (NQR) at same TR and SR in Fig. 3.20. The NQR is defined as

$$\mathrm{NQR}(\tilde{r}; s, t) = \frac{\mathrm{MOS}(s, t, r)}{\mathrm{MOS}(s, t, r_{\max})}, \tag{3.15}$$

As shown in Fig. 3.20, we can observe that the dropping trends of NQR for different SR's and TR's tend to cluster together. This shows that NQR depends on SR but not TR. By examining the general trend of how NQR changes with $\tilde{r}$, we propose a model for NQR data, called MNQR, with a function form of,

$$\mathrm{MNQR}(\tilde{r}; s) = \frac{1 - e^{-\alpha_r \cdot (\tilde{r})^{\beta_r}}}{1 - e^{-\alpha_r}}. \tag{3.16}$$

The parameter α_r controls how fast the NQR drops as $\tilde{r}$ decreases, with a smaller value corresponding to a faster dropping rate. Based on the previous analysis, we assume α_r is

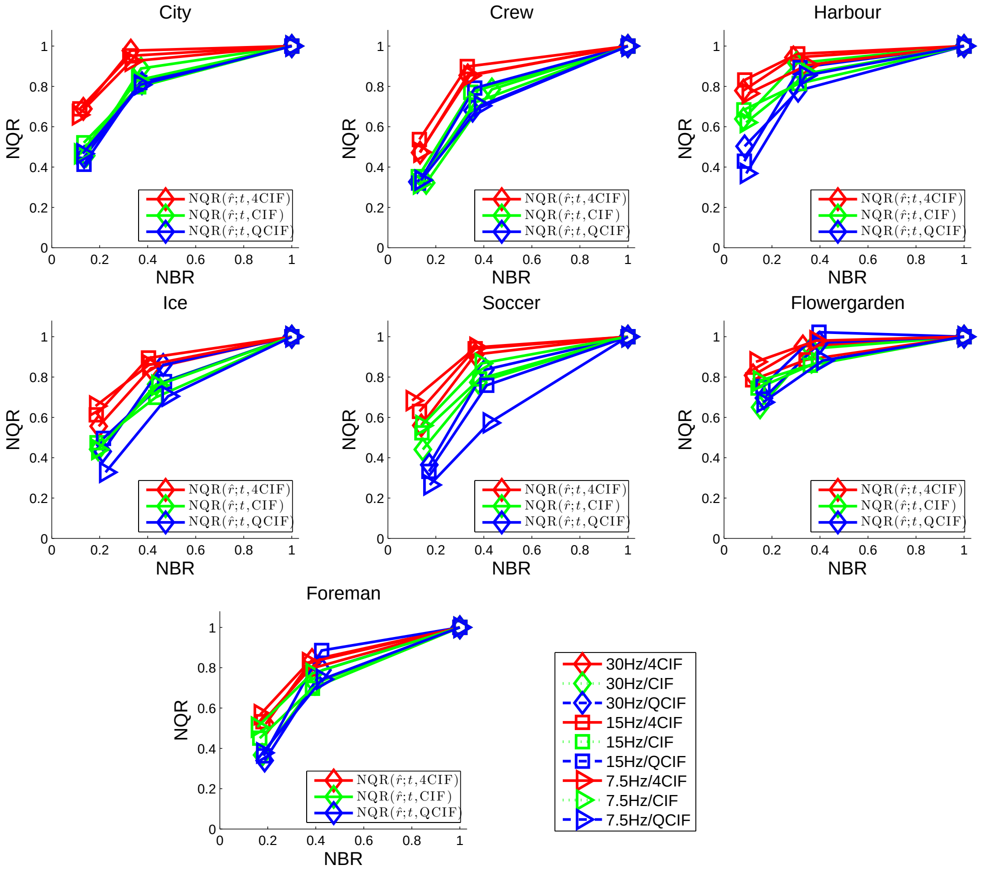


Figure 3.20: Measured NQR under different SR's and TR's. Note that lines with the same color correspond to NQR data at different TR's but the same SR.

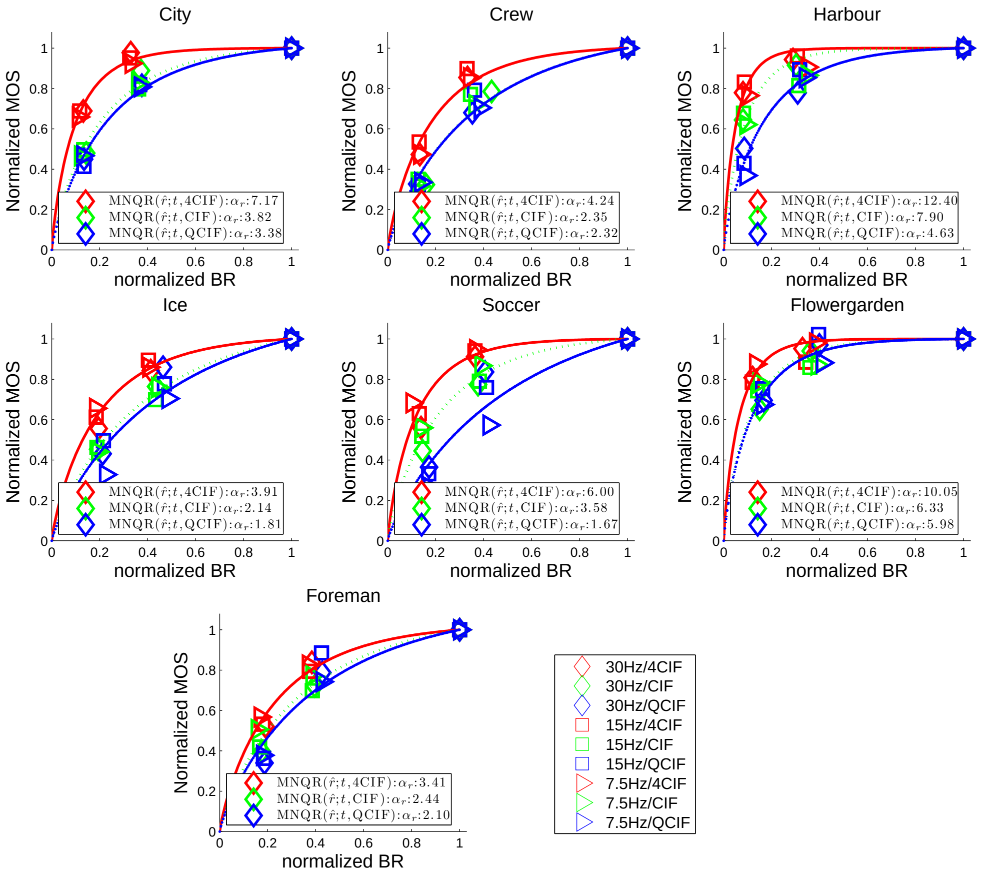


Figure 3.21: Normalized quality v.s. normalized bit rate. Points are measured data under different TR's and SR's. Curves are derived from the model given in (3.16) with β_r=0.86. The model parameter α_r is determined by least squares fitting of data at all SR's and TR's. PCC=0.985, RMSE=0.039.

independent of both SR and TR, and derive its value for each test sequence by least squares fitting using measured NQR data at all SR's and TR's. We also found that β_r is a constant of 0.86 for all seven sequences, which is derived by least square fitting for all NQR data. Figure 3.21 shows that the model curves can capture the data trends well with PCC=0.985, RMSE= 0.039. We also compute the PCC and RMSE of the model when using a best fitting α_r for each different pair of SR and TR. Table 3.10 shows that this brings slight improvement in terms of PCC and RMSE. However, considering that we already achieve high PCC and low RMSE with a parameter that is independent of both SR and TR, we choose to use this option to reduce the model complexity.

3.3.3 The Overall Q-STAR Model

Let $\tilde{s} = s/s_{\max}$, $\tilde{t} = t/t_{\max}$, $\tilde{q} = q/q_{\min}$ denote the normalized SR, TR and QS, respectively. Recall that the overall QSTAR model in [56], which is a function of $\tilde{s}$, $\tilde{t}$, $\tilde{q}$, can be recognized as

$$\begin{aligned} \mathrm{QSTAR}(\tilde{s}, \tilde{t}, \tilde{q}) &= \mathrm{MNQQ}(\tilde{q}; s_{\max})\mathrm{MNQS}(\tilde{s}; \tilde{q})\mathrm{MNQT}(\tilde{t}) \\ &= e^{\alpha_q\left(1-\tilde{q}\right)} \frac{1 - e^{-\hat{\alpha}_s L((\mathrm{QP}(q))(\tilde{s})^{\beta_s}}}{1 - e^{-\hat{\alpha}_s L(\mathrm{QP}(q))}} \frac{1 - e^{-\alpha_t \cdot (\tilde{t})^{\beta_t}}}{1 - e^{-\alpha_t}}. \end{aligned} \tag{3.17}$$

By replacing $\mathrm{MNQQ}(\tilde{q}; s_{\max})$ to $\mathrm{MNQP}(\overline{\mathsf{PSNR}}; s_{\max})$ in (3.13), the new QSTART model can be re-written, i.e.,

$$\begin{aligned} &\mathrm{QSTAR}(\tilde{s}, \tilde{t}, \overline{\mathsf{PSNR}}) = \\ &\left(\beta_{\mathrm{p}_1} - \frac{\beta_{\mathrm{p}1}}{1 + e^{\beta_{\mathrm{p}2}(\overline{\mathsf{PSNR}} - \alpha_{\mathrm{p}})}}\right) \frac{1 - e^{-\hat{\alpha}_s L((\mathrm{QP}(q))(\tilde{s})^{\beta_s}}}{1 - e^{-\hat{\alpha}_s L(\mathrm{QP}(q))}} \frac{1 - e^{-\alpha_t \cdot (\tilde{t})^{\beta_t}}}{1 - e^{-\alpha_t}}, \end{aligned} \tag{3.18}$$

is a function of $\tilde{s}, \tilde{t}, \overline{\mathsf{PSNR}}$. We plot the predicted quality v.s. $\overline{\mathsf{PSNR}}$ in Fig. 3.22 with PCC = 0.991 , RMSE = 0.031. The fitting is still quite good comparing with the model in (3.17) with PCC = 0.992, RMSE = 0.029. Note that PSNR, here, is the average PSNR of decoded video frames. We first predict the quality of videos at different PSNRs for $t_{\max}$ and $s_{\max}$ and then apply the MNQS and MNQT to correct the quality of videos at lower $\tilde{s}$

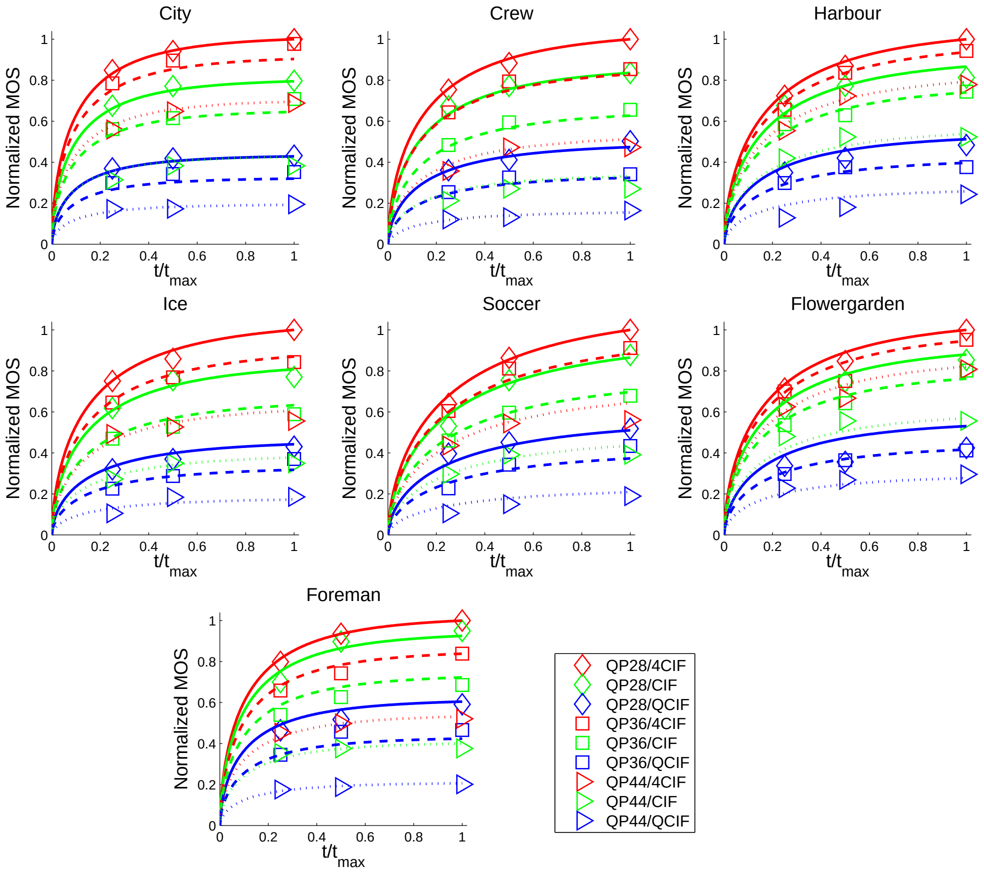


Figure 3.22: Predicted normalized quality (in curves) and measured normalized MOS (in points) v.s. $t/t_{\max}$ under different SR's and TR's using (3.18). The model parameters $\alpha_{\rm p}$, $\hat{\alpha}_s$ and α_t, are obtained by least square fitting.

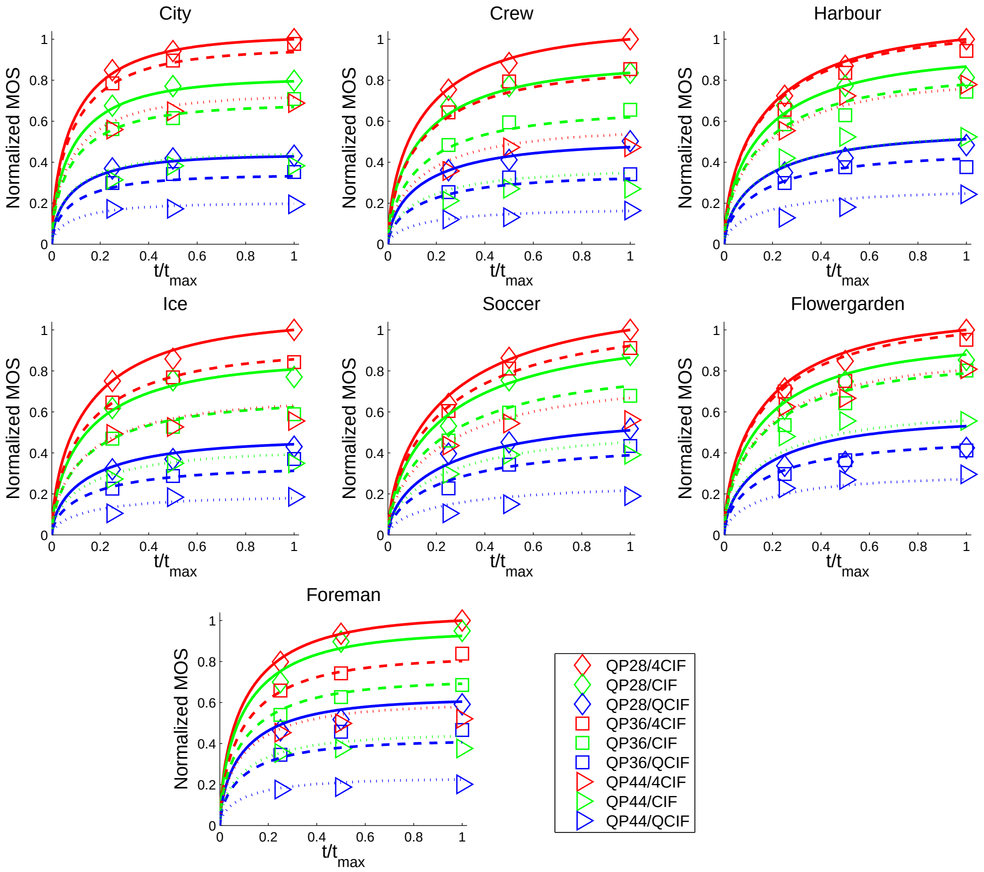


Figure 3.23: Predicted normalized quality (in curves) and measured normalized MOS (in points) v.s. $t/t_{\max}$ under different SR's and TR's using (3.19). The model parameters α_r, $\hat{\alpha}_s$ and α_t, are obtained by least square fitting.

and $\tilde{t}$. This is based on the observations that the PSNR of lower s or t is similar to those at $t_{\max}$ or $s_{\max}$.

By replacing $\mathrm{MNQQ}(\tilde{q}; s, t)$ in (3.17) to $\mathrm{MNQP}(\tilde{r}; s_{\max})$ in (3.16), QSTAR can be addressed to

$$\mathrm{QSTAR}(\tilde{s}, \tilde{t}, \tilde{r}) = \frac{1-e^{-\alpha_r\cdot(\tilde{r})^{\beta_r}}}{1-e^{-\alpha_r}} \frac{1-e^{-\hat{\alpha}_s L((\mathrm{QP}(q))(\tilde{s})^{\beta_s}}}{1-e^{-\hat{\alpha}_s L(\mathrm{QP}(q))}} \frac{1-e^{-\alpha_t\cdot(\tilde{t})^{\beta_t}}}{1-e^{-\alpha_t}}, \tag{3.19}$$

where β_r=0.86. As shown in Fig. 3.23, the predicted quality fits the measured MOS very accurate, with PCC = 0.989, RMSE = 0.035. As can be seen in Fig. 3.20, the PVS with lower s or t has similar $\tilde{r}$ with that at $s_{\max}$ or $t_{\max}$. We first, apply MNQP model in (3.13) to predict the quality with different bit rate levels at $s_{\max}$ and $t_{\max}$, following by the MNQS and MNQT to correct the video quality at lower s and t. Note that in both models (3.18) and (3.19), we still keep those constants remain the same value, i.e., $\beta_s = 0.74$, $\beta_t = 0.63$ and υ_1=$-$0.037, υ_2=2.25 in $L((\mathrm{QP}(q))$ in these two new proposed models while both of them also remain three content-dependent parameters $\alpha_{\mathrm{p}}/\alpha_{\mathrm{r}}$, $\hat{\alpha}_s$ and α_t. Table 5.5 summarizes the model parameters, RMSE and PCC of the proposed QSTAR model in (3.18) and (3.19). In this table we also list the average 95% confidence interval (CI) of user ratings (normalized by maximum possible rating of each source sequence) for each source sequence. We see that the RMSE of the prediction error is much lower than the CI for all sequences. The correlation scatter plots between predicted and measured quality using (3.18) and (3.19) are presented in Fig. 3.24.

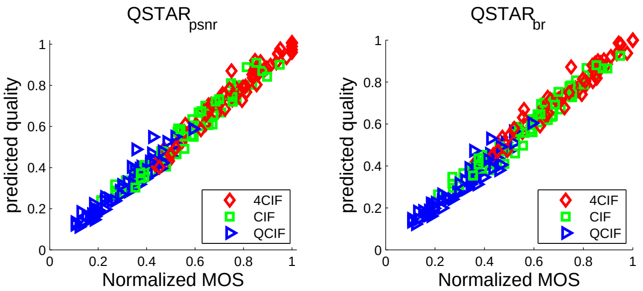


Figure 3.24: Predicted quality using QSTAR model against measured MOS. Left: Predicted quality by (3.18); Right: Predicted quality by (3.19).

Table 3.11: The parameters and performance of QSTAR model.

	city	crew	harbour	ice	soccer	fg	foreman	Avg.
Parameters obtained by least square fitting with MOS data using (3.17)								
α_q	7.25	4.51	9.65	5.61	6.31	10.68	4.57	
$\hat{\alpha_s}$	3.52	4.07	4.58	3.68	4.55	4.83	5.94	
α_t	4.10	3.09	2.83	3.00	2.23	2.80	3.80	
RMSE	0.018	0.025	0.038	0.033	0.032	0.058	0.038	0.035
PCC	0.998	0.996	0.992	0.993	0.992	0.979	0.991	0.991
Parameters obtained by least square fitting with MOS data using (3.18)								
α_{p}	26.67	30.39	24.17	32.41	27.98	23.92	32.59	
RMSE	0.017	0.030	0.034	0.026	0.0304	0.043	0.035	0.029
PCC	0.998	0.996	0.990	0.994	0.991	0.987	0.989	0.991
Parameters obtained by least square fitting with MOS data using (3.19)								
α_r	7.17	4.23	12.40	3.91	6.00	10.04	3.40	
RMSE	0.024	0.033	0.036	0.029	0.037	0.050	0.0341	0.035
PCC	0.995	0.991	0.989	0.993	0.989	0.978	0.9898	0989
avg. CI	0.048	0.049	0.050	0.050	0.050	0.051	0.049	0.050

3.4 Model Verification

In order to verify the model accuracy on other datasets, we apply QSTAR model partially or fully on eight different datasets. The brief description of these datasets are listed in Tab. 3.12. First, we apply the bit rate model MNQR (3.16) and MNQQ (3.7) on DataSet#2 − 3 in Figs. 3.25-3.30. We also evaluate MNQP on DataSet#4 in Fig. 3.31 and 3.32. Table 3.13 summarize the PCC and RMSE for all seven datasets with individual models (e.g., MNQQ, MNQP) and joint model (e.g., QSTAR($\tilde{s}$, $\tilde{q}$;30Hz) and QSTAR($\tilde{s}$,$\tilde{t}$, $\tilde{q}$). These datasets includes different codecs and different experiment settings or configurations. Note that for validation of individual model, sometimes we only include partial data, which meet the requirement of quality metric. For example, MNQT can only be apply for fixed SR, QS or same frame-quality videos.

Table 3.12: Data Set Description

DataSet#1	The 7 source sequences used in this paper, obtained with 3 frame rates (30, 15, 7.5 Hz), 3 spatial resolutions (4CIF, CIF, QCIF), and 3 QP levels (28, 36, 44). A total of 189 PVSs.
DataSet#2	The 5 source sequences used in [4], obtained with 3 frame rates (30, 15, 7.5 Hz), five different bit rate levels (each frame rate has its corresponding 5 QPs) coded by H.263 for CIF resolutions. A total of 75 PVSs.
DataSet#3	The 5 source sequences used in [4], obtained with CIF , 3/4 CIF and QCIF resolutions, five different bit rate levels (each SR has its corresponding 5 QPs) coded by H.263 for 30Hz . A total of 75 PVSs.
DataSet#4	5 QCIF sequences used in [5], each with 3 frame rates (30, 15, 7.5 Hz) and 4 bit rate levels coded by H.264. A total of 54 PVSs.
DataSet#5	5 QCIF sequences used in [5], each with 3 frame rates (30, 15, 7.5 Hz) and 4 bit rate levels coded by H.263. A total of 54 PVSs.
DataSet#6	3 source sequences used in [6], each with 3 frame rates (30, 15, 7.5 Hz), 6 spatial resolutions (in between QCIF and CIF), and 3 bit rate levels coded by H.264/SVC. A total of 54 PVSs.
DataSet#7	Selected 5 source sequences used in [7], each with 1080i spatial resolution, and 3 or 4 QP levels (or bit rates) coded by H.264 and MPEG-2, respectively. Only lossless sequences included. A total of 35 PVSs.
DataSet#8	3 source sequences used in [8], each with 3 spatial resolutions, 4 frame rates (50, 25, 12.5, 6.25) and several bit rate levels coded by H.264/SVC. A total of 26 PVSs.
DataSet#9	7 CIF sequences used in chapter 2, each with 4 frame rates (30, 15, 7.5, 3.75 Hz) and 3 QP levels (28, 36, 40). A total of 100 PVSs.

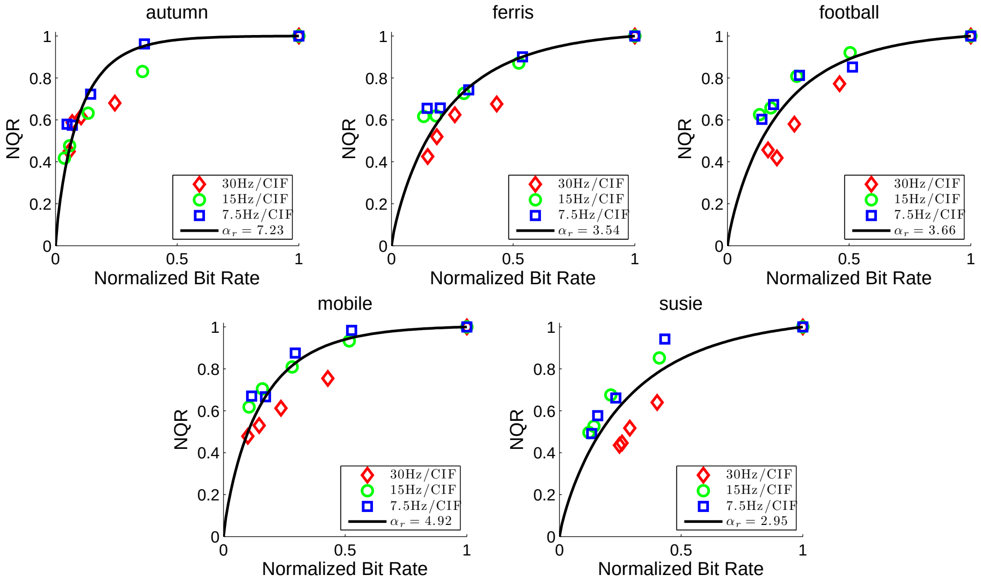


Figure 3.25: Normalized quality v.s. normalized bit rate. Points are measured data under different TR's and SR's from DataSet#2 [4]. Curves are derived from the model given in (3.16) with β_r=0.86. The model parameter α_r is determined by least squares fitting of data at all SR's but same TR. PCC=0.895, RMSE=0.089.

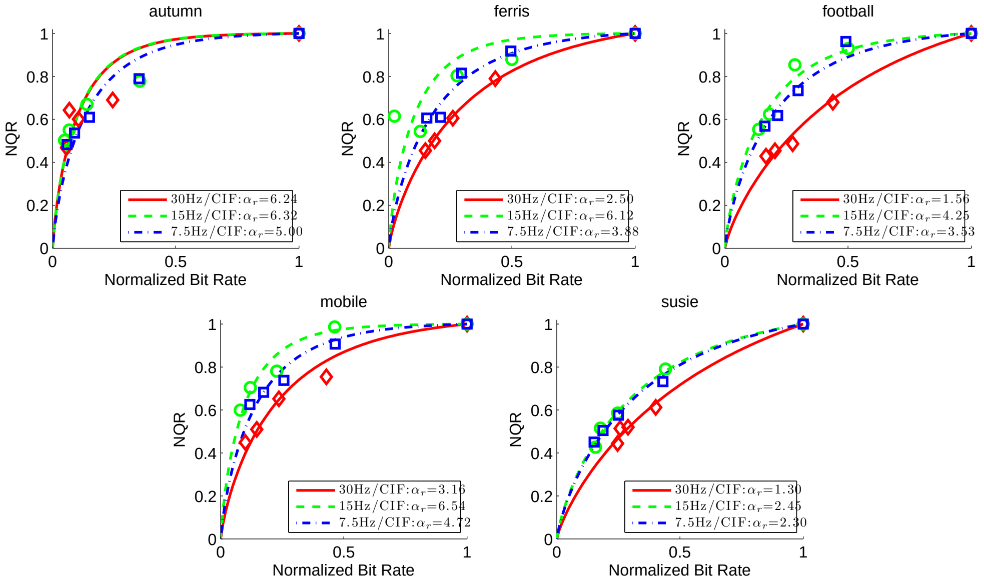


Figure 3.26: Normalized quality v.s. normalized bit rate. Points are measured data under different TR's from DataSet#2 [4]. Curves are derived from the model given in (3.16) with β_r=0.86. The each model parameter α_r is determined by least squares fitting of data at each TR. PCC=0.95, RMSE=0.066. The p-value for TR*BR is 0.01 on NQR data.

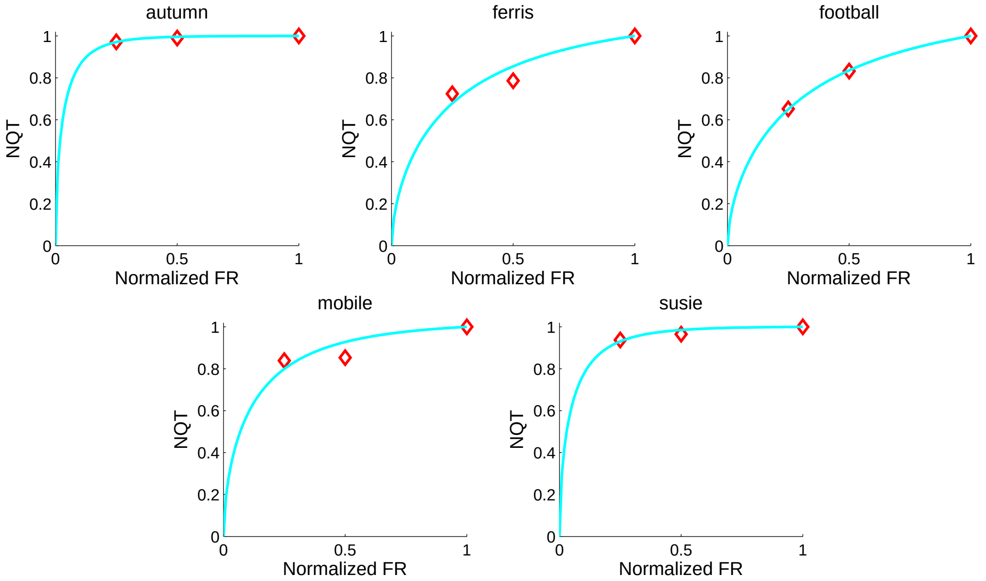


Figure 3.27: Normalized quality v.s. normalized FR. Points are measured data under different TR's from DataSet#2 [4]. Curves are derived from the MNQQ model in (3.17). The model parameter α_q is determined by least squares fitting of data at QS=12 and CIF resolution. PCC=0.964, RMSE=0.031.

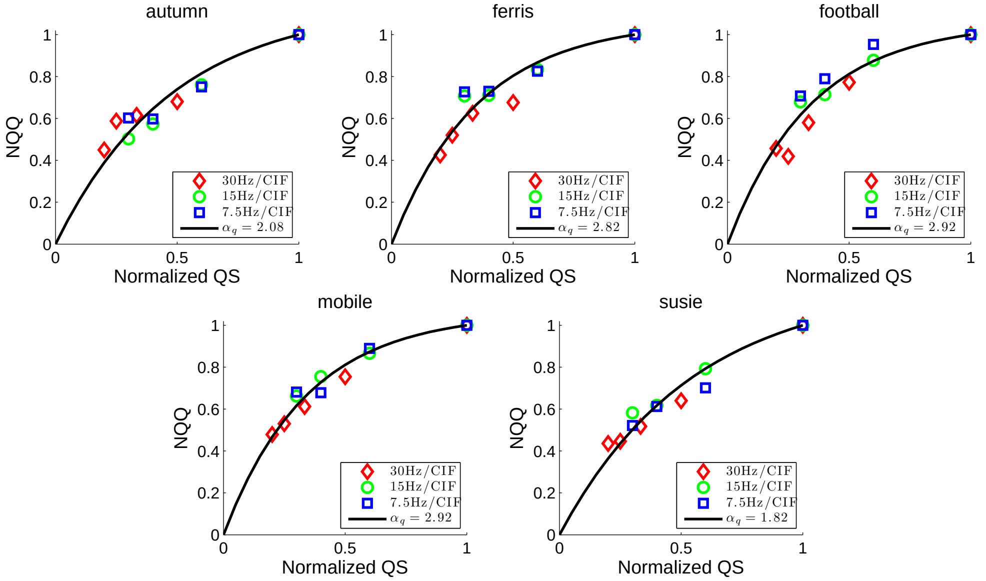


Figure 3.28: Normalized quality v.s. normalized QS. Points are measured data under different TR's from DataSet#2 [4]. Curves are derived from the MNQQ model in (3.7). The model parameter α_q is determined by least squares fitting of data at all TR's. PCC=0.86, RMSE=0.10.

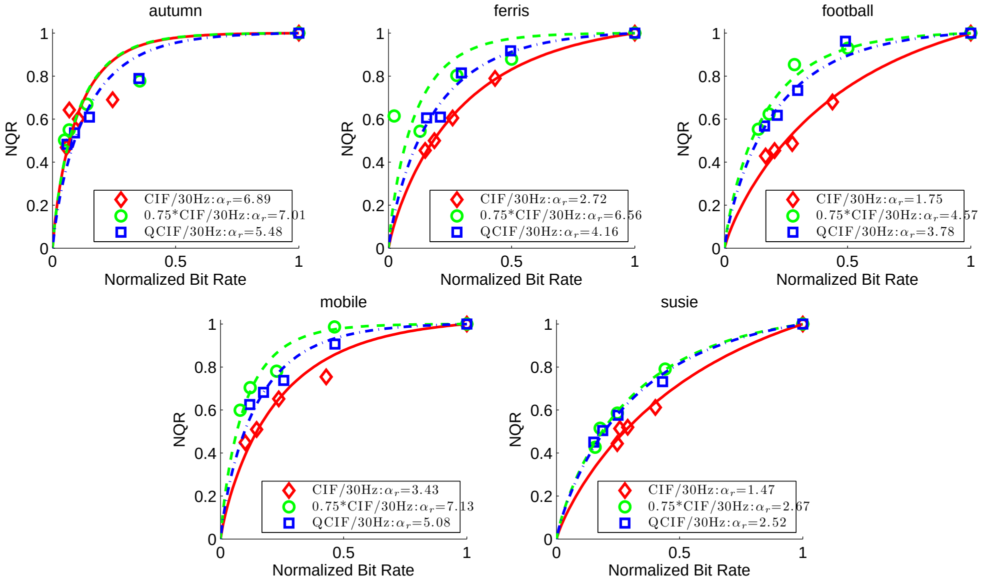


Figure 3.29: Normalized quality v.s. normalized bit rate. Points are measured data under different SR's from DataSet#3 [4]. Curves are derived from the model given in (3.16) with β_r=0.86. The model parameter α_r is determined by least squares fitting of data for each SR. PCC=0.947, RMSE=0.069.

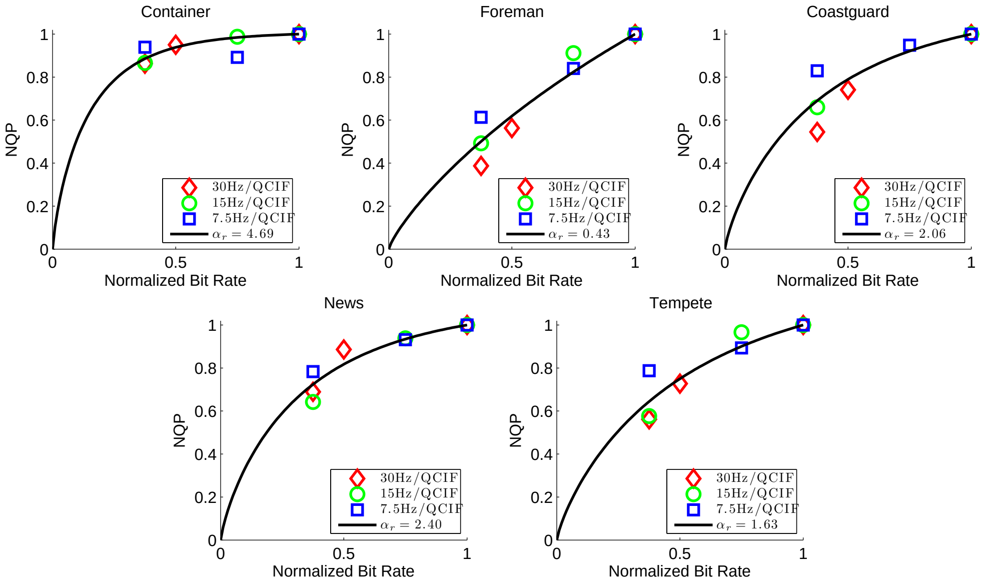


Figure 3.30: Normalized quality v.s. normalized bit rate. Points are measured data under different TR's from DataSet#4 [5] (coded by H.264). Curves are derived from the model given in (3.16) with β_r=0.86. The model parameter α_r is determined by least squares fitting of data at all TR's. PCC=0.934, RMSE=0.061. The p-value for TR*BR is 0.051 on NQR data.

3.5 Summary

In this work, we propose a perceptual quality model considering the impact of SR, TR and QS on mobile display platforms. In this model, we use a one-parameter function to capture the quality decay v.s. SR, TR and QS individually. The parameter in each function is sequence dependent. Interestingly, we found that the dropping rate of the quality with TR, characterized by α_t, is independent of SR and QS, and the dropping rate of quality with both SR and QS, indicated by α_s and α_q, respectively, are both independent of TR. Although the dropping rate α_s with SR is dependent on QP, we found that they are related linearly. The overall model only requires three content-dependent parameters. The model was validated by subjective ratings for compressed video sequences under a large range of SR, TR, and QS, for seven source videos with large variations in their motion and texture

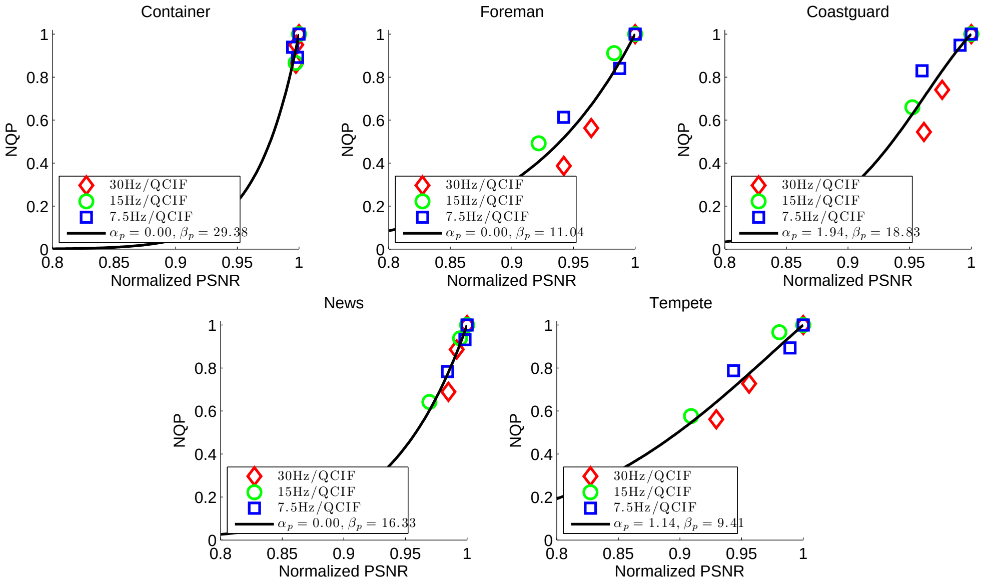


Figure 3.31: Normalized quality v.s. normalized PNSR. Points are measured data under different TR's from DataSet#4 [5]. Curves are derived from the model given in (3.14). The model parameter α_p and β_p are determined by least squares fitting of data at all TR's. PCC=0.930, RMSE=0.063.

characteristics. The model with the content-derived features has a high PCC (=0.988) with subjective ratings.

As a conclusion, it is worth noting the implication of the proposed model form in (3.11). It suggests that the quality of a video is the product of a spatial quality factor (jointly determined by SR and QS) and a temporal quality factor (determined by TR). The spatial term is in turn the product of two factors, MNQQ and MNQS. MNQQ describes how does QS affects the quality when the video is at the maximum SR; and MNQS accounts for the quality degradation due to lower SR. The rate of degradation depends on QP, as indicated by the dependency of the parameter α_s on QP. We would like to note that, in the work presented in chapter 2 we also found that the quality is the product of a spatial factor and a temporal factor, and the parameters of the two factors are independent. In addition to MNQQ model relating NQQ with QS, we further propose MNQP and MNQR when QS is not available. They hold the same merit of MNQQ that the falling trend is independent of TR but SR. Both MNQP and MNQR are derived using inverted exponential function, which

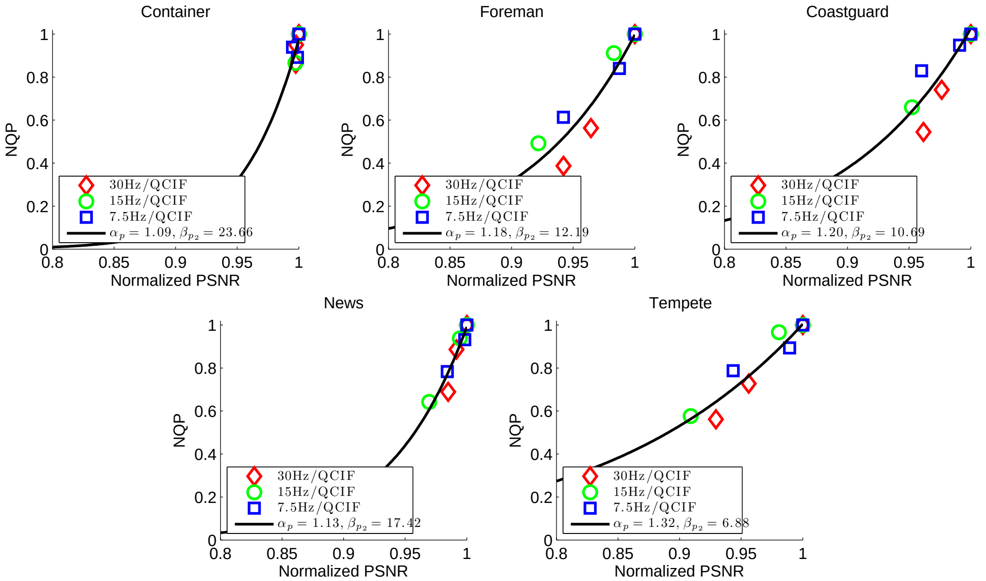


Figure 3.32: Normalized quality v.s. normalized bit rate. Points are measured data under different TR's from DataSet#4 [5]. Curves are derived from the model given in (3.13) with β_{p_1}=10. The model parameter α_{p} and β_{p_2} are determined by least squares fitting of data at all TR's. PCC=0.931, RMSE=0.063.

approximate the measured data very well with PCC=0.98. The proposed NQP or NQR is shown to be highly accurate, compared to the subjective ratings from our own subjective tests as well as test results reported from other datasets.

Although the proposed model is developed for videos generated by the H.264/SVC codec, we expect that the same function form is applicable to scalable videos coded using other codecs and to non-scalable videos coded at different (s,t,q) combinations. However, the model parameters for the same video content may differ, depending on the encoder configurations. This hypothesis needs to be validated in future studies.

The proposed quality models, together with the rate model, also as a function of STAR in [71], can be used to determine the optimal STAR that maximizes the quality given a rate constraint, both for video encoding/transcoding and for scalable video adaptation. Our prior work [54, 72] has investigated a subset of this problem, where SR is fixed, and only TR and QS are adapted, based on quality and rate models as functions of TR and QS only. Extension of this work to include the SR dimension, using the newly developed quality and

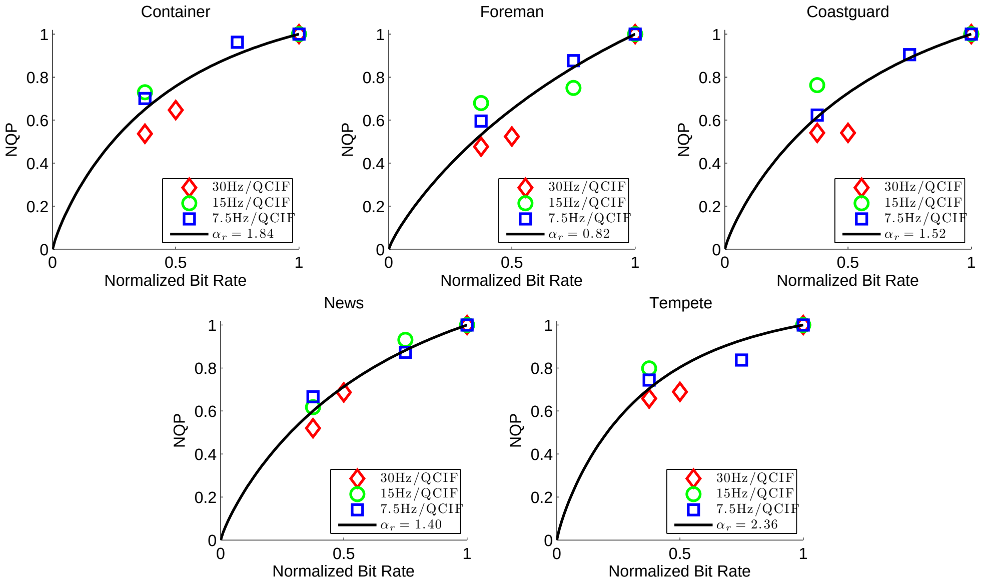


Figure 3.33: Normalized quality v.s. normalized bit rate. Points are measured data under different TR's from DataSet#5 [5] (coded by H.263). Curves are derived from the model given in (3.16) with β_r=0.86. The model parameter α_r is determined by least squares fitting of data at all TR's. PCC=0.913, RMSE=0.074. The p-value for TR*BR is 0.0001 on NQR data.

rate models, both as functions of SR, TR, and QS, is another interesting direction for future research.

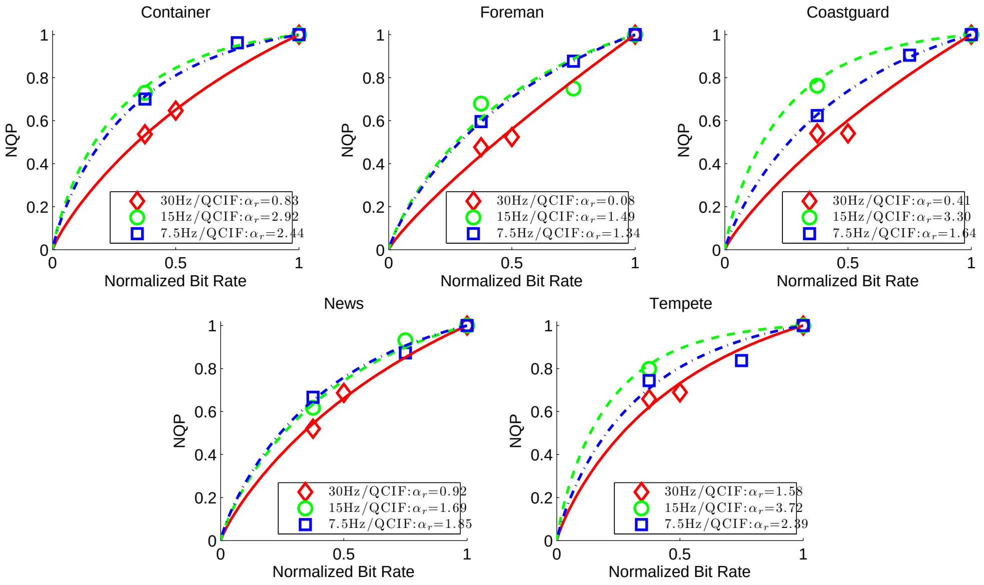


Figure 3.34: Normalized quality v.s. normalized bit rate. Points are measured data under different TR's from DataSet#5 [5] (coded by H.263). Curves are derived from the model given in (3.16) with β_r=0.86. The model parameter α_r is determined by least squares fitting of data at all TR's. PCC=0.976, RMSE=0.039.

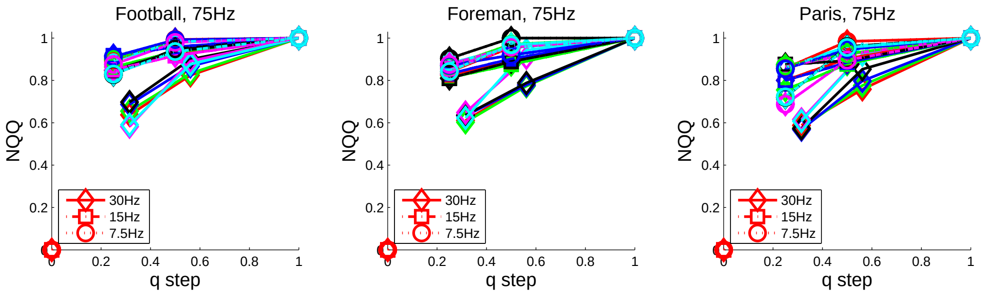


Figure 3.35: Measured NQQ under different SR's and TR's from DataSet#6. Note that lines with the different markers correspond to NQQ data at different TR's but each marker includes five SR's.

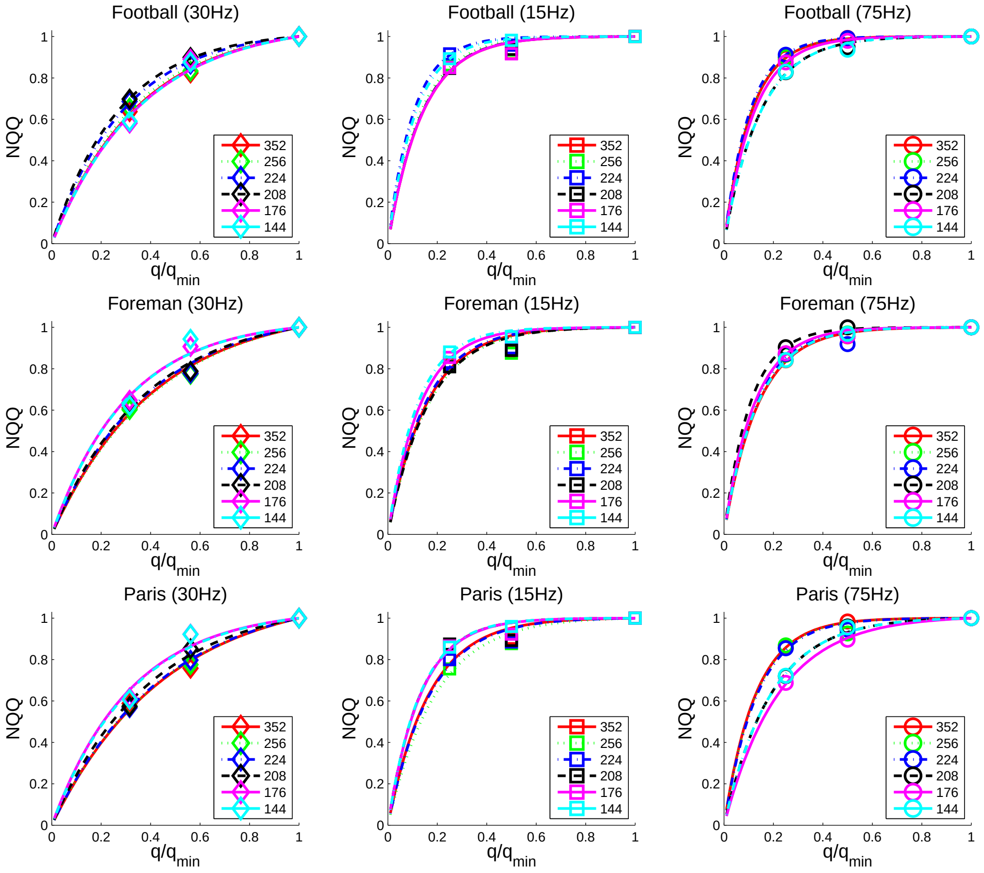


Figure 3.36: Normalized quality v.s. normalized QS. Points are measured data under different TR's and QS's from DataSet#6. Curves are derived from the MNQQ model given in (3.19) with PCC=0.98, RMSE=0.022. QP = 32, 37, 42 for 30Hz; QP = 23, 29, 35 for 15 and 7.5HZ

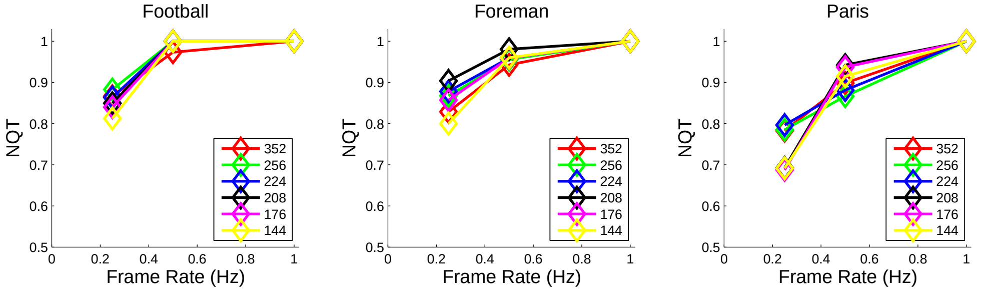


Figure 3.37: Measured NQT under different SR's but same QS from DataSet#6 [6]. Note that lines with the different markers correspond to NQT data at different SR's but the same QS.

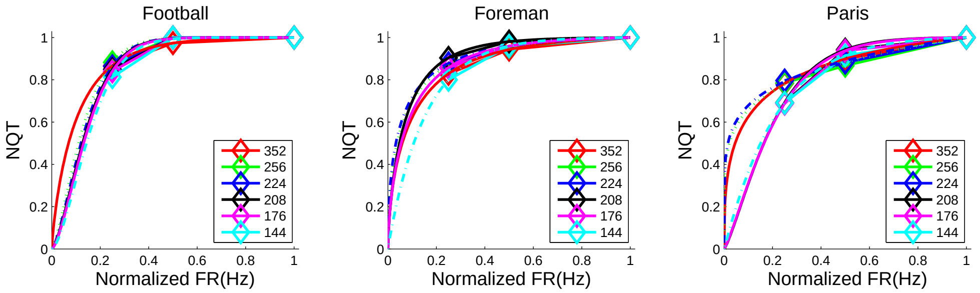


Figure 3.38: Normalized quality v.s. normalized FR. Points are measured data under different QS's but same QP=35 (37 for 30Hz) from DataSet#6. Curves are derived from the MNQT model given in (3.9) with PCC=0.99, RMSE=0.003.

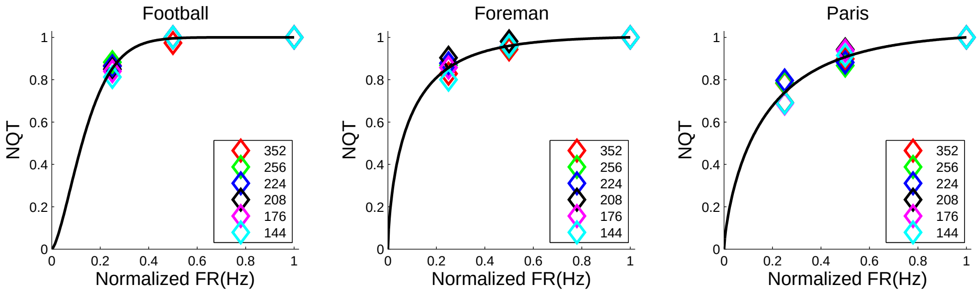


Figure 3.39: Normalized quality v.s. normalized FR. Points are measured data under different SR's but same QP=35 (37 for 30Hz) from DataSet#6. Curves are derived from the MNQT model given in (3.17) usign the same model parameter α_t for all different SR's with PCC=0.96, RMSE=0.023.

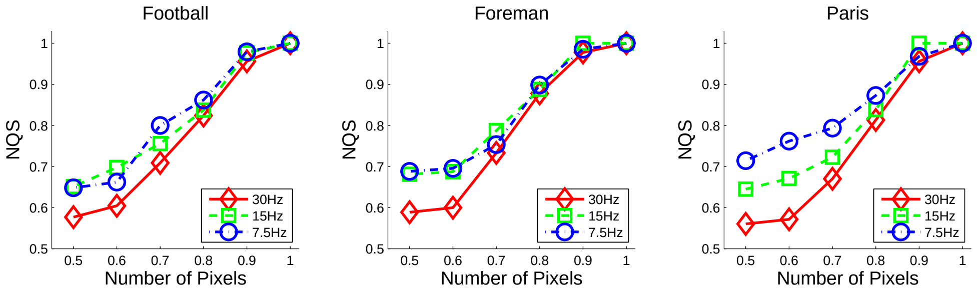


Figure 3.40: Measured NQS under different TR's from DataSet#6 [6]. Note that lines with the different markers are corresponding to different TR's.

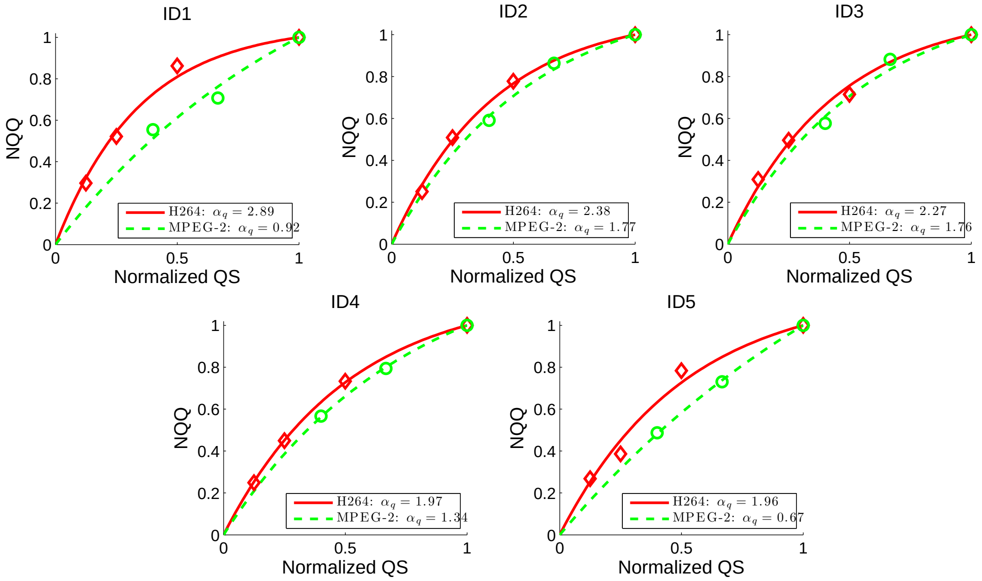


Figure 3.41: Normalized quality v.s. normalized QS. Points are measured data under different codecs (H.264 and MPEG-2) DataSet#7 [7]. Curves are derived from the MNQQ model given in (3.17) with PCC=0.993, RMSE=0.029. Note that lines with the different lines correspond to different codecs.

Table 3.13: The parameters and performance of QSTAR model.

Model	Metrics	DataSet#1	#2	#3	#4	#5	#6	#7	#8	#9
QSTAR(s, t, q)	RMSE	0.035	-	-	-	-	0.039	-	0.08	-
	PCC	0.991	-	-	-	-	0.959	-	0.856	-
QSTAR(t, q;CIF)	RMSE	-	0.120	-	-	-	0.023	-	-	0.044
	PCC	-	0.854	-	-	-	0.975	-	-	0.973
QSTAR(s, q;30Hz)	RMSE	-	-	0.112	-	-	-	-	-	-
	PCC	-	-	0.927	-	-	-	-	-	-
MNQQ(q) (3.7)	RMSE	0.041	0.052	-	-	-	0.022	0.029	0.079	0.046
	PCC	0.982	0.961	-	-	-	0.98	0.993	0.819	0.963
MNQT(t)	RMSE	0.052	0.030	-	-	-	0.023	-	0.032	0.034
	PCC	0.891	0.964	-	-	-	0.960	-	0.987	0.968
MNQS(s)	RMSE	0.030	-	-	-	-	-	-	-	-
	PCC	0.992	-	-	-	-	-	-	-	-
MNQR(r)	RMSE	0.039	0.066	0.069	0.061	0.039	-	-	-	-
	PCC	0.986	0.950	0.941	0.934	0.976	-	-	-	-
MNQP($\overline{\text{PSNR}}$) (3.14)	RMSE	0.037	-	-	0.063	-	-	-	-	0.031
	PCC	0.985	-	-	0.930	-	-	-	-	0.973

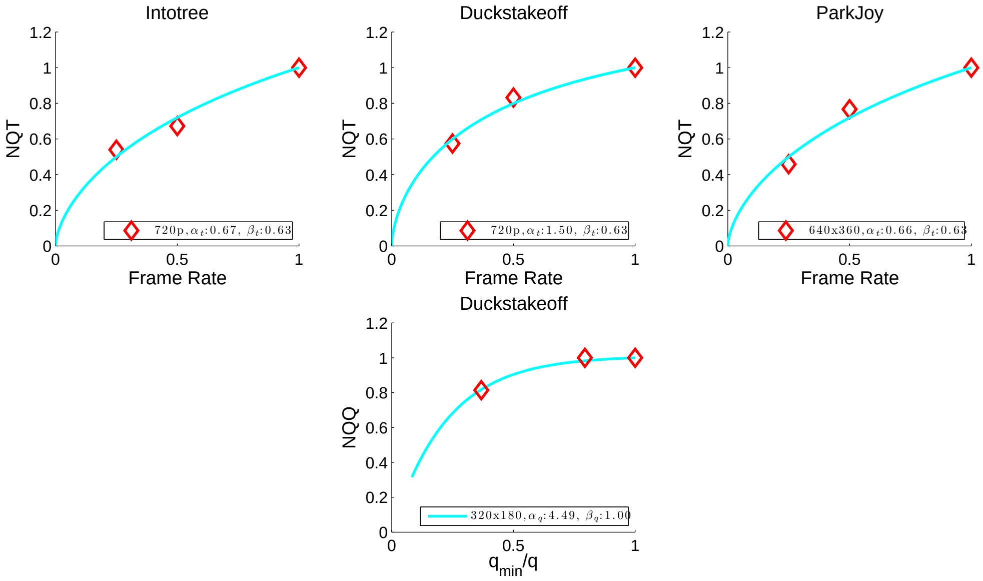


Figure 3.42: Top part: Normalized quality v.s. normalized FR. Points are measured data from DataSet#8 [8]. Curves are derived from the MNQT model given in (3.9) with PCC=0.98, RMSE=0.032. Bottom part: Normalized quality v.s. normalized QS from DataSet#8. Points are measured data and the curve is derived from (3.7) with PCC=0.010, RMSE=0.991.

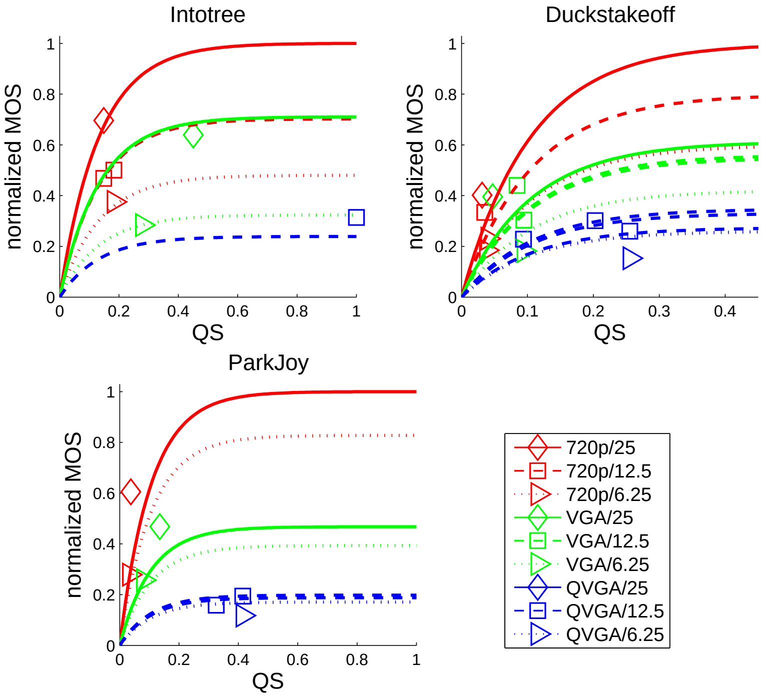


Figure 3.43: Predicted quality v.s. normalized QS. Points are measured data under different TR's, QS's and SR's from DataSet#8 [8].

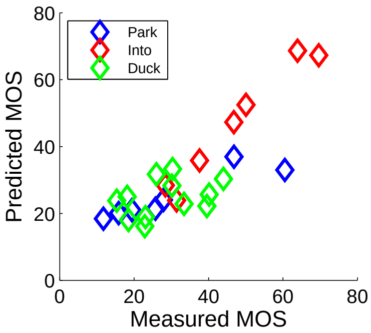


Figure 3.44: Predicted quality v.s. measured MOS. Points are measured data under different TR's, QS's and SR's from DataSet#8 [8]. The scatter plot is based on (3.17). The model parameters are determined by least squares fitting of all data. PCC=0.856, RMSE=0.08.

Chapter 4

Perceptual Quality Modeling of Video with Frame Rate and Quantization Variation

This work investigates the impact of temporal variation of frame rate (FR) or quantization stepsize (QS) on perceptual video quality. Among many dimensions of FR/QS variation, as a first step we focus on videos in which two FR's, or QP's, alternate over fixed intervals. We present subjective test results, and analyze the influence of several factors (including the delta FR/QP, the changing frequencies, and the video content). According the observation and data analysis, we propose to several models to characterize the quality degradation of viewers perception with respect to the variation in FR, QS, or correspondingly in the bit rate. Such quality assessment and modeling are essential in making video adaptation decisions when delivering video over dynamically changing wireless links.

Take for example a hypothetical case where the available bandwidth alternates between R_l and R_h, and the frame rates (FR), denote as t, that can lead to the best perceived quality for constant rate video at R_l and R_h are t_l and t_h, respectively. In this situation, is it better to code the video with alternating FR's of t_l and t_h, or would it better to stay at t_l? More generally, one may want to vary not only the FR, but also the frame size and QS to meet the instantaneous rate constraints. In this work, we focus on the QS or FR variation while keeping the other resolutions fixed. The following description is for the investigation regarding FR variation. A similar study is carried out for QS variation. Among the many dimensions of variations, we consider the simple case where the FR alternates between t_l

and t_h, with each FR staying over a constant time duration Fz. We conduct subjective tests where viewers are asked to rate the quality of video with varying t_l and t_h and Fz. We study the effect of t_h, t_l, their difference $\Delta t = t_h - t_h$ and ratio t_l/t_h on the perceived quality. We include a variety of videos, to further assess the influence of the video content. This study directly addresses the questions we raised for the hypothetical example given earlier. But it also shed lights for more complicated cases where the FR may vary among more than two levels and the variation may not follow a periodic pattern.

This chapter is organized as follows. Section 4.1 describes the subjective test configurations. Section 4.2 and 4.3 presents the subjective test results, the observations, and also proposes a model that relates to the perceived quality with the FR/QS variation. Section 4.4 investigates the statistical significance of impact of FR/QS variation, video content and frequency on perceptual quality using the ANOVA statistical test. Finally Section 4.5 concludes the paper.

4.1 Subjective Test Setup

4.1.1 Testing Material

Our experiment is conducted using five video source sequences, Akiyo, Foreman, Football, Ice, Waterfall, all in CIF (352×288) resolution and at frame rate 30 fps with originally 10 seconds long, which are chosen from JVT (Joint Video Team) test sequence pool [59]. All these sequences are coded using JVT scalable video model (JSVM912) [60]. For each sequence, one bitstream is generated with five temporal layers, with corresponding FR of 1.875, 3.75, 7.5, 15, and 30Hz , and each temporal layer in turn has five quality layers created with QS equal to 28, 32, 36, 40 and 44 (with corresponding to QS = 16, 25, 40, 64, 102), respectively, using the coarse grain scalability (CGS) without QS cascading. For the study reported here, the test videos are obtained by decoding all temporal layers (i.e. FR= 30 Hz) but different number of quality layers, corresponding to the desired QS variation.

Two different experiments, quality impact of FR and QS variation, were implemented.

Table 4.1: Testing configuration for frame rate variation

QS	Fz	t_h(Hz)	t_l(Hz)
16	1/2/3 sec	30	30/15/7.5
		15	15/7.5
		7.5	7.5

Table 4.2: Testing configuration for QS variation

FR	Fz	QS_b	QS_v
30	1/2/3 sec	16	16/25/40/64/102
		40	25/40/64/102
	3 sec	102	25/64/102

For temporal variation, we first make two frame rates switch back and forth periodically through the entire video with changing interval (Fz) of 1, 2, and 3 seconds. Let t_h and t_l denote the higher and lower FR of the video. Table 4.1 details all the test configurations, which leads to a total of 90 processed (encoded and decoded) video sequences (PVS). For QS variation, we fix FR to 30Hz but allow QS to switch back and forth periodically through the entire video with Fz of 1, 2, and 3 seconds. In Tab. 4.2, QS_b indicates the beginning QS while QS_v denotes the deviated QS, which could be either higher or lower than QS_b. There are a total of 130 PVS's

4.1.2 Experiment Configuration

The subjective quality assessment is carried out by using a protocol similar to ACR (Absolute Category Rating) described in [17]. Basically, each viewer is presented a series of video in a random order, and the viewer is asked to give overall rating of each video in the range of 0 to 100. Each test for one subject consists of two sessions, a training session and a test session. The training session (about 2 minutes) is used for the subject to accustom him/herself to the rating procedure and ask questions if any. The PVS's in the test session (about 12 minutes) are ordered randomly so that each subject sees the video clips in a different order. Most of the viewers are engineering students from Polytechnic Institute of New York University, with age 21 to 33. Other details regarding each experiment are given below.

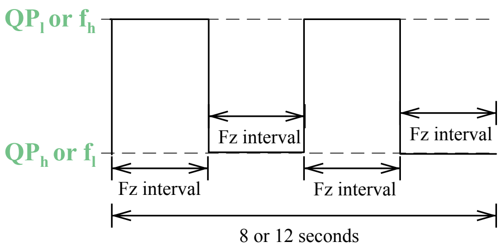


Figure 4.1: The Variations of QS or FR for a video. The sequence is 8 sec long for Fz=1,2, and 12 sec long for Fz=3.

1. The first experiment-QS variation:

   There are two tests included in this experiment. First one has two base QS (16 and 40) with Fz of 1 and 2 seconds. The second one has three base QS (16, 40, and 102) with Fz of 3 seconds. Note that these two tests will be later combined together to explore the quality impact of all Fz's with QS variation as shown in Tab. 4.2. Four common sequences are selected for each source sequence and the selections of testing points are uniformly distributed among the entire range. In order to shorten the duration of the test, each test is divided into two subgroups. Each of them contains around 36 processed video sequences and lasts about 14 minutes. Thirty one non-expert viewers who had normal or corrected-to-normal vision acuity participated in one or two subgroup tests. There are on average 22 ratings for each PVS and a total of 90 and 60 PVS's for the first and second test, respectively.

2. The second experiment-FR variation:

   There are also two tests included in this experiment. First one has all the frame rate variation (30, 15 and 7.5Hz) with Fz of 1 and 3 seconds, defined as Fz. The second one has Fz of 2 and 3 seconds and the rest are the same as first test. These two test are later combined together for studying the quality impact of all Fz's with FR variation as shown in Tab. 4.1. Three common sequences for datasets combination are selected for each source sequence. Similar to first experiment, each test are divided into several subgroups, and each of which contains 36 PVS's. The sequences in the test session are also ordered randomly. Forty two non-expert viewers who had normal or corrected-to-normal vision acuity participated in one or two subgroup tests. There are on average 20 ratings for each PVS and a total of 45 PVS's for each test.

Note that we synthesize the 12 second video with repeating content since the original sequence is only 10 second long. In order to eliminate the annoying effect of scene change while looping back from the start point of video, all 12 second long PVSs start from the rest 6 seconds of the original sequences, and then the first 6 seconds for PVS with Fz of 3 sec. By doing this, the scene change occurs at the same time as the QP/FR switches. For 8 seconds long videos, which is designed for Fz=1 and 2, it is the first 8 second of the original video.

4.1.3 Data Post Processing

The raw ratings are converted to Z-scores [68] based on the mean and standard deviation of all the scores of each viewer, given by

$$Z_{mij} = \frac{X_{mij} - \mathrm{MEAN(X_i)}}{\mathrm{STD(X_i)}}. \tag{4.1}$$

Here, X_{mij} and Z_{mij} denote the raw rating and Z-score of m^{th} sequence at j^{th} STAR combination, from i^{th} viewer, respectively. X_i denotes all ratings from i^{th} viewer. $\mathrm{MEAN}(\cdot)$ and $\mathrm{STD}(\cdot)$ represent the operator for taking the mean and the standard deviation of a given set, respectively.

Post Screening

In order to remove "noisy" ratings or outliers, we adopted, with some modification, two post screening methods in concatenation. We first perform BT.500-11 post screening method [16] in Z-score domain to remove all ratings by certain viewers because their ratings are outside the range of the majority of the viewers. On average, one viewer is eliminated for each PVS. We then conduct the second step to the remaining ratings in the raw score domain using a ratio/averaging method. This step is only applied to sequences without TR or QS variations. We make use of the fact that a video coded at a lower FR (or

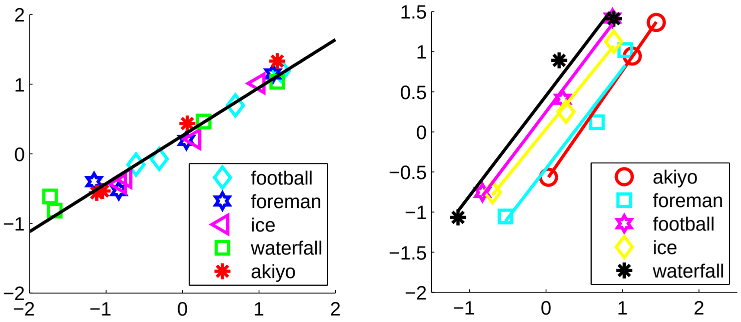


Figure 4.2: The mapping plot from test 1 MOS to test 2 MOS (in Z-domain). Left: For experiment 1, where we used a single mapping for all source sequences. Right: For experiment 2, where lines are the linear mapping function of each sequence.

higher QP) would not have a rating higher than a video coded at a higher FR (or lower QP), if the viewer's judgement is consistent. Therefore, we calculate the ratio of ratings by the same viewer for each pair of PVS's with adjacent QS or TR, respectively. For each source video and each viewer, we count the number of times that the ratio is greater than a threshold ($= 1.1$) for all possible pairs in each dimension, and then we remove all the ratings by a viewer for the same source video (including those PVSs with TR or QS variation) if the outlier counter in any dimension is larger than 2. For the remaining pairs of ratings by each viewer, if the ratio is larger than 1, we replace both ratings by their average. After this step, on average, 17 ratings remain for each PVS.

After the post-processing, we map all the Z-scores between two tests of each experiment using the method recommended in [22]. In experiment 1, we map from test 2 to test 1 as illustrated in Fig. 4.2 (left part). It shows that all the sequences can be fitted using the same linear function, while in experiment 2, each source sequence need their own mapping function as shown in Fig. 4.2 (right part). Finally, we scale the mapped Z-scores back to [0 10] scale.

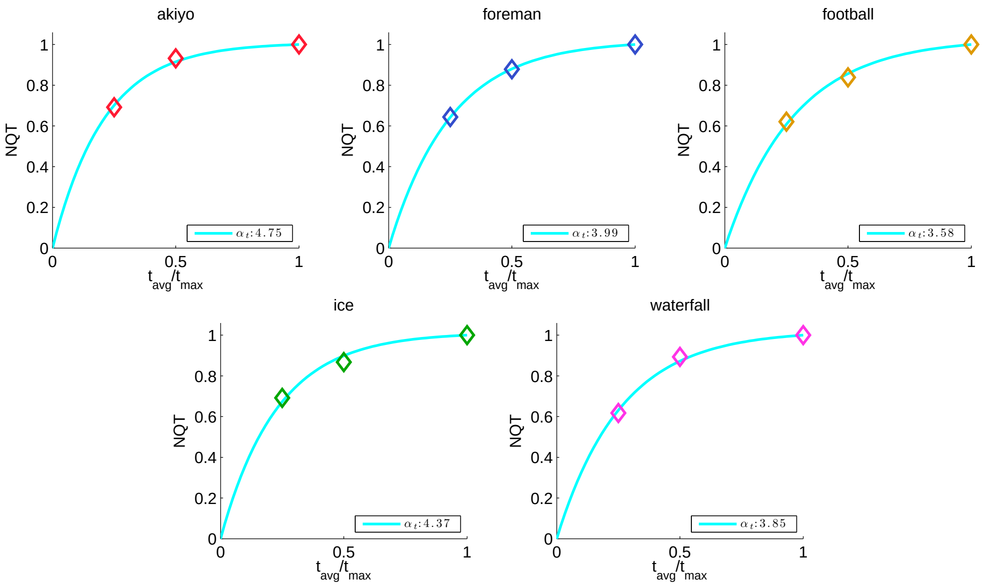


Figure 4.3: MOS vs. normalized FR ($t/t_{\max}$). Points are the measured MOS and curve are obtained using Eq. (4.2) with PCC=0.995, RMSE=0.013

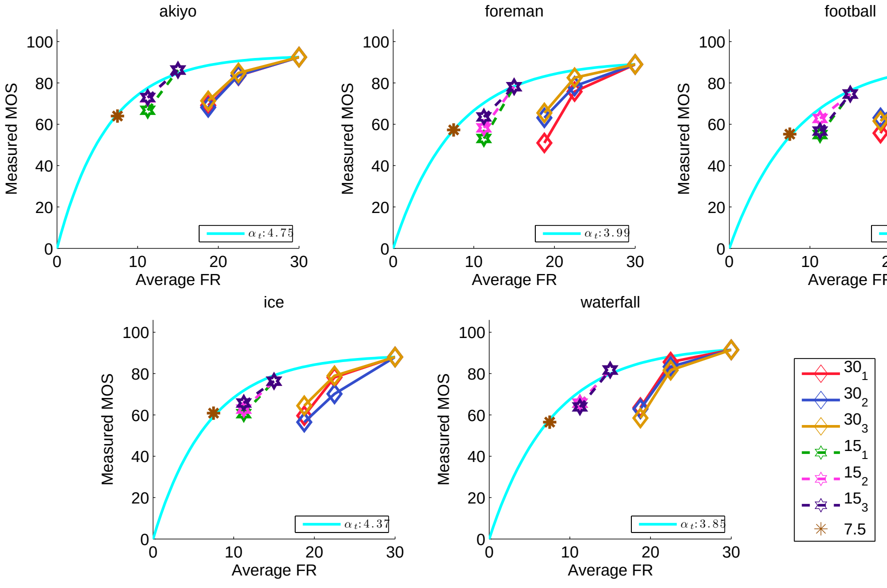


Figure 4.4: Q(t_h, t_l) vs. average FR. Points are the measured data of FR variations with different markers and colors corresponding to different t_l and t_h. Curves are predicted quality for sequences with constant FR by (4.2).

4.2 The Test Results of Frame Rate Variation

4.2.1 Impact of Constant Frame Rate

First we investigate the influence of the frame rate on the perceptual quality of a video with a constant frame rate, i.e. t_h=t_l $= t$. Let Q(t_h, t_l) = MOS(t_h, t_l)/MOS($t_{\max}, t_{\max}$)),

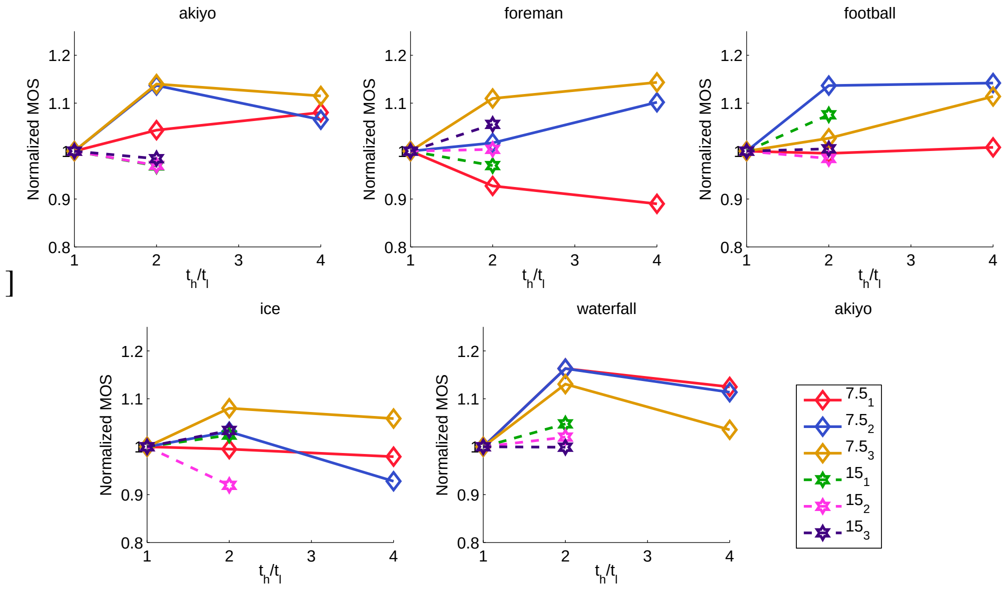


Figure 4.5: Q(t_h, t_l)/Q(t_l, t_l) vs. FR ratio (t_h/t_l) when t_l is fixed. Points are the measured data. Curves with different markers and colors are corresponding to different t_l and Fz

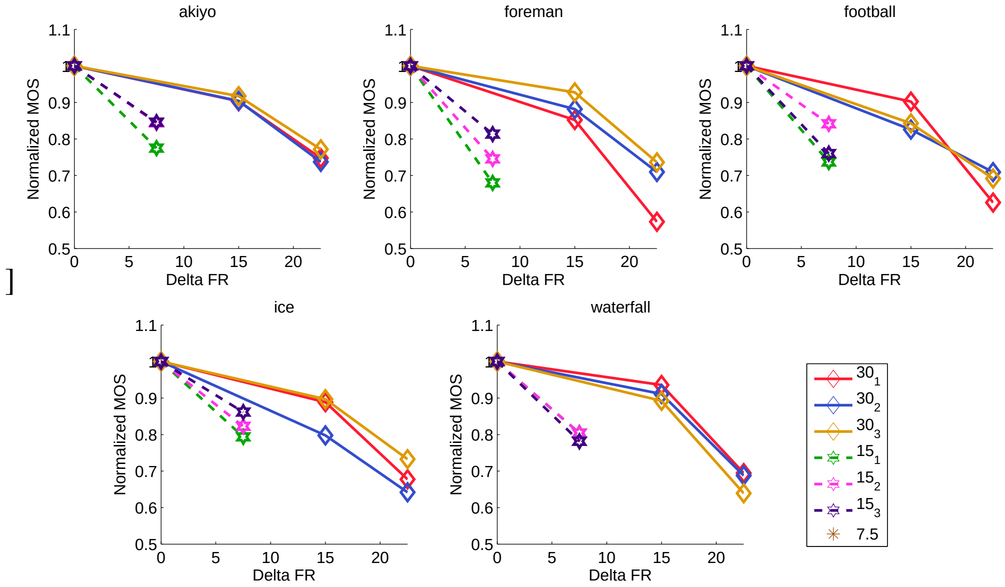


Figure 4.6: Q(t_h, t_l)/Q(t_h, t_h) vs. delta FR. Points are the measured data. Curves with different markers and colors are corresponding to different t_h and Fz

which is usually from 0 to 1. Figure 4.3 shows Q(t_l, t_h) vs. normalized FR, $\tilde{t} = \frac{t}{t_{\max}}$ (here $t_{\max}$=30), of all the testing sequences. For sequences with constant FR, as expected, the MOS reduces as the frame rate decreases. Based on the NQT model presented in Chapter 3 [56], we model the impact of constant frame rate (i.e., $t = t_h$=t_l) using

$$\mathrm{MNQT}_c(t) = \frac{1 - e^{-\alpha_t \cdot (\frac{t}{t_{\max}})^{\beta_t}}}{1 - e^{-\alpha_t}}. \tag{4.2}$$

Here it is found that β_t = 1 for all five source sequences gives the best fitting. As can be seen from Fig. 4.3, this model fits with measured data quite well with PCC=0.995, RMSE=0.013. Note that the parameter α_t characterizes how fast the quality drops as the frame rate reduces with smaller value corresponding to faster dropping rate. We further apply the model 4.2 using β_t = 0.63 (as introduced in chapter 3), the fitting is also very well with PCC=0.987, RMSE=0.025. In the following section, we will compare the quality of constant FR and FR variation and we will also investigate how to model the overall quality of a video considering both effects as due to these two different temporal artifacts.

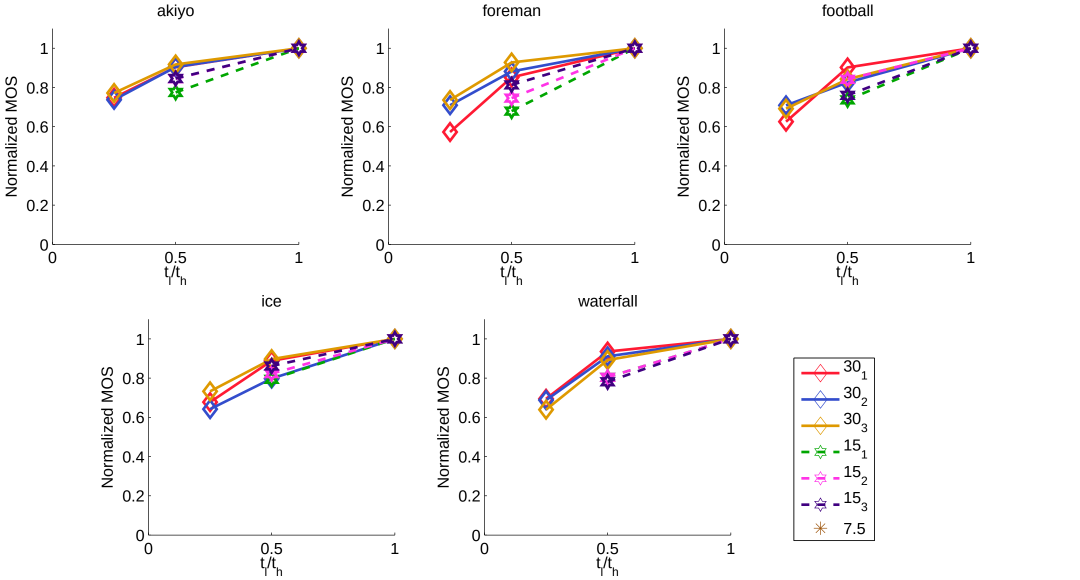


Figure 4.7: Q(t_h, t_l)/Q(t_h, t_h) vs. FR ratio (t_l/t_h). Points are the measured data. Curves with different markers and colors are corresponding to different t_h and Fz

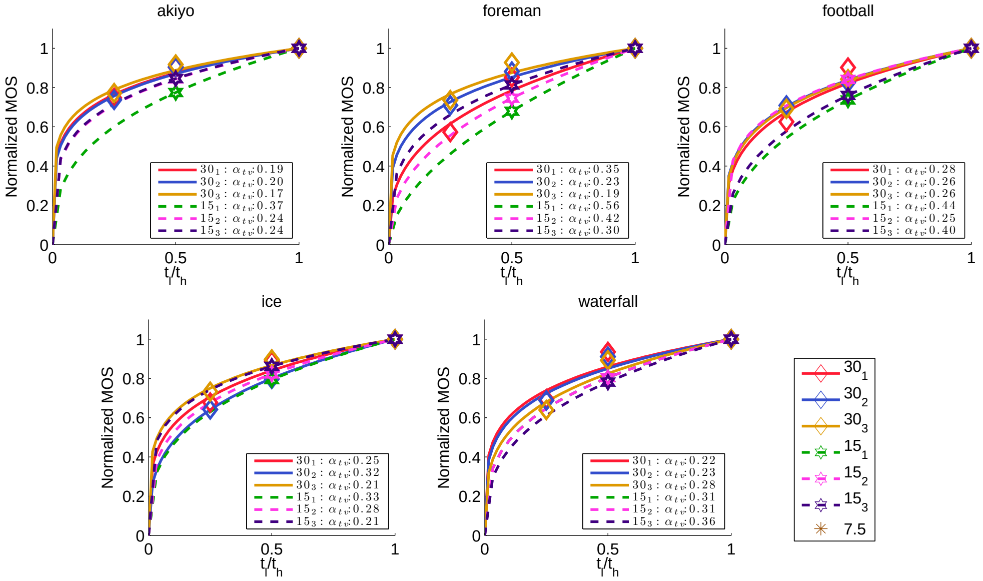


Figure 4.8: Q$(t_h, t_l)/$Q(t_h, t_h) vs. FR ratio (t_l/t_h) by a power law model given in Eq. (4.3) with PCC = 0.980, RMSE = 0.025. Different model parameters are used for different t_h and Fz.

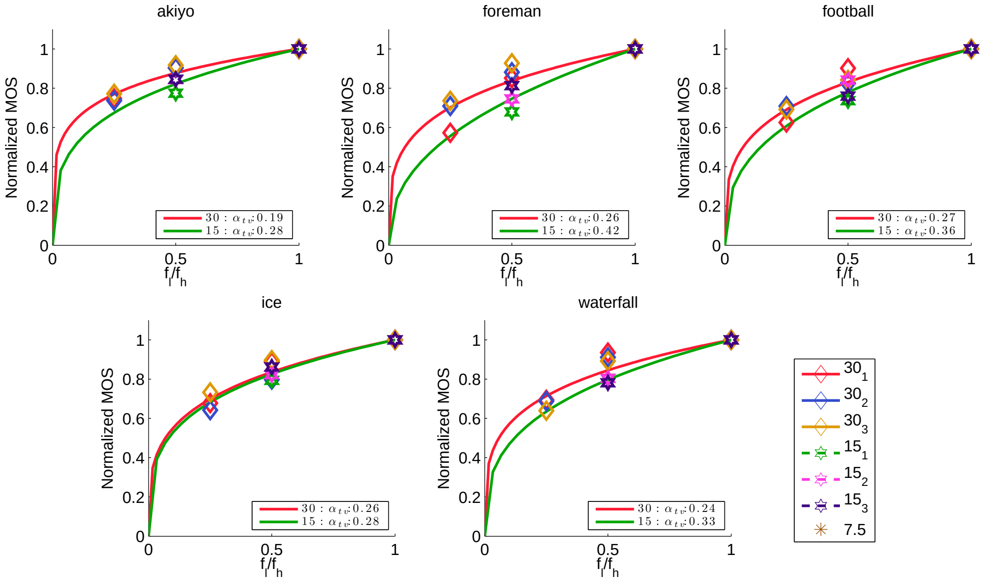


Figure 4.9: Q$(t_h, t_l)/$Q(t_h, t_h) vs. FR ratio (t_l/t_h) by a power law model given in Eq. (4.3) with PCC = 0.958, RMSE = 0.036. Different model parameters are used for different t_h and Fz.

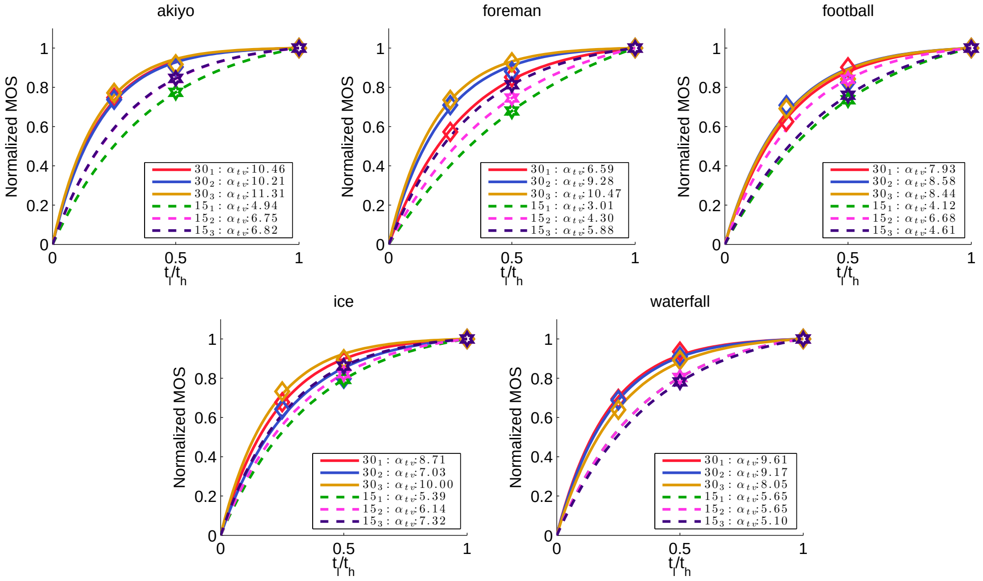


Figure 4.10: Q(t_h, t_l)/Q(t_h, t_h) vs. FR ratio (t_l/t_h) by an exponential model given in Eq. (4.4) with PCC = 0.991, RMSE = 0.016. Points are measured MOS and curves with same marker are the same t_h.

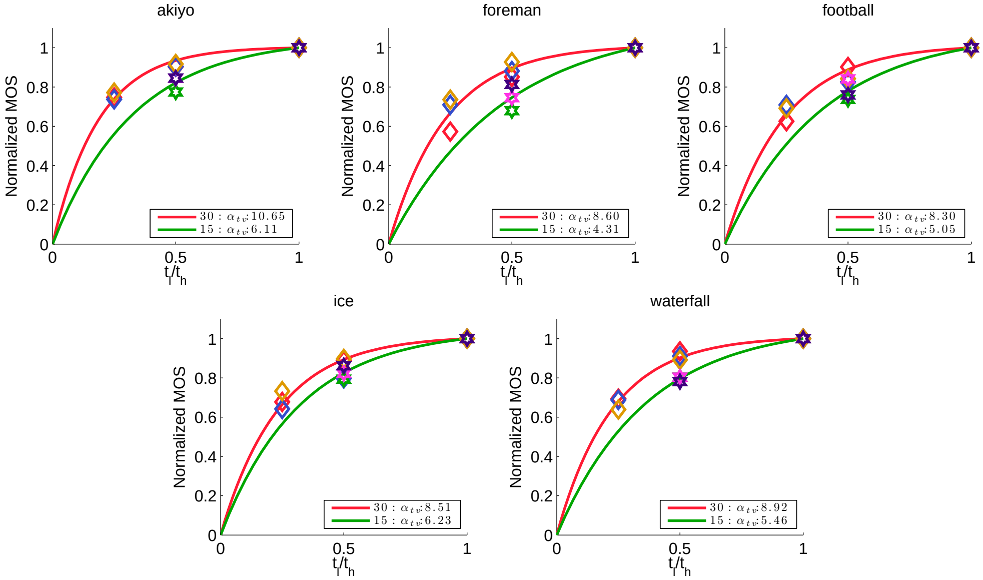


Figure 4.11: Q(t_h, t_l)/Q(t_h, t_h) vs. FR ratio (t_l/t_h) by an exponential model given in Eq. (4.4) with PCC = 0.971, RMSE = 0.029. Each model parameter is used for same t_h but all different Fz's.

4.2.2 Impact of Frame Rate Variation using FR

We now consider sequences in which the frame rate alternates between t_h and t_l. We first discuss, under the same average frame rate, t_{avg} = (t_h+t_l)/2, how does frame rate variation affects the perceived quality. Figure 4.4 shows that, when the t_{avg} is the same, the MOS for a video with a constant frame rate is higher than that with frame rate variation. Note that the cyan solid lines in Fig. 4.4 are predicted quality of constant-frame-rate videos based on Eq. (4.2). The degradation due to frame rate change is more severe when Δt is higher (e.g., MOS difference between (30,7.5) and constant frame rate of (30+7.5)/2=18.75 is greater than MOS difference between (30,15) and constant frame rate of (30+15)/2=22.5.). This result is as expected, as large frame rate variation induces noticeable jitter. It is interesting to note that points corresponding to t_l=t_h/2 are quite close to the operational quality-frame rate curves achievable by using constant frame rates, for most of the sequences. But those with t_l lower than t_h/2 are much below the curve.

We next look at when t_l is fixed due to the lowest available bandwidth, whether alternating between t_l and t_h leads to better quality than staying at t_l, when the available bandwidth fluctuates between the lowest bandwidth and a high bandwidth. This is the question we raised in the introduction as a motivation for this study. Let Q(t_h, t_l)/Q(t_l, t_l) denote as NQT_v, we plot NQT_v against FR ratio (t_h/t_l) in Fig. 4.5, where it shows that alternating between t_l and t_h is generally better than staying at t_l when $t_h/t_l \leq 2$. The slope of improvement reduces as t_l increases, and the degree of improvement is inconsistent (e.g., Football and Waterfall have higher slope in some Fz cases). When t_h is more than double of t_l, the quality improvement is also inconsistent. However, the quality improvement become saturated as $t_h/t_l > 2$. This suggests that when the lowest bandwidth limits the lowest frame rate to t_l, even when available bandwidth at a later time allows a frame rate beyond $2t_l$, it may be better to limit the highest frame rate to $t_h = 2t_l$. Note that there is no significant effect of Fz's from the observation of the results.

Now, if we fix t_h, how the quality changes with different t_l. Figure 4.6 demonstrates the Q(t_h, t_l)$/$Q(t_h, t_h) against Δt to study the impact of the strength of frame rate variation. We observe that higher t_h has slower dropping trend than lower t_h along the Δt trajectory.

This again implies that when t_h is already low, further inducing frame rate variation leads to more visual distortion, than when t_h is higher.

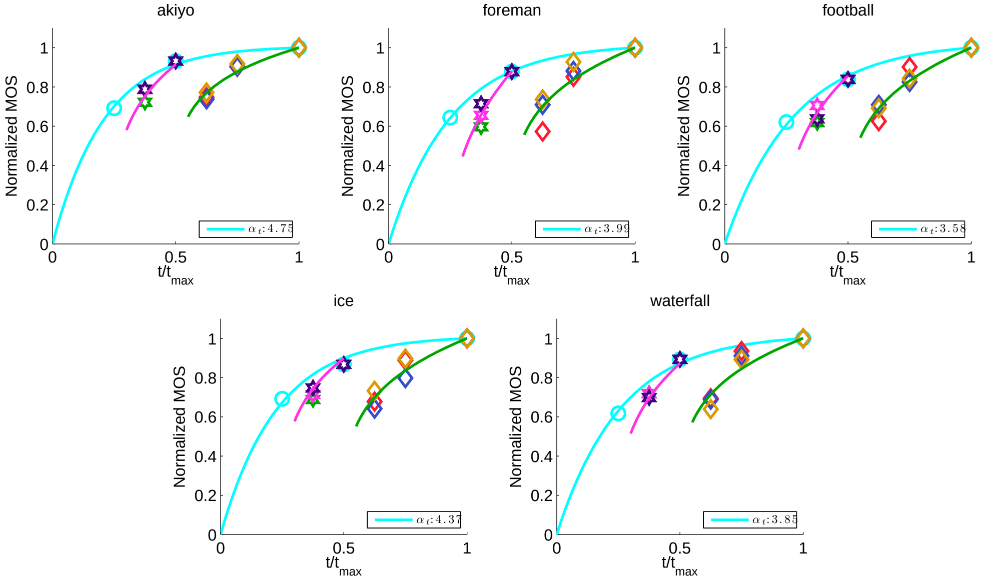


Figure 4.12: Q(t_h, t_l) vs. $t_{avg}/t_{\max}$ using Eq. (4.3) as $\overline{Q}_v$ in (4.5), with PCC = 0.960, RMSE = 0.035. Points are measured MOS and curves with same marker are the same t_h.

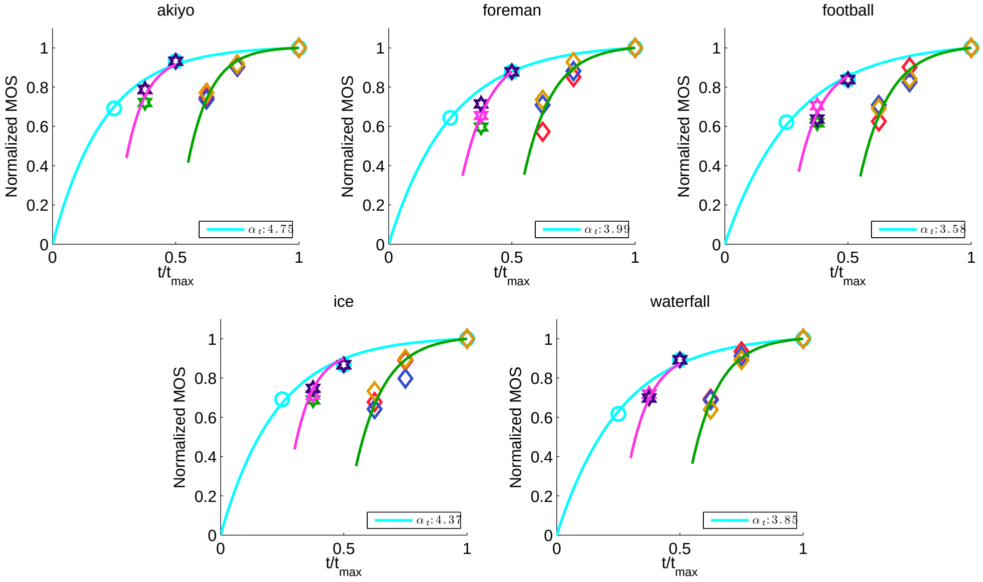


Figure 4.13: Q(t_h, t_l) vs. $t_{avg}/t_{\max}$ using Eq. (4.4) as $\overline{Q}_v$ in (4.5), with PCC = 0.978, RMSE = 0.027. Each model parameter is used for the same t_h but different Fz's.

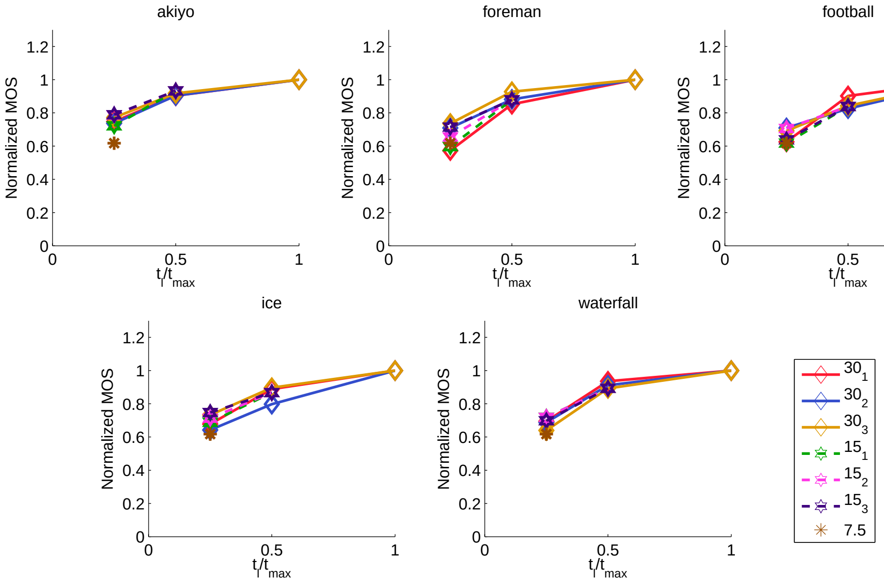


Figure 4.14: Q(t_h, t_l) vs. $t_l/t_{\max}$. Points are the measured MOS. Curves with different markers and colors corresponding to different t_h and Fz.

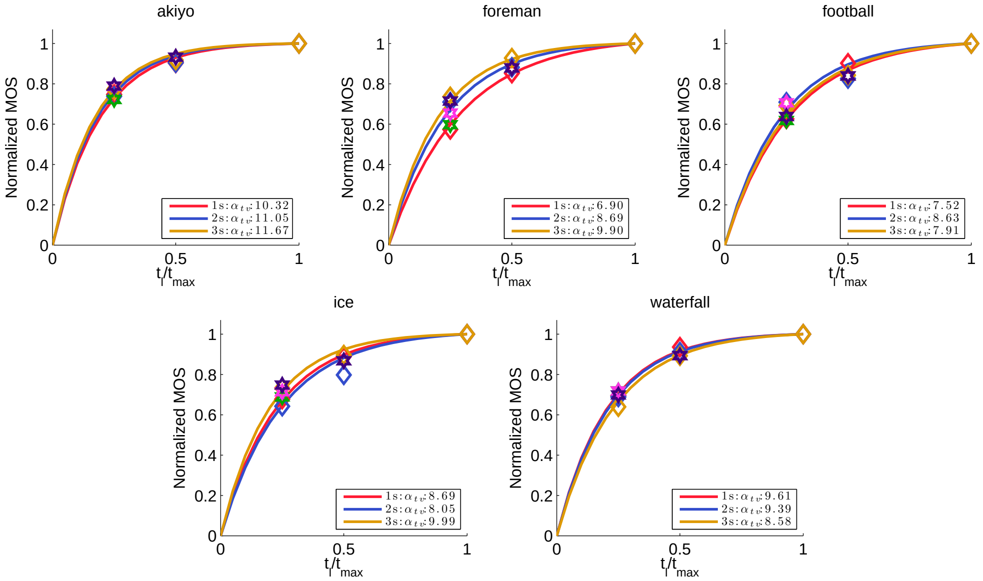


Figure 4.15: Q(t_h, t_l) vs. normalized FR ($t_l/t_{\max}$) using an exponential model given in Eq. (4.7) with PCC = 0.980, RMSE = 0.026. Points are measured MOS and three model curves are used for all f_h and Fz.

Instead of measuring the frame rate variation strength by the difference in frame rate, Δt, Fig. 4.7 uses the frame rate ratio, t_l/t_h. Figure 4.7 shows how does the normalized

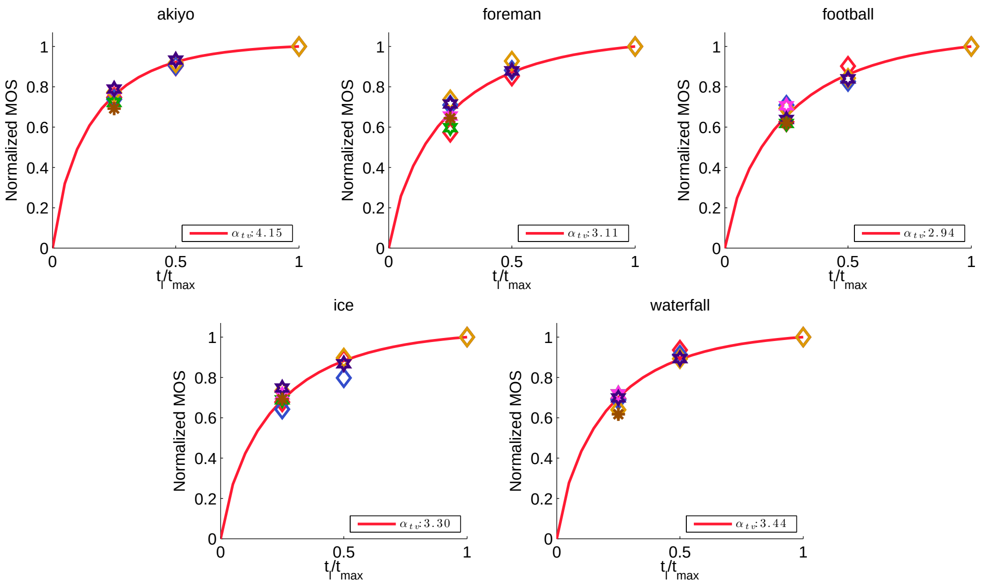


Figure 4.16: Q(t_h, t_l) vs. $t_l/t_{\max}$ using an exponential model given in Eq. (4.7) with PCC = 0.966, RMSE = 0.033. One model curve is used for all f_h and Fz.

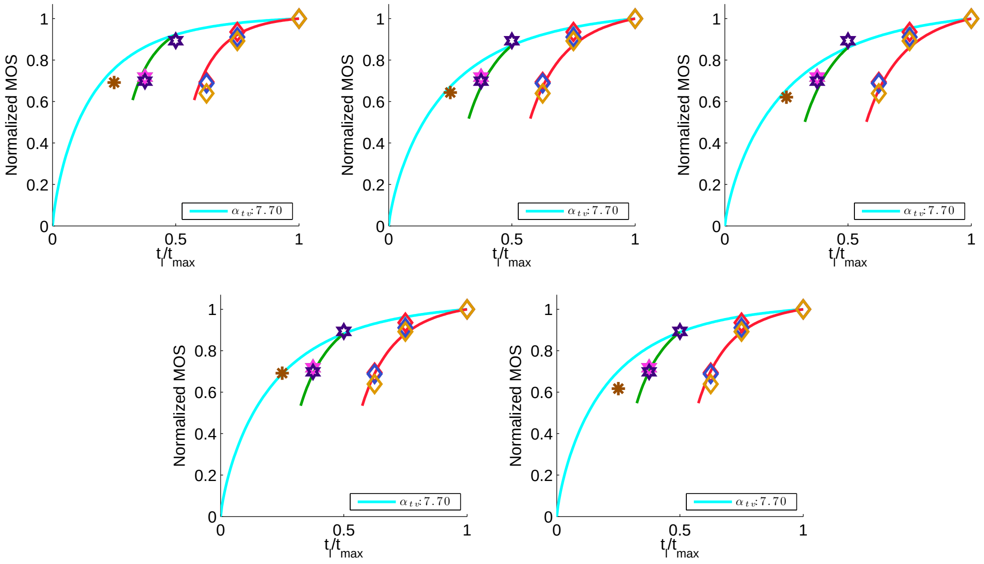


Figure 4.17: Q(t_h, t_l) vs. $t_{avg}/t_{\max}$ using Eq. (4.7) with PCC = 0.966, RMSE = 0.033. Points are measured data.

MOS decrease with t_l/t_h. It is interesting that the dropping trends under different t_h are similar. This implies that the impact of frame rate variation can be well captured by the frame rate ratio, independent of t_h. We found that the variation of quality with the frame rate ratio can be modeled using a power law function, i.e.,

$$\mathrm{MNQT}_v(t_h, t_l) = (t_l/t_h)^{-\alpha_{tv}(\mathrm{Fz}, t_h)} \tag{4.3}$$

where α_{tv} is the model parameter. Fig. 4.8 shows both the measured quality (with points) as well as the predicted one (with curves) using this simple model when α_{tv} depends on both Fz and t_h (with a total of 6 parameters for each sequence to account for different t_l, t_h, and Fz combinations). Besides the power law function, we also examine the accuracy of the inverse exponential function, i.e.,

$$\mathrm{MNQT}_v(t_h, t_l) = \frac{1 - e^{-\alpha_{tv}(\mathrm{Fz}, t_h) \cdot (t_l/t_h)^{\beta_{tv}}}}{1 - e^{-\alpha_{tv}(\mathrm{Fz}, t_h)}}. \tag{4.4}$$

Where the parameter α_{tv} depends on both Fz and t_h. In order to further analyze the relationship between model parameters and model performance, we first conduct the statistical test on normalized MOS Q(t_h, t_l) using two-way repetition ANOVA test. Particularly, we only emphasize the statistical significance for each pair of Fz with the interaction to Δt. As shown in Tab. 4.5 there is no significant difference for all the cases, and the dropping trend of Fz=1,2 and 3 are mostly indistinguishable (p-value $>$ 0.05) for a given t_h. It is note that within each t_h, the dropping trends are prone to cluster together but separable between one t_h to another t_h. For better understanding and quantifying the performance of two models in (4.3,4.4), we compare with all the models and parameter dependency in Tab. 4.3. We see that the model in Eq. (4.4) is slightly more accurate than the mode in (4.3) under the same number of parameters, and the model in (4.4) with α_{tv} depending on t_h only gives a good trade off between accuracy and model complexity.

In order to predict the overall quality for a known pair of t_h and t_l, we propose to use the product of two models, $\mathrm{MNQT}_c(t_h)$ and $\mathrm{MNQT}_v(t_l, t_h)$, i.e.,

$$\mathrm{QTV}(t_h, t_l) = \mathrm{MNQT}_c(t_h)\mathrm{MNQT}_v(t_l, t_h). \tag{4.5}$$

Based on this function form and specific order in Eq. (4.5), the first term is designed to predict the quality at constant FR (referred to t_h), and the second term estimates the degradation of the quality as FR fluctuates between t_h and t_l. Substitute Eq. (4.2) and (4.4) into

(4.5), i.e.,

$$\mathrm{QTV}_1(t_h, t_l) = \frac{1 - e^{-\alpha_t \cdot (\frac{t}{t_{\max}})^{\beta_t}}}{1 - e^{-\alpha_t}} \frac{1 - e^{-\alpha_{tv}(t_h) \cdot (\frac{t_l}{t_h})^{\beta_{tv}}}}{1 - e^{-\alpha_{tv}}}, \tag{4.6}$$

where α_{tv} only depends on t_h and independent of Fz, $\beta_{tv} = \beta_t = 1$. Figure. 4.13 illustrates that the predicted Q(t_h, t_l) v.s. $t_{avg}/t_{\max}$ fits the measured MOS quite well with PCC=0.978, RMSE= 0.027. Here, the model in (4.6) only needs three parameters, since β_{tv} and β_t both are constants.

Besides studying the quality impact directly from FR variation strength of t_h and t_l, we plot the measured Q(t_h, t_l) against $t_l/t_{\max}$ in Fig. 4.14. Note that we normalize the MOS by the MOS at $t_{\max}$ instead of the quality at t_h so that the quality effect is independent of t_h. It is interesting to observe that the dropping trends under different t_h and Fz's are clustered together. This implies that the human eyes are mostly dominated by the effect of t_l in FR variation and the overall quality can be well approximated with an inverse exponential function of $t_l/t_{\min}$ without knowing the effect of t_h, i.e.,

$$\mathrm{QTV}_2(t_l, t_h) = \frac{1 - e^{-\alpha_{tv} \cdot (\frac{t_l}{t_{\max}})^{\beta_{tv}}}}{1 - e^{-\alpha_{tv}}}, \tag{4.7}$$

where β_{tv} is a constant value of 0.8 and α_{tv} is a model parameter, characterizing the dropping trend of the curve with smaller value corresponding to higher dropping trend. As shown in Fig. 4.16, we use one curve to fit all the measured MOS with PCC=0.970. Although the PCC is not as high as the model in (4.6), the advantage of this simple model (4.7) is that it requires only one content-dependent parameter, whereas the model in (4.6) requires three.

Table 4.3: The model performance of $\mathrm{MNQT}_v(t_h, t_l)$

Model Function	Assumption	# par	PCC	RMSE
$\frac{1-e^{-\alpha_{tv}\cdot(\frac{t_l}{t_h})^{\beta_{tv}}}}{1-e^{-\alpha_{tv}}}$	$\alpha_{tv}\succ t_h$, Fz; $\beta_{tv}\succ t_h$	8	0.991	0.010
	$\alpha_{tv}\succ t_h$, Fz	6	0.981	0.016
	$\alpha_{tv}\succ t_h$	2	0.971	0.029
$\left(\frac{t_l}{t_h}\right)^{-\alpha_{tv}}$	$\alpha_{tv}\succ t_h$, Fz	6	0.980	0.025
	$\alpha_{tv}\succ t_h$	2	0.958	0.036

$\succ$:Depends on

Table 4.4: The model performance of $\overline{Q}_{rv}(r_{th}, r_{tl})$

Function	Assumption	# par	PCC	RMSE
$\frac{1-e^{-\alpha_{rv}\cdot(\frac{r_{tl}}{r_{th}})^{\beta_{rv}}}}{1-e^{-\alpha_{rv}}}$	$\alpha_{rv}\succ r_{th}$, Fz ; $\beta_{rv}\succ r_{th}$	8	0.990	0.002
	$\alpha_{rv}\succ t_h$, Fz	6	0.990	0.017
	$\alpha_{rv}\succ t_h$	2	0.968	0.032

$\succ$:Depends on

Table 4.5: The ANOVA test for changing interval:FR variation

Target pair	Factors	F-value	p-value
when t_h = 30Hz			
Fz = 1 and 2	Δt*Fz	1.76	0.19
Fz = 1 and 3	Δt*Fz	1.44	0.25
Fz = 2 and 3	Δt*Fz	0.49	0.61
Fz =1, 2, 3	Δt*Fz	1.28	0.29
when t_h = 15Hz			
Fz = 1 and 2	Δt*Fz	3.43	0.08
Fz = 1 and 3	Δt*Fz	3.41	0.08
Fz = 2 and 3	Δt*Fz	0.001	0.96
Fz =1, 2, 3	Δt*Fz	2.44	0.10

4.2.3 Impact of Bit Rate Variation Due to Frame Rate Variation

We also compare quality of the videos with different frame rate variations under the same average BR in Fig. 4.8. It is clear that, a constant frame rate video has a better quality than a video with frame rate variation (especially when the frame rate variation is large), under the same average bit rate. It is a very useful observation when we try to optimized the perceptual quality of the video bitstream under fluctuate rate constraint.

Although the Eq. 4.6 can predict the impact of quality on FR variations for a given pair of t_l and t_h, it may not be applicable if the FR information is not available. It would be useful to predict the perceptual quality against BR variation if the BR information is accessible. Let r_{tl} and r_{th} denote the BR at t_l and t_h, respectively. We first plot the measure data, defined as Q(r_{th}, r_{tl}) = (MOS(r_{th}, r_{tl})/MOS($r_{\max}, r_{\max}$)), with measured data at fixed FR in terms of BR (When the bit rate is constant, i.e., $r_{tl} = r_{th}$) in Fig. 4.18. We found that the quality degradation can be approximated well using the exponential function for given

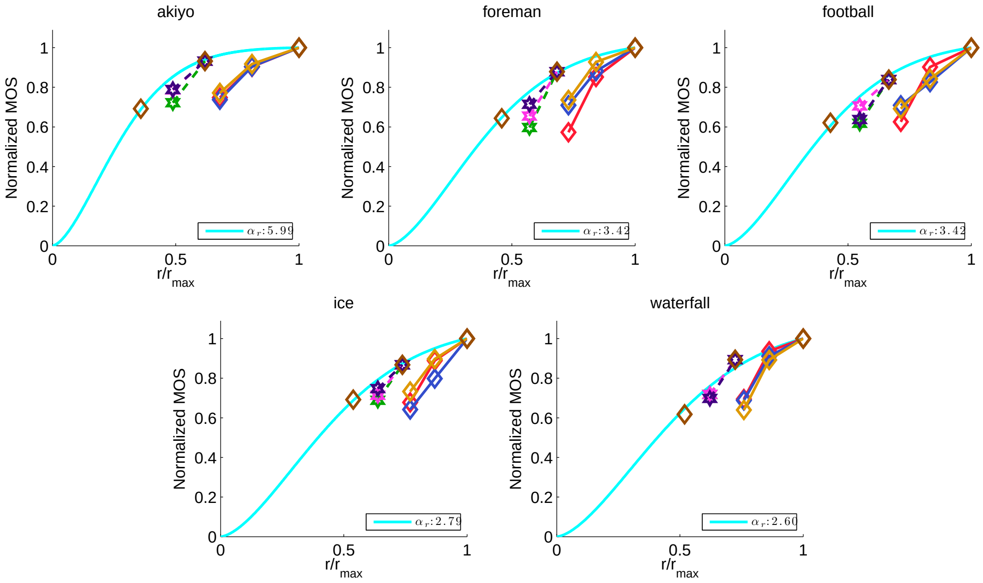


Figure 4.18: Q(r_{th}, r_{tl}) vs. $r_{avg}/r_{\max}$. Points are the measured data and curves with different markers and colors corresponding to different t_h and Fz. The fitting curves are predicted quality obtained from (4.8) with PCC = 0.992, RMSE = 0.035.

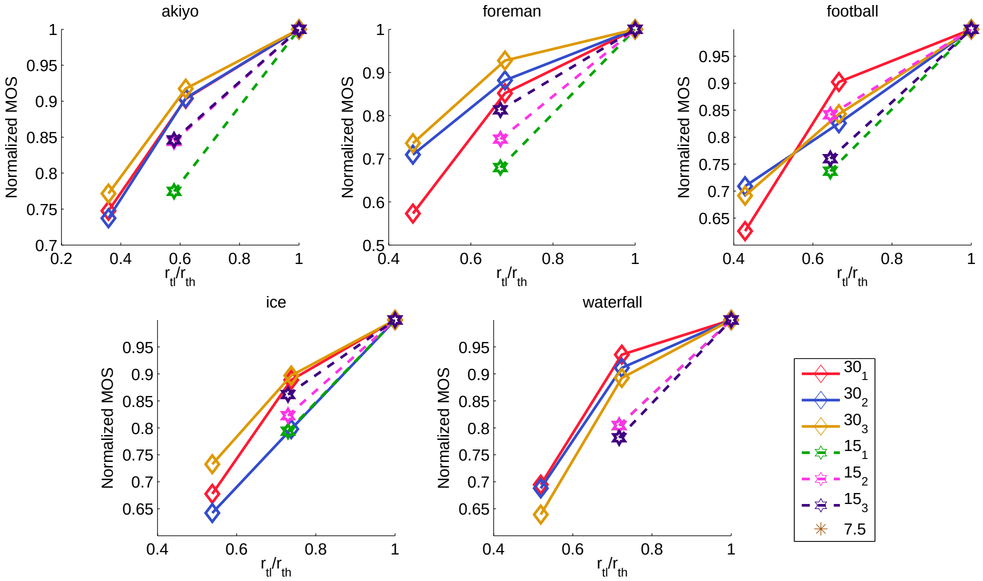


Figure 4.19: Q(r_{th}, r_{tl})/Q(r_{th}, r_{th}) vs. r_{tl}/r_{th}. Points are the measured data. Curves with different markers and colors corresponding to different t_h and Fz.

spatial resolution s and QS q, i.e.,

$$\text{MNQT}_{cr}(r; s, q) = \frac{1 - e^{-\alpha_r (\frac{r}{r_{\max}})^{\beta_r}}}{1 - e^{-\alpha_r}}, \tag{4.8}$$

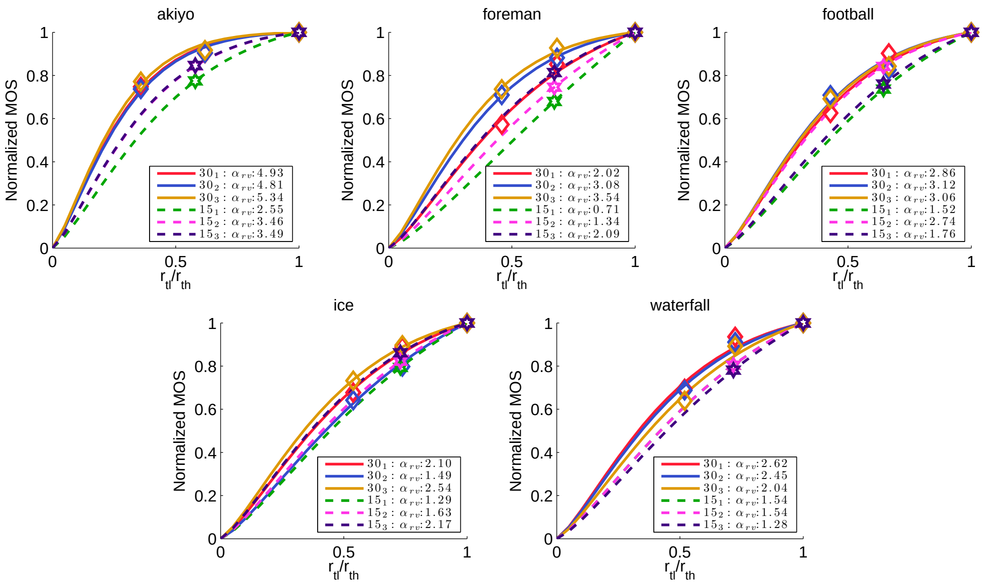


Figure 4.20: Q(r_{th}, r_{tl})/Q(r_{th}, r_{th}) vs. r_{tl}/r_{th} using an exponential model given in Eq. (4.9) with PCC = 0.990, RMSE = 0.017. Points are measured MOS and curves with same marker are the same t_h.

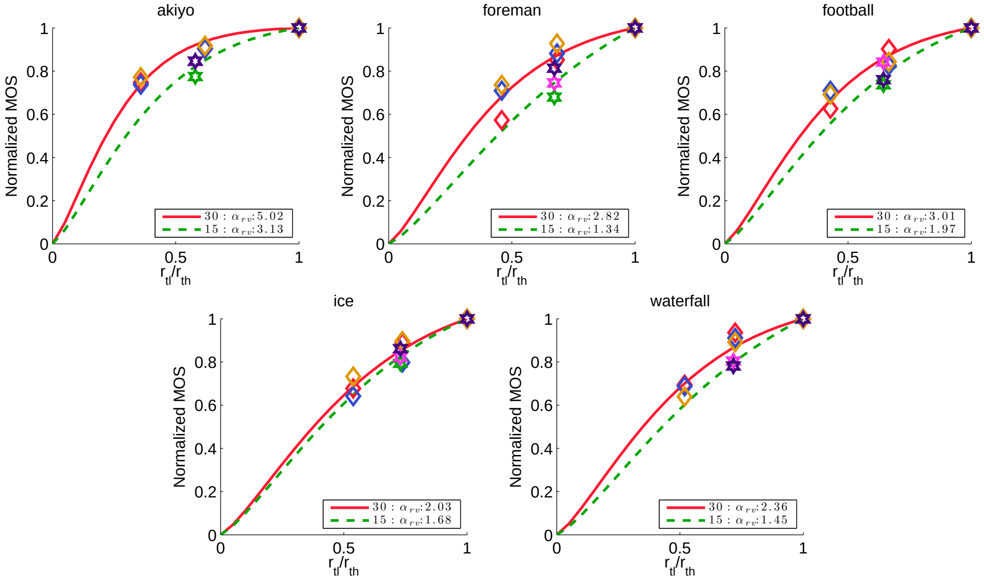


Figure 4.21: Q(r_{th}, r_{tl})/Q(r_{th}, r_{th}) vs. BR ratio (r_{tl}/r_{th}) by an exponential model given in Eq. (4.9) with PCC = 0.968, RMSE = 0.032. Points are measured data. Each model parameter is used for the same t_h but different Fz's.

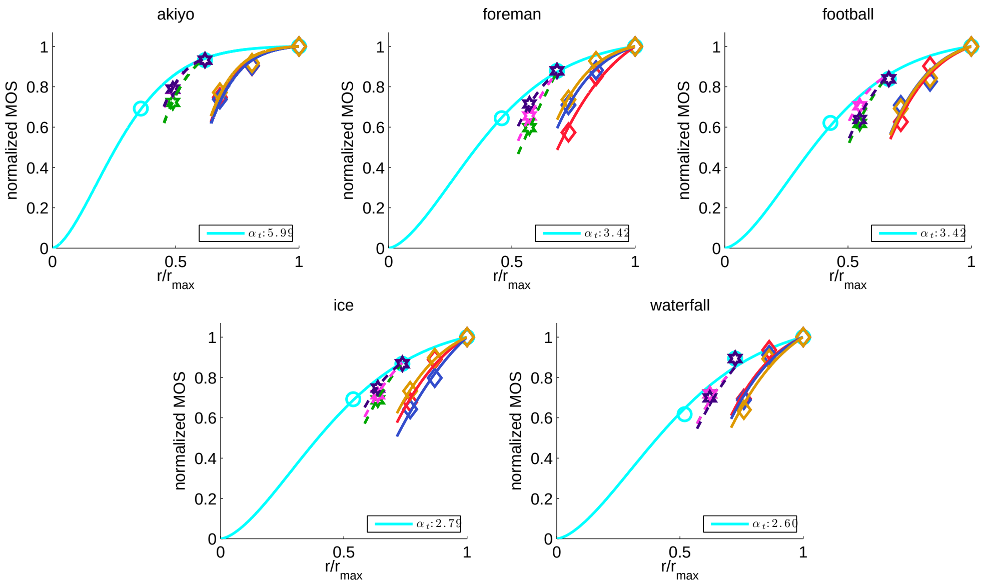


Figure 4.22: Q(r_{th}, r_{tl})/Q(r_{th}, r_{th}) vs. normalized BR ($r_{avg}/r_{\max}$) for high FR base using Eq. (4.11) with PCC = 0.985, RMSE = 0.021. Points are measured MOS and curves with same marker are the same t_h.

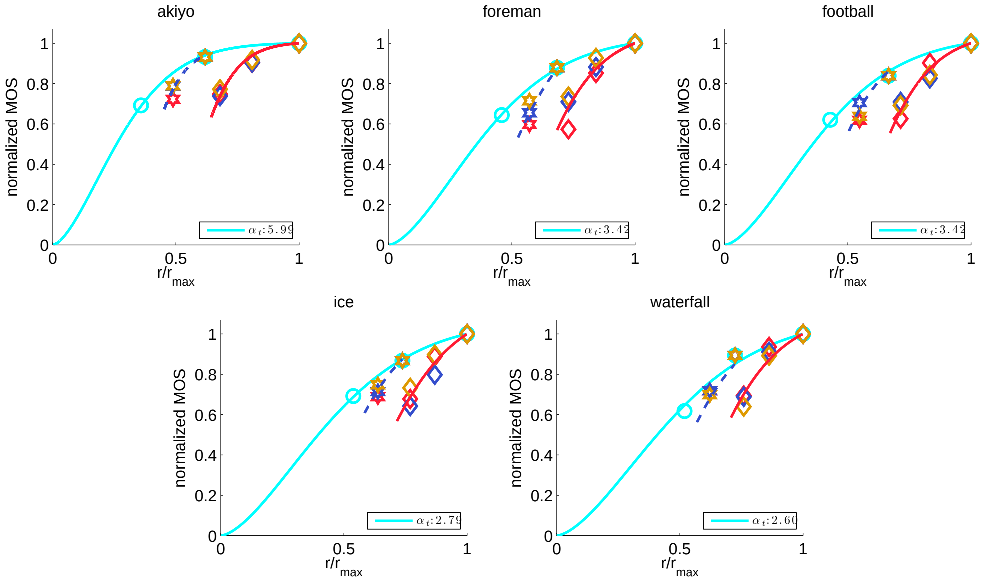


Figure 4.23: Q(r_{th}, r_{tl}) vs. normalized BR ($r_{avg}/r_{\max}$) by using Eq. (4.11) with PCC = 0.965, RMSE = 0.032. Points are measured data.

where r represents the BR, and r_{max} is the bit rate at $t_{\max} = 30$. The model parameter α_r control the falling rate of the curve and we found that β_r is a constant value of 1.6 for

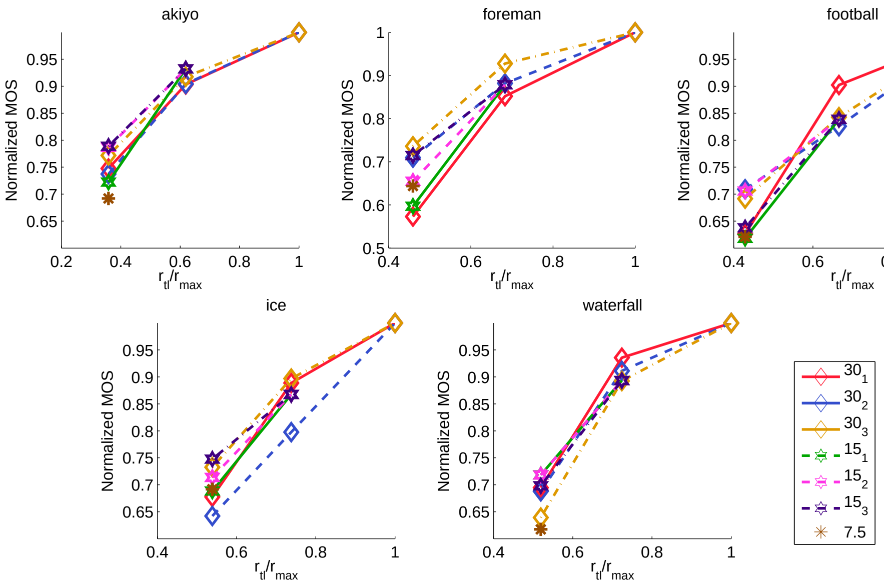


Figure 4.24: Q(r_{th}, r_{tl}) vs. $r_{tl}/r_{\max}$. Curves with different markers and colors are corresponding to different t_h and Fz.

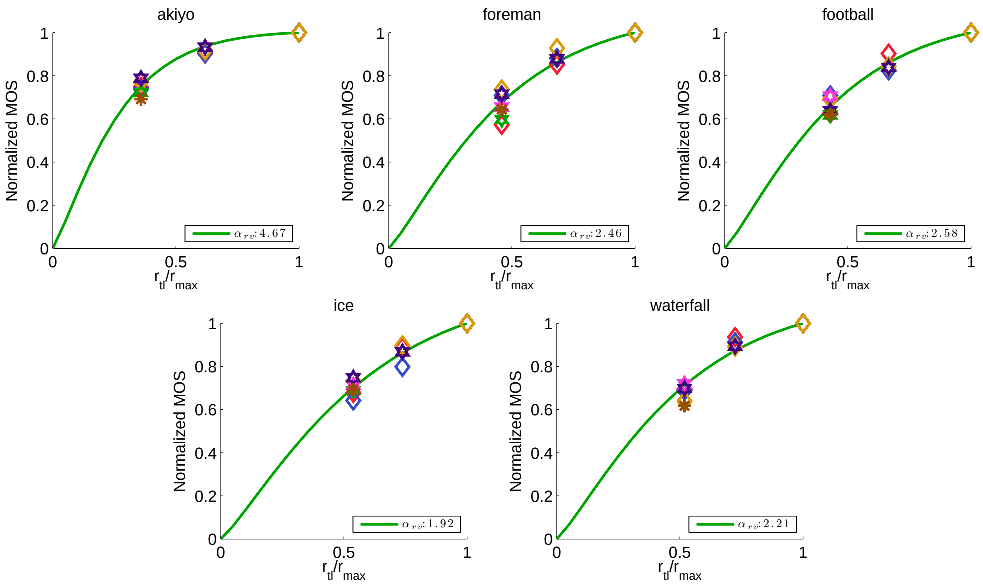


Figure 4.25: Q(r_{th}, r_{tl}) vs. $r_{tl}/r_{\max}$ by an exponential model given in Eq. (4.22) with PCC = 0.966, RMSE = 0.032. Points are measured MOS and curves with same marker are the same t_h.

all the source sequences. Figure 4.18 shows the predicted curves fit measured MOS quite accurate with PCC=0.992.

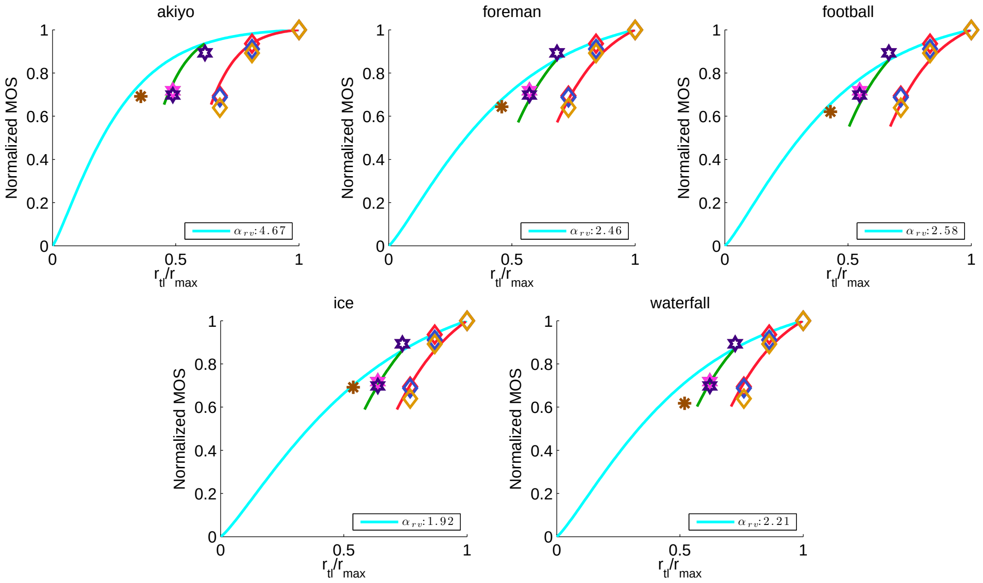


Figure 4.26: Q(r_{th}, r_{tl}) vs. $r_{avg}/r_{\max}$ for high FR base using Eq. (4.22) with PCC = 0.966, RMSE = 0.032. Points are measured data.

Besides the quality impact of constant BR, we next move to the BR variation due to the FR changing. We plot Q(r_{th}, r_{tl})/Q(r_{th}, r_{th}) against r_{tl}/r_{th} in Fig. 4.19. Similar to (4.4), we also propose to model the Q(r_{th}, r_{tl})/Q(r_{th}, r_{th}) in Fig. 4.19 by

$$\text{MNQT}_{rv}(r_{tl}, r_{th}) = \frac{1 - e^{-\alpha_{rv} \cdot (\frac{r_{tl}}{r_{th}})^{\beta_{rv}}}}{1 - e^{-\alpha_{rv}}}, \tag{4.9}$$

where α_{rv} is model parameter that may depend on r_{th}, Fz, and β_{rv} is a constant value of 1.3 for all five source sequences. Figure 4.21 demonstrates the fitting curves with measure data using (4.9), where we plot 2 fitting curves for each source sequence with high PCC = 0.97. Note that in Fig. 4.20 we also fit the measure data using 6 different curves corresponding to different t_h and Fz with PCC = 0.99, but it requires too many model parameters. Due to this manner, the tradeoff between number of parameters and model accuracy is a crucial issue to be investigated. Therefore, we conduct a comparison of the model performance in Tab. 4.4 regarding the dependency of parameter α_{rv} with Fz and r_{th}. After considering the number of parameter and accuracy, we conclude that the model in (4.9) with two parameters only can (varying with t_h, and independent of Fz) predict measured data very well with PCC=0.968, RMSE=0.032.

Finally, for a given pair of r_{th} and r_{tl}, we can estimate the overall quality by the product of constant BR and BR variation model, i.e.,

$$\mathrm{QTVBR}(r_{tl}, r_{th}) = \mathrm{MNQT}_{cr}(r; s, q)\mathrm{MNQT}_{rv}(r_{tl}, r_{th}), \tag{4.10}$$

Substituting $\mathrm{MNQT}_c r(r; s, q)$ and $\mathrm{MNQT}_r v(r_{tl}, r_{th})$ by (Eqs. (4.8) and (4.9), respectively, and the new form of (4.10) can be written as

$$\mathrm{QTVBR}_1(r_{tl}, r_{th}) = \frac{1 - e^{-\alpha_r(\frac{r}{r_{\max}})^{\beta_r}}}{1 - e^{-\alpha_r}} \frac{1 - e^{-\alpha_{rv}\cdot(\frac{r_{tl}}{r_{th}})^{\beta_{rv}}}}{1 - e^{-\alpha_{rv}}}. \tag{4.11}$$

We plot Q(r_{th}, r_{tl}) v.s. r_{avg}/r_{max} (r_{avg}=($r_{tl} + r_{th}$)/2) in Fig. 4.23 and the predicted quality captures the measured quality well with PCC = 0.965, RMSE = 0.032.

Moreover, we plot Q(r_{th}, r_{tl}) against $r_{tl}/r_{\max}$ in Fig. 4.24. Note that we normalize the measure MOS by the MOS at $r_{\max}$ instead of the quality at r_{th} so that the falling trends is independent of r_{th}. It is interesting to observe that the dropping trends under different r_{tl} and Fz's are clustered together. This implies that the QS variation can be well approximated by the r_{tl} without knowing r_{th}. We found that the variation of quality with $r_{tl}/r_{\max}$ can be modeled quite well using an inverted exponential function, i.e.,

$$\mathrm{QTVBR}_2(r_{tl}) = \frac{1 - e^{-\alpha_{rv}(\frac{r_{tl}}{r_{\max}})^{\beta_{rv}}}}{1 - e^{-\alpha_{rv}}}, \tag{4.12}$$

where only one model parameter α_{rv} and a constant value of $\beta_{rv} = 1.2$ are needed to capture the quality variation with normalized QS and both these two parameters are independent of Fz and t_l but video content. Figure 4.25 shows the fitting curves are accurate with PCC=0.966, RMSE=0.032.

4.3 The Test Results of Quantization Variation

4.3.1 Impact of constant QS

First we investigate the influence of the QS on the perceptual quality of a video with a constant QS, i.e. $q = q_h = q_l$. Let Q(q_h, q_l) = MOS (MOS(q_h, q_l)/MOS($q_{\min}, q_{\min}$)) denote as the NQQ, which is usually from 0 to 1. Figure 4.27 shows Q(q_l, q_h) vs. normalized QS, $q_{\min}/q$ of all the testing sequences. As expected, the quality reduces as the QS increases. According to the MNQQ model presented in chapter 3, we model the impact of constant QS (i.e., $q = q_h = q_l$) by

$$\mathrm{MNQQ}_c(q) = \frac{1 - e^{-\alpha_q (\frac{q_{\min}}{q})^{\beta_q}}}{1 - e^{-\alpha_q}}, \tag{4.13}$$

where q represents the QS, $q_{\min}$ is the minimum QS ($q_{\min} = 16$). We also found that β_q is constant for all five sequences with a value of 1 (same as the MNQQ in chapter 3). As can be seen from Fig. 4.27, this model fits with measured data quite well. Note that the parameter α_q characterizes how fast the quality drops as the QS increases.

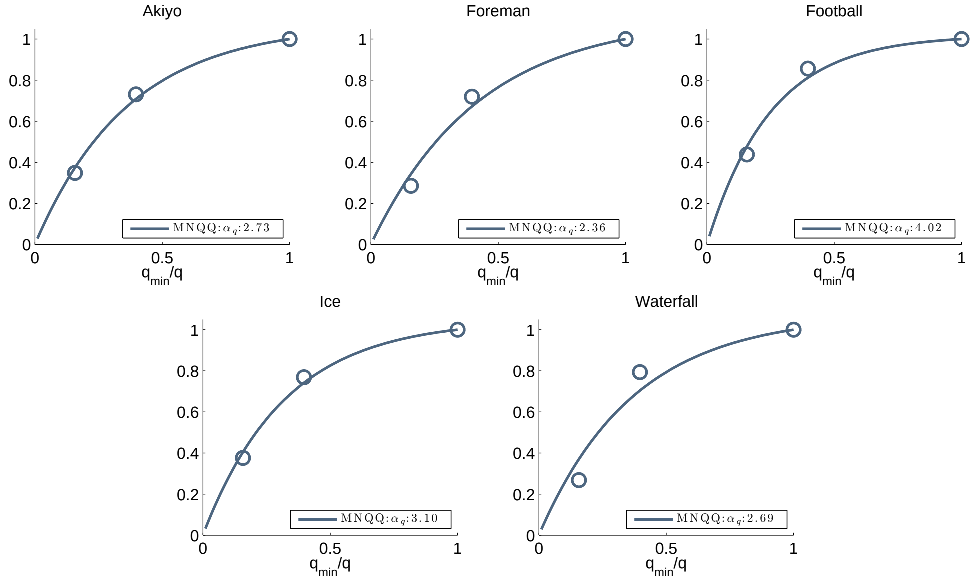


Figure 4.27: Q(q_h, q_l) vs. $q_{\min}/q$. Points are the measured data and curve are obtained using Eq. (4.13) with PCC=0.991, RMSE=0.037. The β_q = 1.

4.3.2 Impact of QS Variation

We now consider sequences in which the QS alternates between q_h and q_l. We first discuss, under the same average QS, q_{avg} = (q_h+q_l)/2, how does QS variation affects the perceived quality. Figure 4.28 shows that, when the q_{avg} is the same, the quality for a

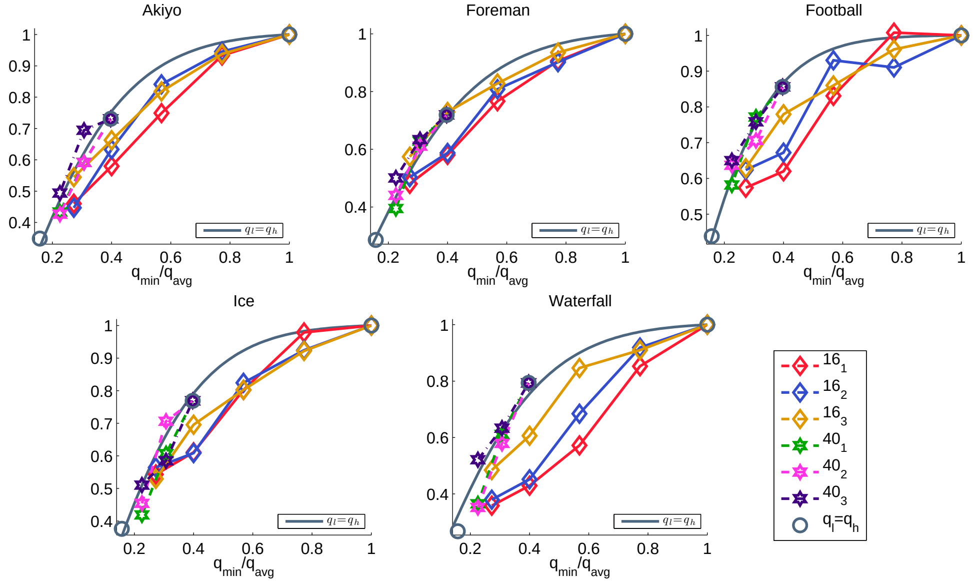


Figure 4.28: Q(q_l, q_h) vs. $q_{\min}/q_{avg}$. Each line connects points with the same q_l and Fz. For example, 16_3 indicates q_l=16, Fz=3. Note that lines for q_l=40 mostly overlap with the quality curve for the constant QS case.

video with a constant QS, MOS(q_{avg}, q_{avg}), is higher than that with QS variation, MOS(q_h, q_l). Note that the solid curve in Fig. 4.28 are predicted quality of constant-QS videos based on (4.13). Interestingly for q_l=16, the degradation due to QS change is more severe when Δq=q_h-q_l is in the middle range, i.e., Q(16,40) and (16, 64) deviate from their constant-QS counterparts more than Q(16,102) and Q(16,25), and that between Q(16, 25) and $\mathrm{MNQQ}_c(16)$, is larger than the difference between Q(16, 40) and Q(16, 102). In addition, the mean quality of MOS(q_l) and MOS(q_h) is higher than Q(q_l, q_h) as shown in Fig. 4.34.

We next examine when when q_h is fixed due to the lowest available bandwidth, whether alternating between q_l and q_h leads to better quality than staying at q_h, when the available bandwidth fluctuates between the lowest bandwidth and a high bandwidth. This is the question we raised in the introduction as a motivation for this study. We plot Q(q_l, q_h)/Q(q_h, q_h) against the ratio q_h/q_l in Fig. 4.29. For q_h=102, alternating between q_l and q_h, is consistently better than staying at q_h=102, and the degree of improvement depends on Fz and the texture details of the video (e.g., Waterfall and Foreman have higher improvement ratio,

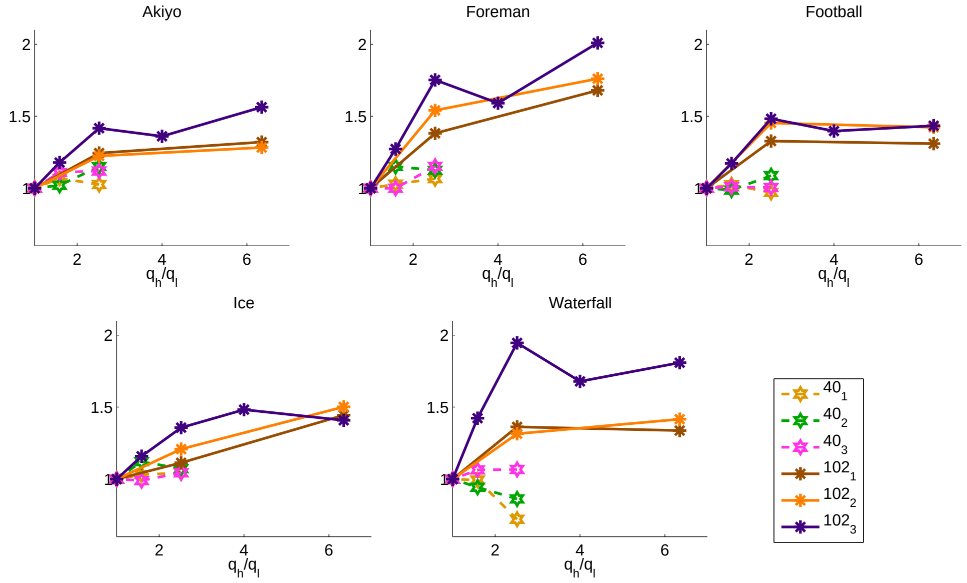


Figure 4.29: Q(q_h, q_l)/Q(q_h, q_h) vs. q_l/q_h when q_h is fixed. Points are measured data. Each line connects points with the same q_h and Fz. For example, 102_3 indicates q_h=102, Fz=3.

e.g., up to 2). It can be observed that shorter Fz leads to less improvement. This is as expected as shorter Fz corresponds to more rapid QS switching, which can be more annoying to the human viewer. Interestingly, the slope of improvement saturates and become inconsistent as q_l further decreases starting when q_h/q_l becomes higher than 2.5. This suggests that when the lowest bandwidth limits the highest QS to q_h, even when available bandwidth at a later time allows a QS below $0.4q_h$, it may be better to limit the q_l to $0.4q_h$. When q_h is already low (e.g. q_h=40), switching to q_l provides inconsistent gain. Our ANOVA analysis (described in Sec. 4.4) shows that the quality variation observed for both q_l=25 and q_l=16 is statistically insignificant (see entries for P_2 and P_3 in Tab. 4.10).

Based on the observation above, it would be essential to understand the joint effect of constant and variation effect of QS on perceptual quality and quantify the quality level by deriving the function forms. The intuitive way to understand the quality degradation is to investigate Q(q_l, q_h) v.s. $q_{\min}/q_{avg}$ as shown in Fig. 4.30. By applying the model in (4.13), the predicted quality with measured MOS are illustrated in Fig. 4.31 with PCC=0.942, RMSE=0.067. This is not an accurate way since we learn that the quality impact is quite dependent on both q_h and q_l, not just the average of q_h and q_l.

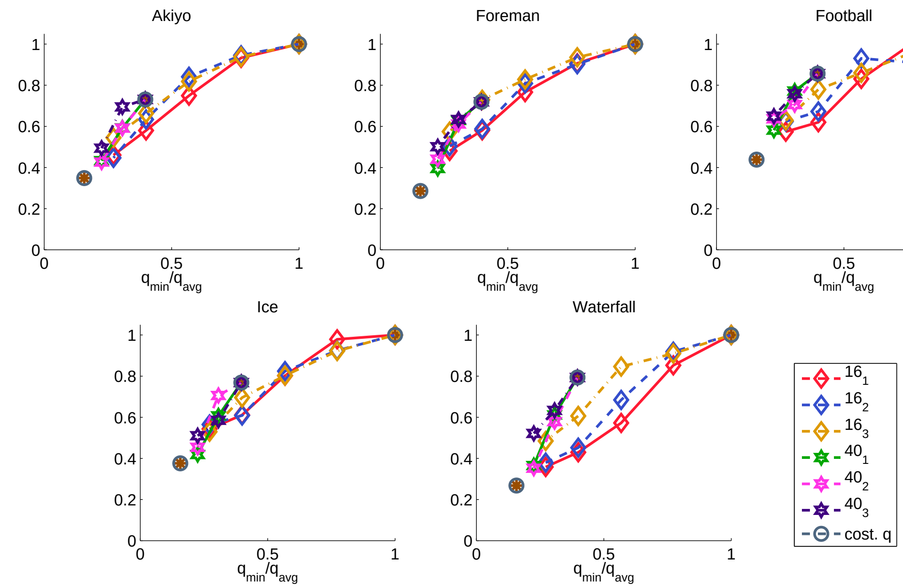


Figure 4.30: Q(q_h, q_l) vs. normalized average QS ($q_{\min}/q_{avg}$). Points are measured data and curves different markers and colors are corresponding to different q_l and Fz.

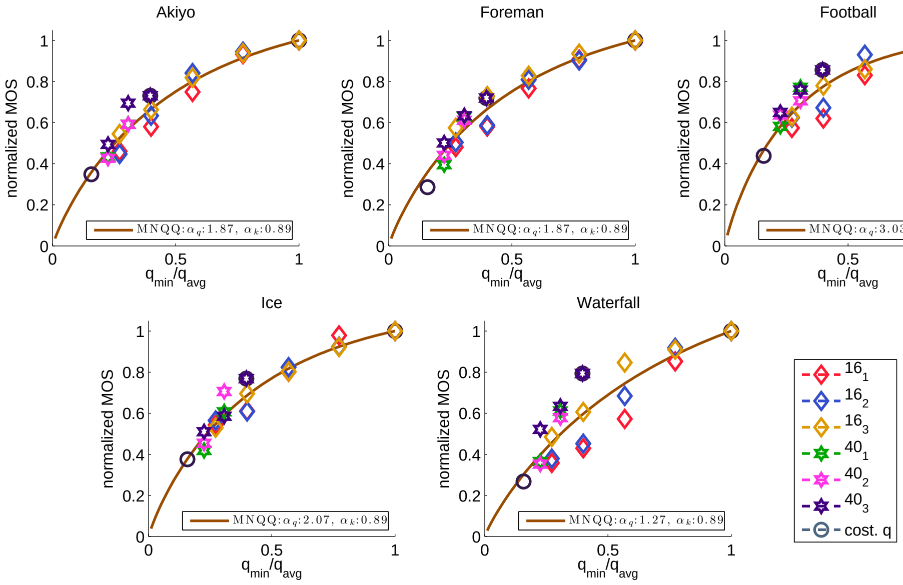


Figure 4.31: Q(q_h, q_l) vs. $q_{\min}/q_{avg}$ using an exponential model given in Eq. (4.13) with PCC = 0.942, RMSE = 0.067. Points are measured data and one curve is used to fit all different q_h and Fz. The model uses the same $\beta_{qv} = 0.89$ for all source sequences.

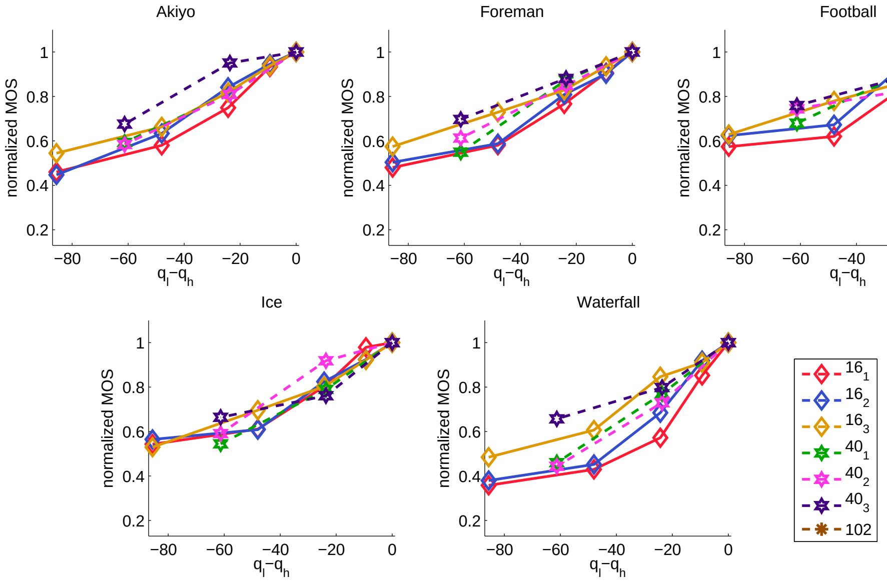


Figure 4.32: Q(q_h, q_l)/Q(q_l, q_l) vs. Δq ($q_l - q_h$) for low QS base.

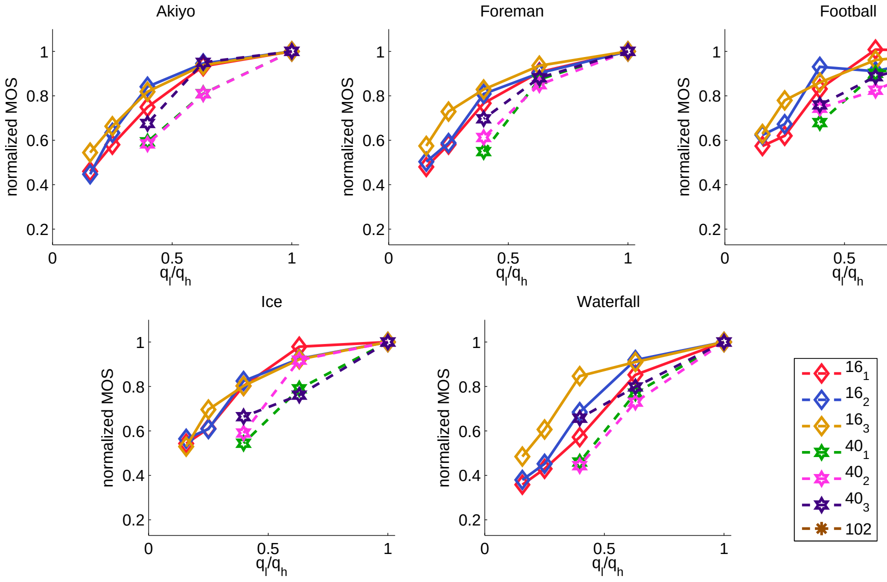


Figure 4.33: Q(q_h, q_l)/Q(q_l, q_l) vs. QS ratio (q_l/q_h) when q_l is fixed. Points are measured data and urves with different markers and colors are corresponding to different q_l and Fz

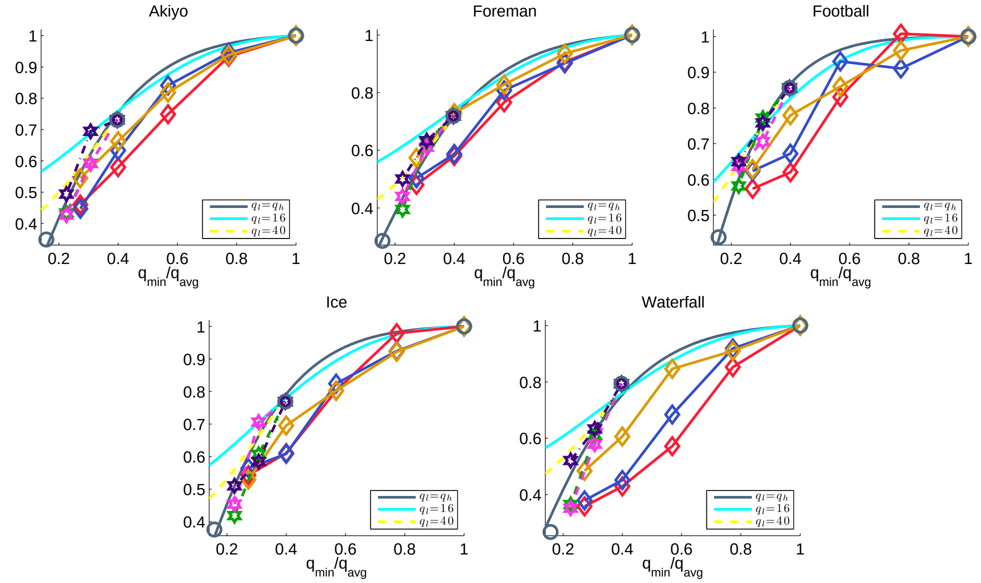


Figure 4.34: Q(q_l, q_l) vs. inverse normalized QS ($q_{\min}/q$). The dark blue curve is for constant QS. The cyan curve is determined by (MNQQ$_c$(q_l=16)+MNQQ$_c$(q_h))/2, while the orange curve is the (Q(q_l=40)+Q(q_h))/2.

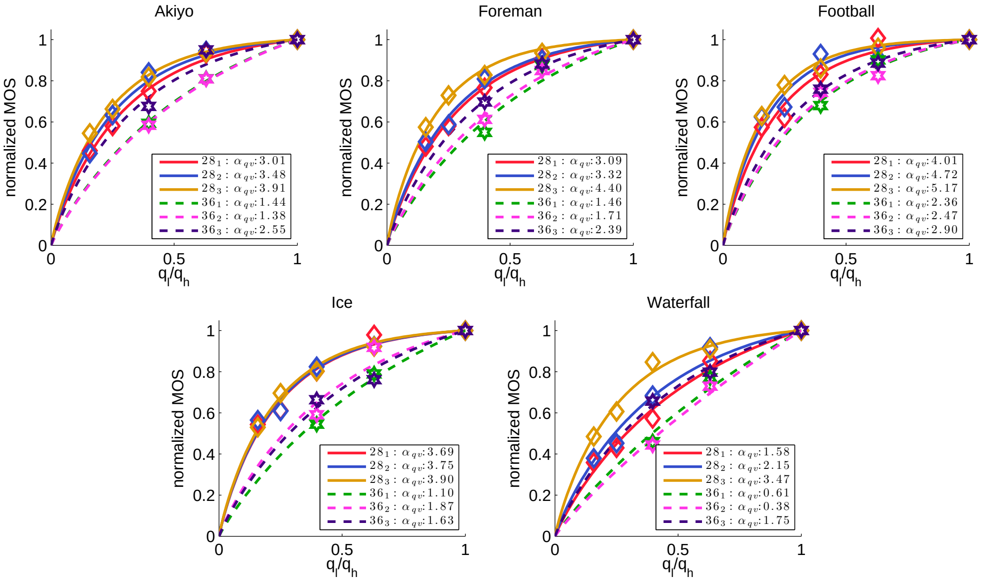


Figure 4.35: NQQ$_v$ vs. QS ratio (q_l/q_h) when q_l is fixed. Points are measured MOS and curves are obtained using Eq. (4.14) with PCC = 0.986, RMSE = 0.030. The parameter α_{qv} depends on Fz and q_l, and β_{qv} is 0.92.

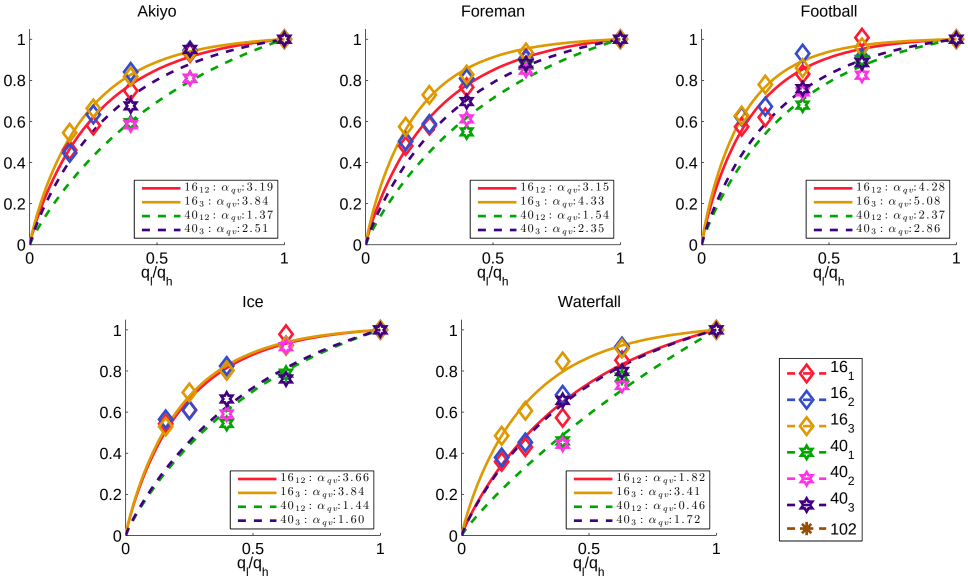


Figure 4.36: NQQ_v vs. QS ratio (q_l/q_h) when . Points are measured data and the curves are obtained using (4.14) with PCC = 0.983, RMSE = 0.035. Parameter α_{qv} depends on q_l and Fz but is the same at Fz=1,2 and β_{qv} is 0.92, respectively.

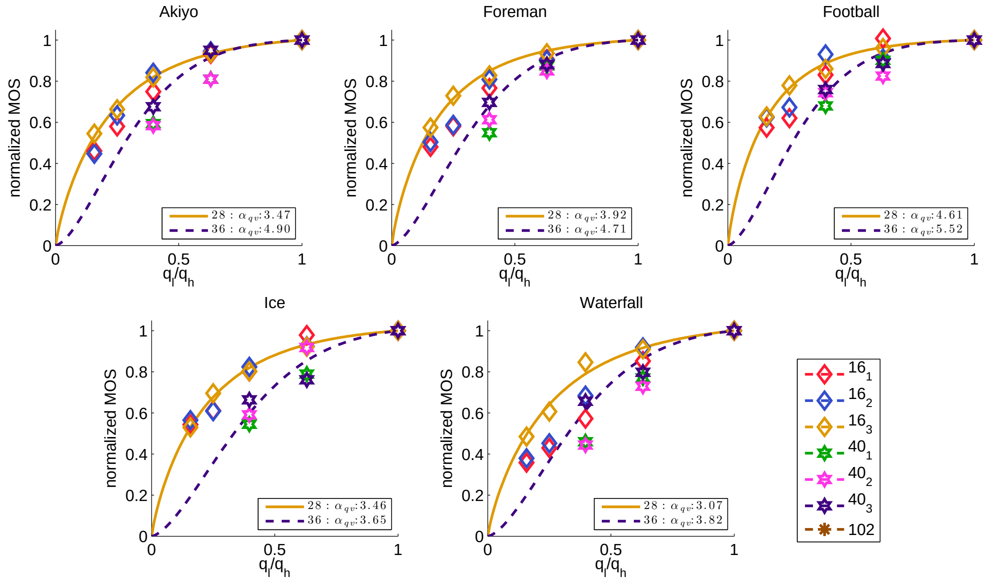


Figure 4.37: NQQ_v vs. QS ratio (q_l/q_h) using an exponential model given in (4.14) with PCC = 0.982, RMSE = 0.028. Points are measured data. Each model parameter is used for the same q_l but different Fz's.

Table 4.6: The model performance of $\overline{Q}_v(q_h, q_l)$

Function	Assumption	# par	PCC	RMSE
$\frac{1-e^{-\alpha_{qv}\cdot(q_r)^{\beta_{qv}}}}{1-e^{-\alpha_{qv}}}$	$\alpha_{qv}\succ q_l$, Fz; $\beta_{qv}\succ q_l$	8	0.991	0.022
	$\alpha_{rv}\succ q_l$, Fz but Fz=1,2	4	0.983	0.035
	$\alpha_{qv}\succ q_l$	2	0.967	0.047
$e^{\alpha_{qv}(\mathrm{Fz},q_l)\cdot(1-\bar{q}_r)}$	$\alpha_{qv}\succ q_l$, Fz	6	0.981	0.034
	$\alpha_{qv}\succ q_l$	2	0.965	0.049

$\succ$:Depends on

Table 4.7: The model performance of $\overline{Q}_v(r_{qh}, r_{ql})$

Function	Assumption	# par	PCC	RMSE
$\frac{1-e^{-\alpha_{rv}\cdot(\frac{r_{qh}}{r_{ql}})^{\beta_{rv}}}}{1-e^{-\alpha_{rv}}}$	$\alpha_{rv}\succ r_{ql}$, Fz ; $\beta_{rv}\succ r_{ql}$	8	0.984	0.033
	$\alpha_{rv}\succ q_l$, Fz	6	0.983	0.034
	$\alpha_{rv}\succ q_l$, Fz but Fz=1,2	4	0.980	0.038
	$\alpha_{rv}\succ q_l$	2	0.965	0.034
	α_{rv} not $\succ q_l$, Fz	1	0.955	0.056

$\succ$:Depends on

Table 4.8: The ANOVA test for changing interval:QS variation

Target pair	Factors	F-value	p-value
when q_l = 16			
Fz = 1 and 2	Δq*Fz	0.56	0.69
Fz = 1 and 3	Δq*Fz	2.37	0.06
Fz = 2 and 3	Δq*Fz	1.38	0.25
when q_l = 40			
Fz = 1 and 2	Δq*Fz	0.19	0.82
Fz = 1 and 3	Δq*Fz	3.85	0.03
Fz = 2 and 3	Δq*Fz	1.54	0.23

In stead of the quality variation in terms of Q(q_l, q_h)/Q(q_h, q_h) , we further investigates the normalized quality Q(q_l, q_h)/Q(q_l, q_l) v.s. q_l/q_h when q_l = 16 and 40 with 3 different Fz's in Fig. 4.33. It is interesting to observe that the falling trend of the the normalized MOS with Fz of 1 and 2 are similar except for those with Fz of 3. We found that the quality variation with QS ratio, $q_r = \frac{q_l}{q_h}$ can be modeled quite accurately by the inverted exponential function, i.e.,

$$\overline{Q}_v(q_h, q_l) = \frac{1 - e^{-L\cdot(q_r)^{\beta_{qv}}}}{1 - e^{-L}}.$$
$$\text{with } L = \begin{cases} \alpha_{qv}(1, q_l), & \text{if Fz} = 1, 2, \\ \alpha_{qv}(\text{Fz}, q_l), & \text{if Fz} = 3, \end{cases} \tag{4.14}$$

where α_{qv} is a model parameter and β_{qv} is constant value of 0.92. In Fig. 4.36 it shows that the fitting is very well with PCC= 0.983 and RMSE = 0.035. In order to further analyze the relationship between model parameters and model performance, we investigate the dependency of α_{qv} with Fz's and q_l with respect to their model performance in Tab. 4.6 as well as the ANOVA test in Tab. 4.8. In ANOVA test, we conduct two-way repetition ANOVA test on normalized MOS Q(q_h, q_l). Particularly, we only emphasize the statistical significance for each pair of Fz with the interaction to Δq. As shown in Tab. 4.8 that there is no significant difference between Fz 1 and 2 (p-value $>$ 0.05), while the dropping trend of Fz=1 and 3 are mostly distinguishable (p-value $<$ 0.05). It is also noted that the quality impact of Fz=2 and 3 is somewhat separable due to its p-value is closer to 0.05 than Fz=1 and 3. Note that instead of modeling the quality variation with QS ratio, we also want to learn, assuming q_l is fixed, how the quality changes with different Δq. Nevertheless, its sigmoid-like trend is not easy to model and it takes more number of parameters than other models.

Instead of modeling the quality variation with QS ratio, we would like to learn, assuming q_l is fixed, how the quality changes with different Δq. Figure 4.41 demonstrates Q(q_l, q_h)/Q(q_l, q_l), against Δq, to study the impact of the strength of frame rate variation. We observe that higher q_l has slower dropping trend than smaller q_l along the Δq trajectory. This again implies that when q_l is high, further inducing QS variation leads to more visual improvement, than when q_l is lower. However, it is note that if we look at the normalized MOS v.s. QS ratio in Fig. 4.33, the higher q_l has faster dropping trend than smaller q_l along the QS ratio trajectory. It could be because QS is the power law version of QP, the arithmetic distance between q_l and q_h becomes smaller when q_l is larger as q_h is fixed. We further to model the normalized MOS (MOS(q_h, q_l)/MOS(q_l, q_l)) against Δq by utilizing the sigmoid function, i.e.,

$$\mathrm{MNQQ}_v(q_h, q_l) = \frac{1.1}{1 + \beta_{qv} e^{\alpha_{qv}(q_h - q_l)}}, \tag{4.15}$$

where α_{qv} is model parameter and β_{qv} is a constant value of 0.11. As shown in Fig. 4.42 the model curves does not fit with measured quality very well, with PCC=0.945, RMSE = 0.072.

We summarize the performance of different models and its corresponding parameters with respect to different Fz's or q_l in Tab. 4.6. According to the number of model parameter and the accuracy, all these columns are very similar and we should choose one with fewest model parameters. Therefore, we conclude that Eq. (4.14) gives the promising results with less model parameters.

In order to predict the overall quality with the impact of both constant QS and QS variation, we propose to use the product of two models, $\overline{Q}_c(q)$ and $\overline{Q}_v(q_h, q_l)$, i.e.,

$$\mathrm{QQV}(q_h, q_l) = \mathrm{MNQQ}_c(q_l, q_l)\mathrm{MNQQ}_v(q_l, q_h), \tag{4.16}$$

where MNQQ_c can be replaced by the model in Eq. (4.13), while models in Eqs. (4.14) for MNQQ_v. After comparing with all the models and parameter dependency in Tab. 4.6, it suggests that we replace $\mathrm{MNQQ}_v(q_l, q_h)$ with the model in Eq. (4.14), i.e.,

$$\mathrm{QQV}_1(q_h, q_l) = \frac{1 - e^{-\alpha_q(\frac{q_{\min}}{q})^{\beta_q}}}{1 - e^{-\alpha_q}} \frac{1 - e^{-L\cdot(q_r)^{\beta_{qv}}}}{1 - e^{-L}}, \tag{4.17}$$

where, Fig. 4.39 illustrates that the predicted quality v.s. q_{avg} fits the measured data quite well with PCC=0.981, RMSE= 0.038. Here, α_{qv} depends on q_l and Fz's but is the same for both Fz=1,2, while β_{qv} is a constant value of 0.91. Base on this function form and specific order in Eq. (4.17), the first term is responsible for predicting the quality at lower constant QS, from which the second term deduces the quality of QS variations.

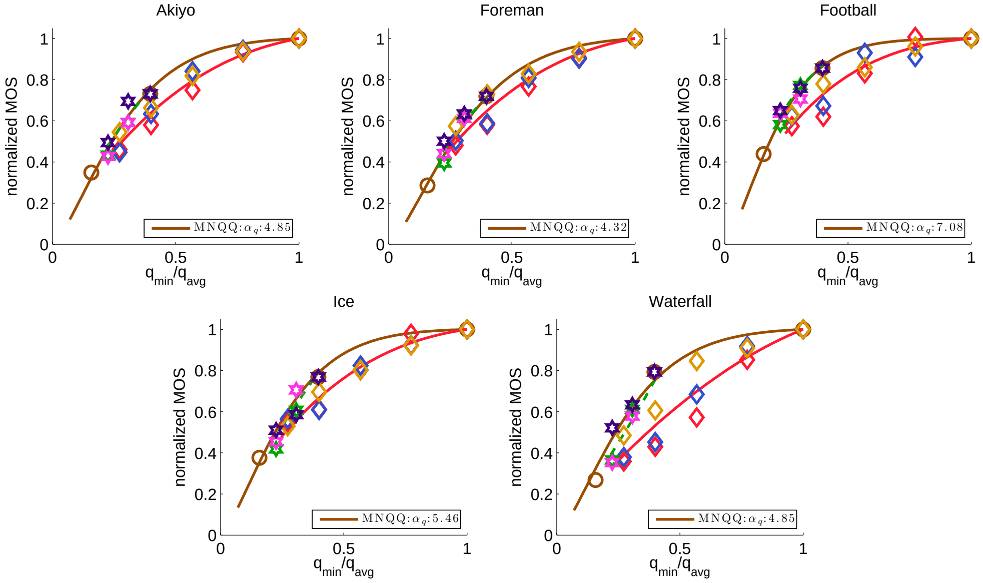


Figure 4.38: Q(q_h, q_l) vs. $q_{\min}/q_{avg}$ using Eq. (4.17) with PCC = 0.971, RMSE = 0.044. Points are measured data and predicted curves are from least square fitting. The parameter α_{qv} only depends on q_l and $\beta_{qv} = 0.9$ is a constant.

Moreover, we plot Q(q_h, q_l)/Q($q_{\min}, q_{\min}$) against $q_{\min}/q_h$ in Fig. 4.43. Note that we normalize the MOS by the MOS at $q_{\min}$ instead of the quality at q_l so that the falling trends is independent of q_l but $q_{\min}$. It is interesting to observe that the dropping trends under different q_l and Fz's are clustered together. This implies that the QS variation can be well approximated by the $q_{\min}/q_h$ without knowing the effect of q_l. We found that the variation of quality with $q_{\min}/q_h$ can be modeled quite well using an inverted exponential function,

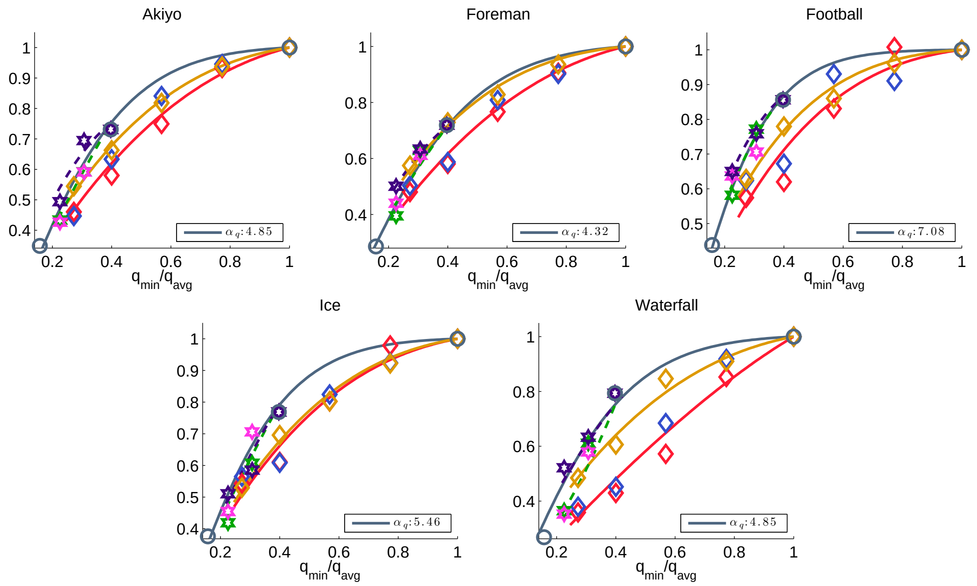


Figure 4.39: Q(q_h, q_l) vs. $q_{\min}/q_{avg}$ using Eq. (4.17) with PCC = 0.983, RMSE = 0.034. Points are measured data and predicted curves are from least square fitting. The parameter α_{qv} depends on Fz but is the same at Fz=1,2 and $\beta_{qv} = 0.91$ is a constant.

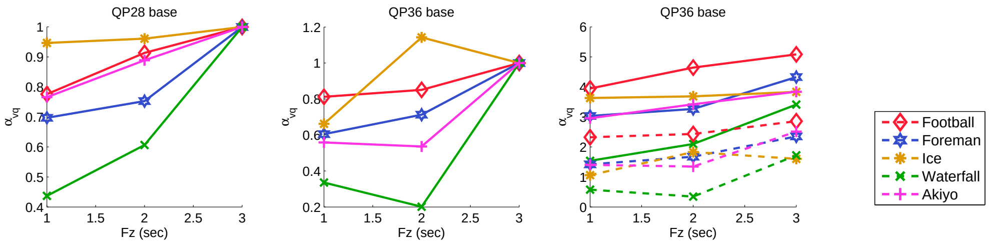


Figure 4.40: The relationship between α_{qv} and frequency using Eq. (4.17) for base QS 16 and 40, while using $\beta_{qv} = 0.91$ for both base QP16 and 40.

i.e.,

$$\mathrm{QQV}_2(q_h) = \frac{1 - e^{-\alpha_{qv}(\frac{q_{\min}}{q_h})^{\beta_{qv}}}}{1 - e^{-\alpha_{qv}}}, \tag{4.18}$$

where only one model parameter α_{qv} and a constant value of $\beta_{qv} = 0.96$ are needed to capture the quality variation with normalized QS and both these two parameters are independent of Fz and q_l but video content. Figure 4.44 shows the fitting curves are accurate with PCC=0.965, RMSE=0.049. By comparing with the result of normalized MOS against $q_{\min}/q_{avg}$, (4.18) more accurately reflects quality degradation than the model in (4.13),

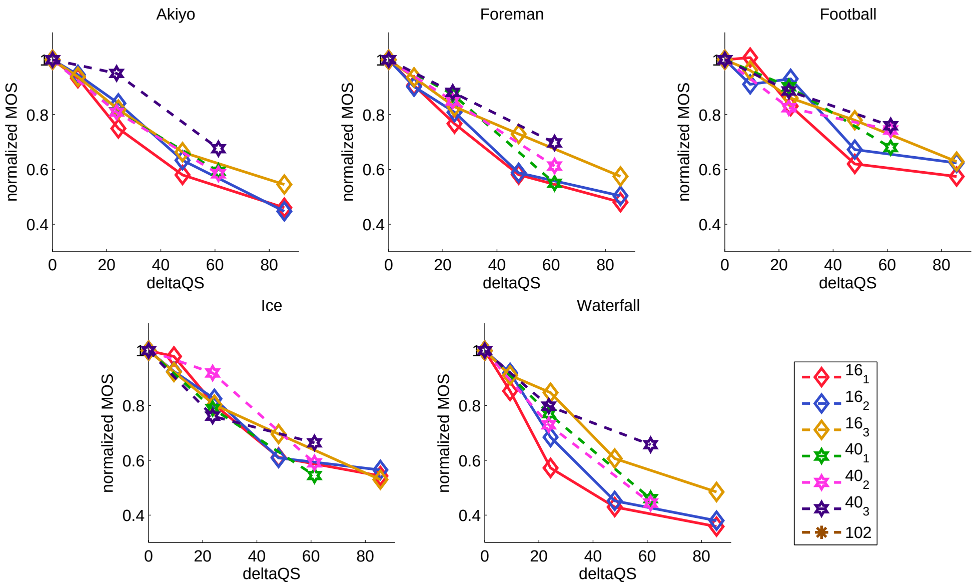


Figure 4.41: Q(q_h, q_l)/Q(q_l, q_l) v.s. Δq.

which considers $q_{\min}/q_{avg}$ as variable. However, we observe that human eyes are still sensitive to the variation of QS for both q_h and q_l so that we assume the we can only apply this model in (4.18) on certain range of q_l. Note that we could also develop the quality model of QS variation when q_h is fixed, nevertheless, the insufficiency of dataset (q_h=102 includes only one Fz, and q_h=40 just includes one measured point in each source sequence.) cannot validate and derive a robust model.

4.3.3 Impact of Bit Rate Variation Due to QS Variaiton

We also compare quality of the videos with different QS variations under the same average bit rate ($r_{avg}/r_{\max}$) in Fig. 4.47, where $r_{\max}$ is the maximum average BR at $q_{\min} = 16$. It is clear that, a constant QS video has a better quality than a video with QS (especially when the QS variation is large), under the same average bit rate.

Although we already explored several models to predict the impact of quality on QS variations for a given pair of q_l and q_h, it may not be applicable if the QS information is not available. Let r_{ql} and r_{qh} denote the BR at q_l and q_h, respectively. Besides the Eq. 4.13 in terms of QS, we model the normalized quality, Q(r_{qh}, r_{ql})=MOS(r_{qh}, r_{ql})/MOS($r_{\max}, r_{\max}$),

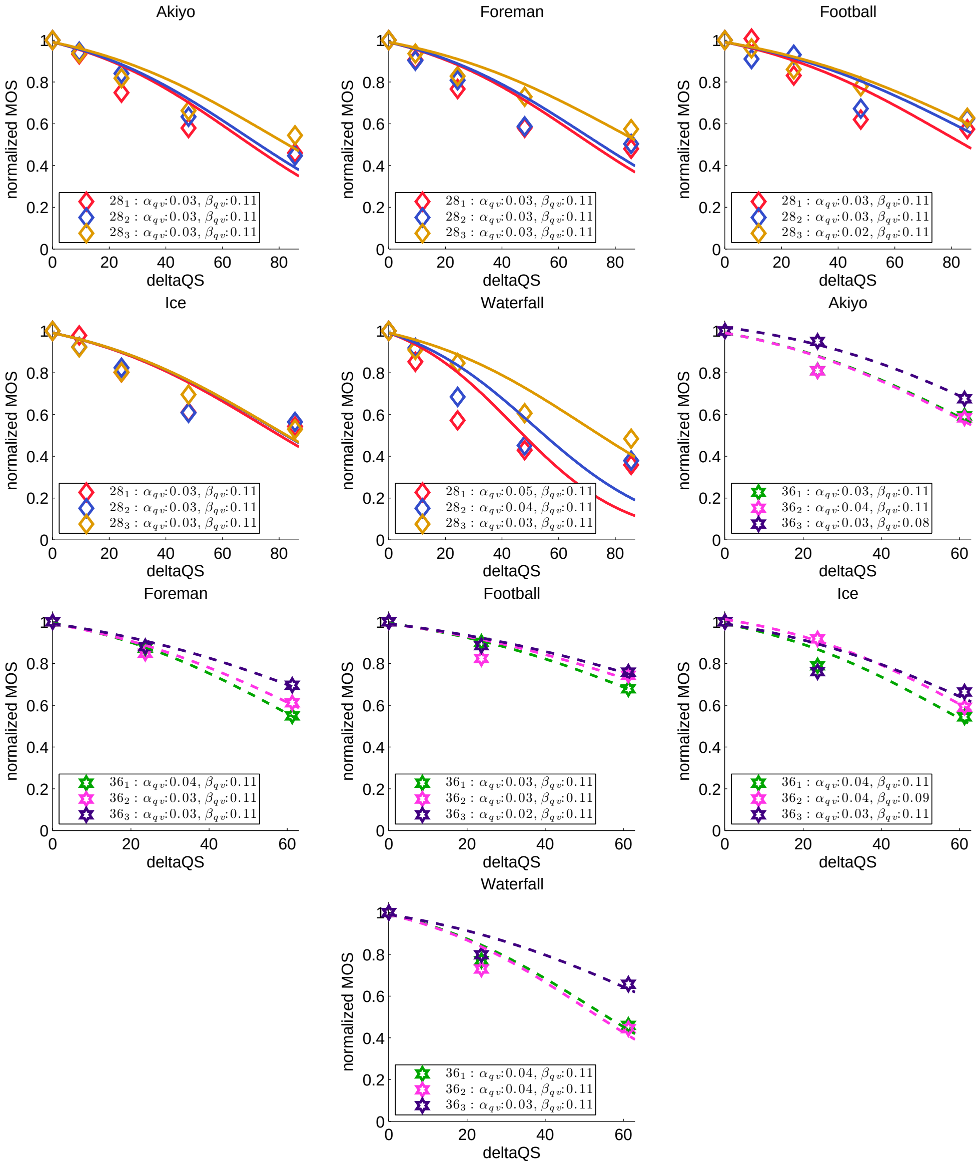


Figure 4.42: Q(q_h, q_l)/Q(q_l, q_l) vs. $Delta$QS using a sigmoid function given in Eq. (4.15) with PCC = 0.945, RMSE = 0.072. Points are measured MOS and curves with same marker are the same q_l.

v.s. normalized BR ($r/r_{\max}$) for constant BR, i.e., $r_{qh} = r_{ql}$, using the exponential function for a given spatial resolution s and FR t by

$$\mathrm{MNQQ}_{cr}(r; s, t) = \frac{1 - e^{-\alpha_r (\frac{r}{r_{\max}})^{\beta_r}}}{1 - e^{-\alpha_r}}, \tag{4.19}$$

,where β_r is constant for all five sequences with a value of 0.82. We plot the predicted

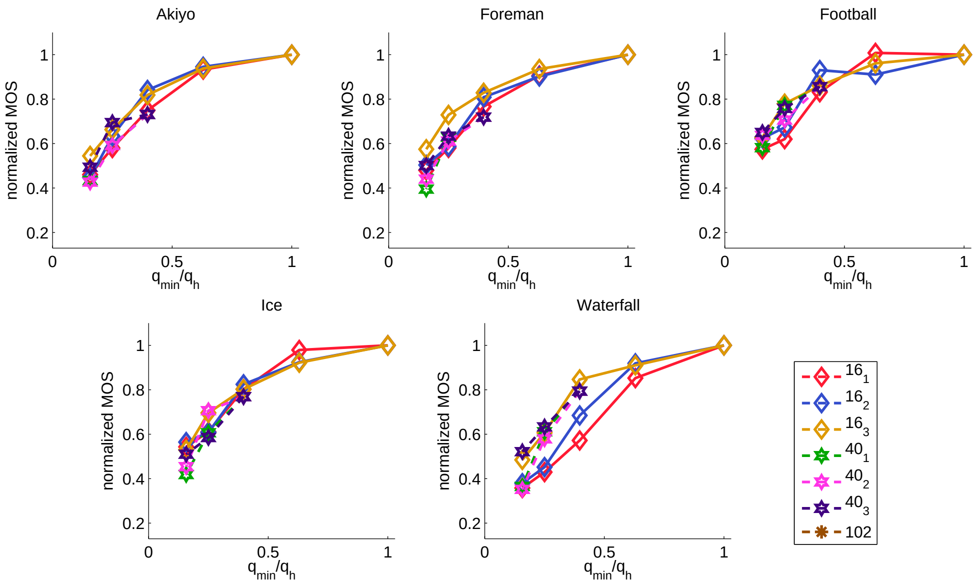


Figure 4.43: Q(q_h, q_l) vs. $q_{\min}/q_h$.

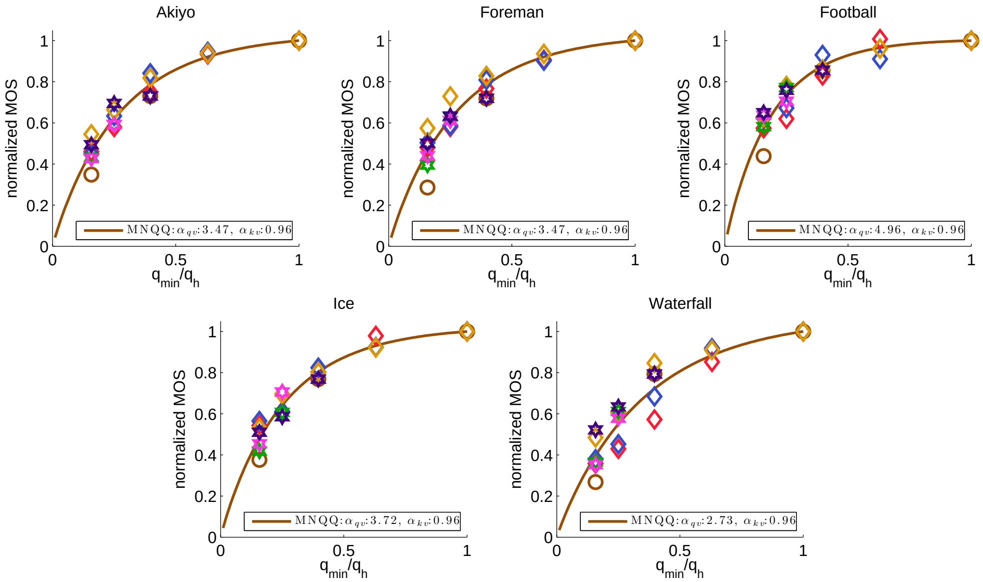


Figure 4.44: Q(q_h, q_l) vs. $q_{\min}/q_h$ using Eq. (4.18) with PCC = 0.965, RMSE = 0.049. Points are measured data and only one model curve is obtained to fit all the measured data.

quality with measured data in Fig. 4.47 and the fitting is very good with PCC=0.992, RMSE=0.035.

Regarding the quality impact of QS variation when only BR info is available, let r_{ql} and r_{qh} denote the BR at q_l and q_h, respectively. Figure 4.48 illustrates the measured

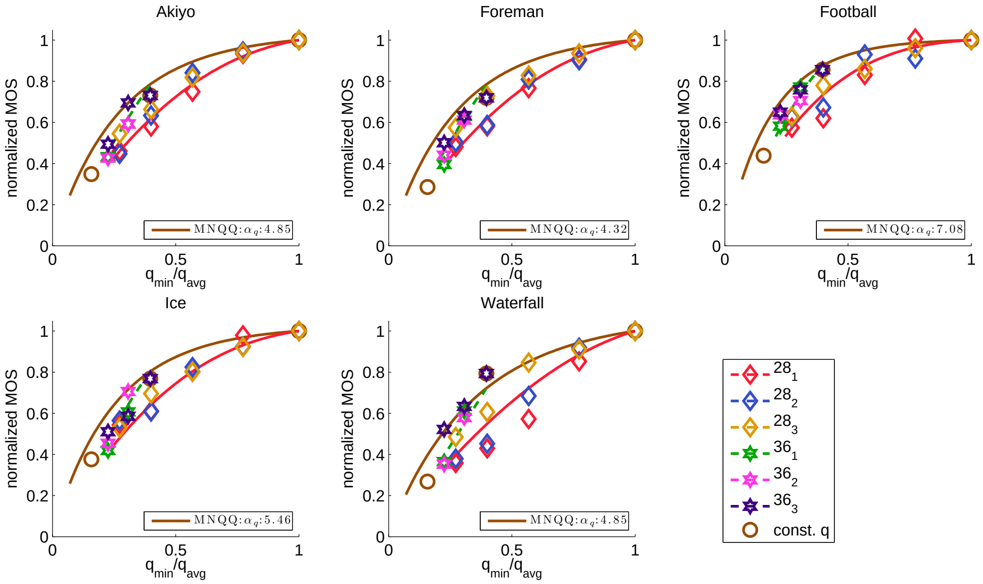


Figure 4.45: Q(q_h, q_l) vs. $q_{\min}/q_{avg}$ using Eq. (4.18) with PCC = 0.965, RMSE = 0.049. The parameter α_{qv} is independent of Fz and $\beta_{qv} = 0.93$ for both base QP16 and 40. Points are measured MOS and curves with same marker are the same q_l.

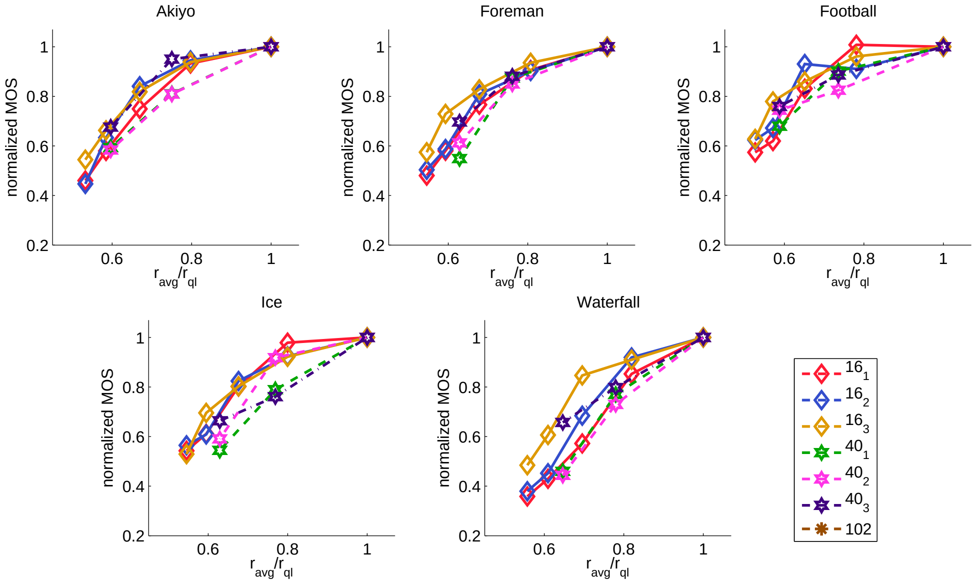


Figure 4.46: Q(q_h, q_l) vs. r_{avg}/r_{ql} for low QS base.

Q(r_{qh}, r_{ql})/Q(r_q, r_{ql}) (in points) v.s. normalized BR (r_{qh}/r_{ql}). It is found that the quality

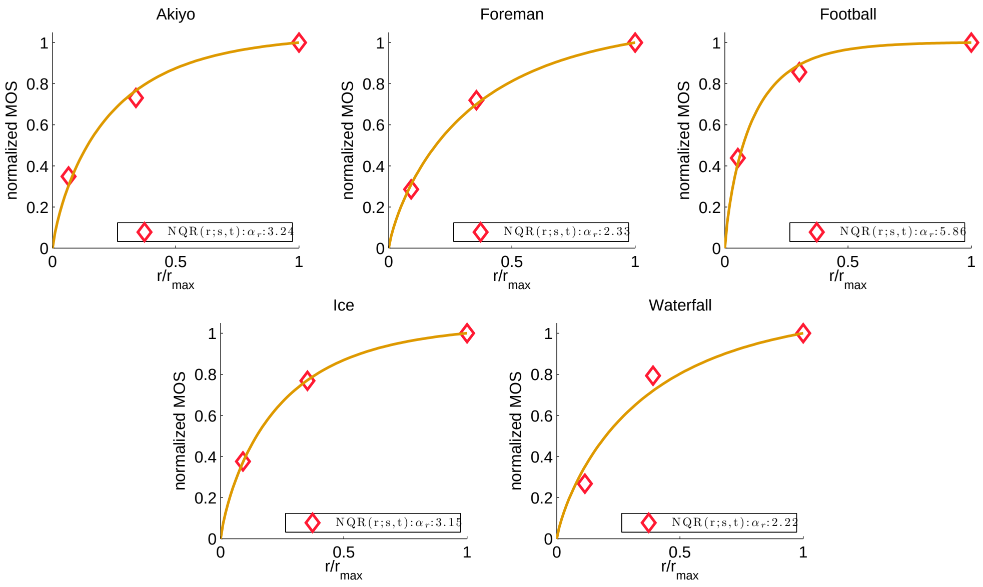


Figure 4.47: Q(r_{qh}, r_{ql}) vs. $r/r_{\max}$ using an exponential model given in Eq. (4.19) with PCC = 0.992, RMSE = 0.035. Points are measured data and predicted curves are from least square fitting.

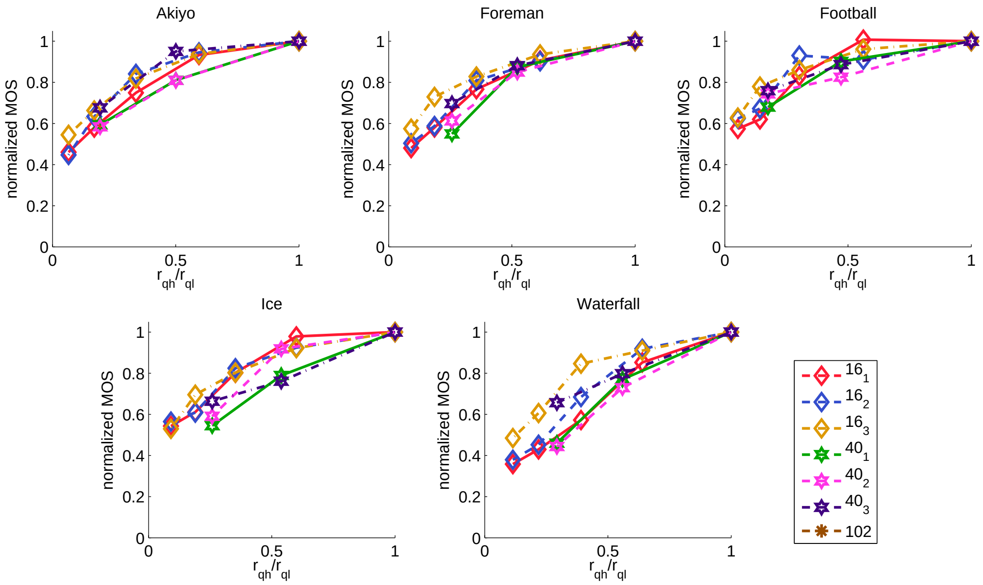


Figure 4.48: Q(r_{qh}, r_{ql})/Q(r_{ql}, r_{ql}) vs. r_{qh}/r_{ql} when r_{ql} is fixed. Curves with different markers and colors are corresponding to different q_l and Fz.

degradation along with r_{qh}/r_{ql} can be well captured by exponential function,

$$\overline{Q}_{rv}(r_{ql}, r_{qh}) = \frac{1 - e^{-K\cdot(\frac{r_{qh}}{r_{ql}})^{\beta_{rv}}}}{1 - e^{-K}}.$$

$$\text{with } K = \begin{cases} \alpha_{rv}(1, r_{ql}), & \text{if Fz} = 1, 2, \\ \alpha_{rv}(\text{Fz}, r_{ql}), & \text{if Fz} = 3, \end{cases} \tag{4.20}$$

Where $\alpha_{rv}(\mathrm{Fz}, r_{ql})$ is model parameter and β_{rv} is a constant value of 0.54 for all five source sequences. Figure 4.51 demonstrates the fitting curves with measure MOS using (4.20) with high PCC = 0.98. This model requires only 4 parameters for each source sequence. In addition, we compare the model performance in Tab. 4.7 regarding the dependency of parameter α_{rv} with Fz and r_{ql}. When considering the number of parameter and accuracy, we conclude that the model in (4.20) can predict quality very well with lower model complexity.

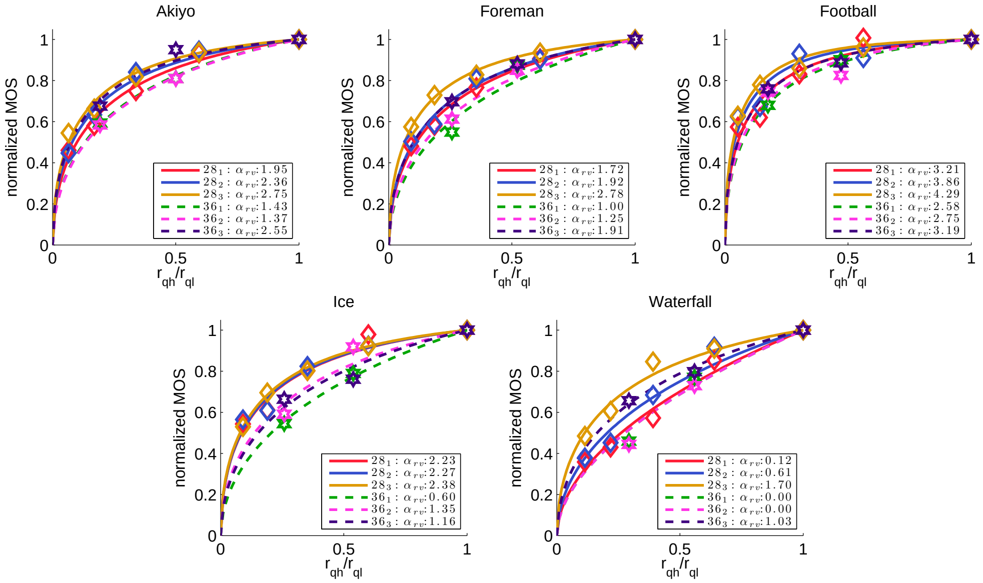


Figure 4.49: Q(r_{qh}, r_{ql})/Q(r_{ql}, r_{ql}) vs. r_{qh}/r_{ql} using an exponential model given in Eq. (4.20) with PCC = 0.983, RMSE = 0.034, when α_{rv} depends on both Fz and r_{ql}. Points are measured MOS and curves with same marker are the same q_l. The model uses a total of 6 model parameters for each sequence. The same $\beta_{qv} = 0.54$ is for all Fz's and q_l.

According to Eq. (4.17), the overall quality can also be derived as the product of Eqs (4.19) and (4.20), i.e.,

$$\mathrm{QQVBR}_1(r_{ql}, r_{qh}) = \frac{1 - e^{-\alpha_r (\frac{r}{r_{\max}})^{\beta_r}}}{1 - e^{-\alpha_r}} \frac{1 - e^{-K \cdot (\frac{r_{qh}}{r_{ql}})^{\beta_{rv}}}}{1 - e^{-K}}. \tag{4.21}$$

We plot the predicted quality v.s. average BR in Fig. 4.54 with PCC=0.976, RMSE = 0.041.

Moreover, we plot the Q(r_{qh}, r_{ql}) against $r_{qh}/r_{\max}$ in Fig. 4.55. Note that we normalize the MOS by the MOS at $r_{\max}$ instead of the quality at r_{ql} so that the falling trends is

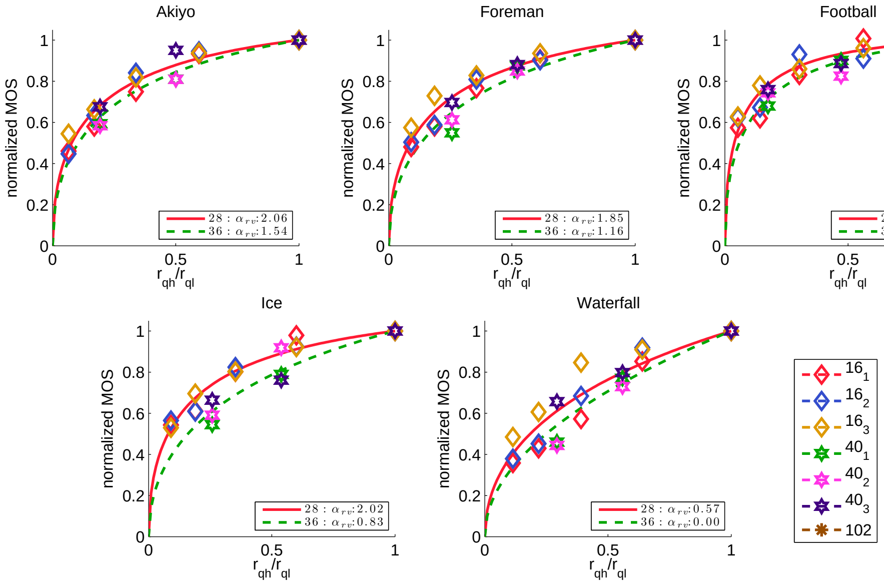


Figure 4.50: Q(r_{qh}, r_{ql})/Q(r_{ql}, r_{ql}) vs. r_{qh}/r_{ql} using Eq. (4.20) with PCC = 0.965, RMSE = 0.034, when α_{rv} only depends on r_{ql}. Points are measured MOS and curves with same marker are the same q_l. They model uses a total of 2 model parameters for each sequence. The same $\beta_{qv} = 0.5$ is for all Fz's and q_l.

independent of r_{ql}. It is interesting to observe that the dropping trends under different r_{qh} and Fz's are clustered together. This implies that the QS variation can be well approximated by the $q_{\min}/q_h$ without knowing the effect of q_l. We found that the variation of quality with $q_{\min}/q_h$ can be modeled quite well using an inverted exponential function, i.e.,

$$\mathrm{QVVBR}_2(r_{qh}) = \frac{1 - e^{-\alpha_{rv}(\frac{r_{qh}}{r_{\max}})^{\beta_{rv}}}}{1 - e^{-\alpha_{rv}}}, \tag{4.22}$$

where only one model parameter α_{rv} and a constant value of $\beta_{rv} = 0.47$ are needed to capture the quality variation with normalized QS and both these two parameters are independent of Fz and r_{qh} but video content. Figure 4.56 shows the fitting curves are accurate with PCC=0.961, RMSE=0.052.

It is also interesting to observe that the falling trend of $\mathrm{MNQT}_{cr}(r; s, t)$ is more exponential law, but the falling trend of $\mathrm{MNQQ}_{cr}(r; s, q)$ is more like power law or sigmoid function.

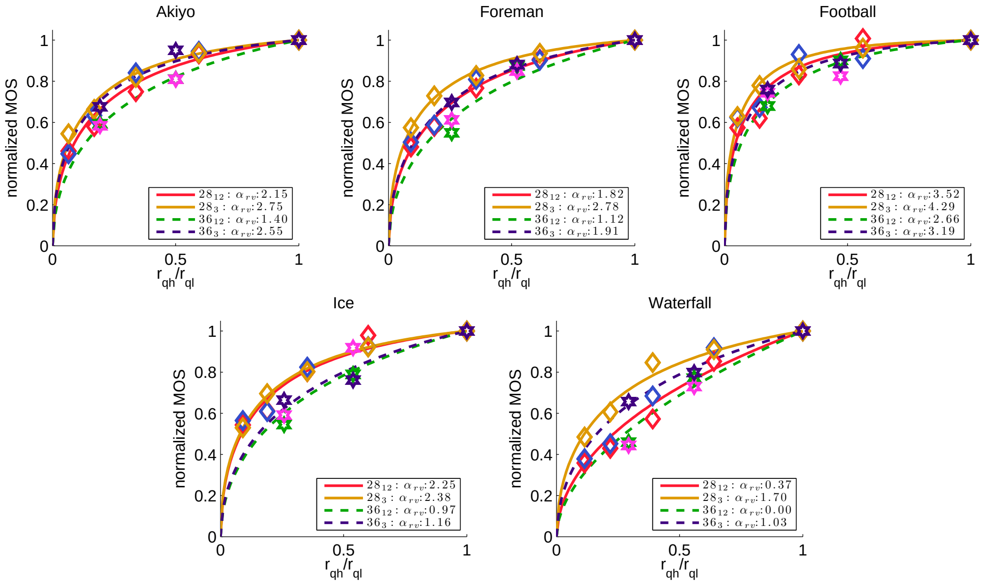


Figure 4.51: Q(r_{qh}, r_{ql})/Q(r_{ql}, r_{ql}) vs. r_{qh}/r_{ql} using Eq. (4.20) with PCC=0.980, RMSE = 0.038. Points are measured data and model curves are obtained by least square fitting. The model uses a total of 4 model parameters for each sequence and β_{qv} is a constant value of 0.54 for all five source sequences.

4.4 Statistic Analysis

The results presented in Sec. 4.2 and 4.3 show that FR/QS variation and video content both affect the perceptual quality very well. To evaluate whether the changes in quality ratings due to these factors are statistically significant, we perform the three-way Analysis of Variance (ANOVA) [69]. With ANOVA, we compute the probability (p-value, which is derived from the cumulative distribution function of F based on the F-value) of the event that the difference in MOS when a particular variable is changed is due to chance. If this probability is low (p-value $<$ 0.05), we consider this variable as having statistically significant difference on raw MOS.

Instead of performing ANOVA on all possible FR/QS variation cases, we focus on a few interesting pairs of FR/QS variation, listed in Table 4.9 and 4.10. For each considered case, we evaluate the p-value due to $\Delta q/\Delta t$ Fz (frequency), video content (i.e. across five video sources), and their interactions, respectively.

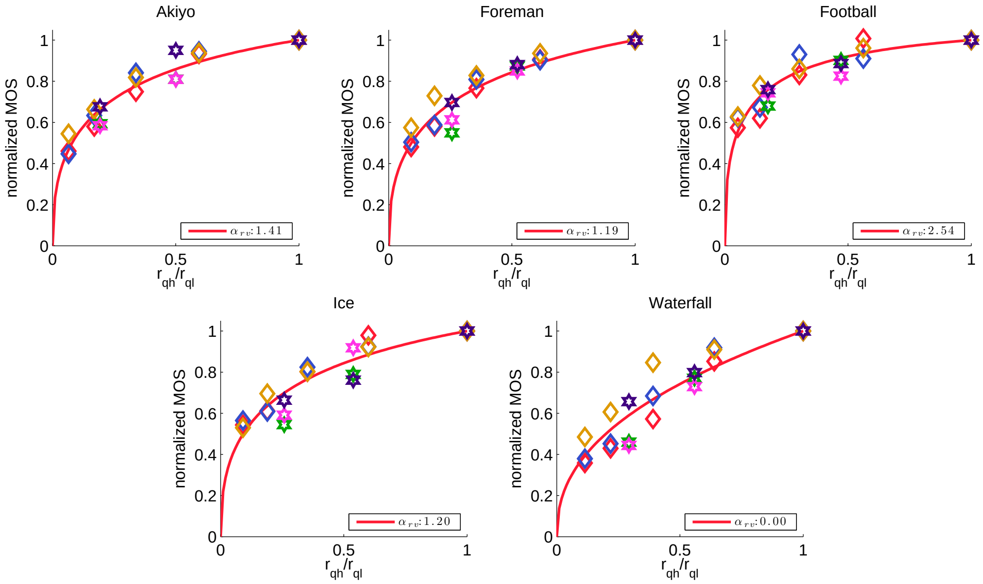


Figure 4.52: Q(r_{qh}, r_{ql})/Q(r_{ql}, r_{ql}) vs. r_{qh}/r_{ql} using Eq. (4.20) with PCC = 0.953, RMSE = 0.056, when α_{rv} only depends on r_{ql}. Points are measured data and predicted curves are obtained from least square fitting. They model uses a total of 2 model parameters for each sequence. β_{qv} is a constant value of 0.43 for all five source sequences.

4.4.1 Statistical Significance for Frame Rate Variation

As shown in Tab. 4.10, the impact of FR variation is significant in all the cases except P_1. It explains that viewers are not sensitive to the quality impact on FR variation between 30 and 15Hz, so that when t_h is high, switch back and forth to slightly lower the FR regularly doesnt affect the subjective rating. But it is interesting to observe that the effect of Fz does not happen in P_4. This indicates that when the constant FR is relatively low, the changes in variation interval wont significantly influence the perceptual quality. The effect of video content is significant for P_1-P_4. This suggests that the video content does affect viewer ratings for all the cases, but it influences viewer ratings less significantly at lower t_l (smaller F-value) . As we next look at the three-way ANOVA interaction between Δt and frequency, all the cases show significant differences (p-value of all the cases are < 0.05). In addition, the interaction between content and frequency also has significant difference for all the cases P_1 - P_4, while there is no significant interaction between content and Δt.

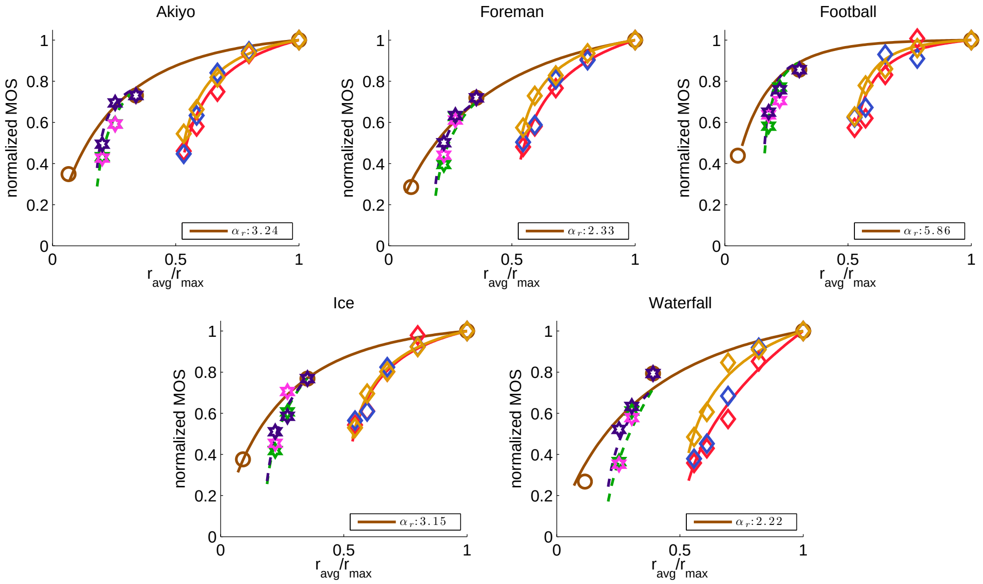


Figure 4.53: Q(r_{qh}, r_{ql}) vs. $r_{avg}/r_{\max}$ using Eq. (4.17) with PCC=0.976, RMSE = 0.041. Points are measured data and curves with same marker are the same q_l. The model uses a total of 5 model parameters for each sequence.

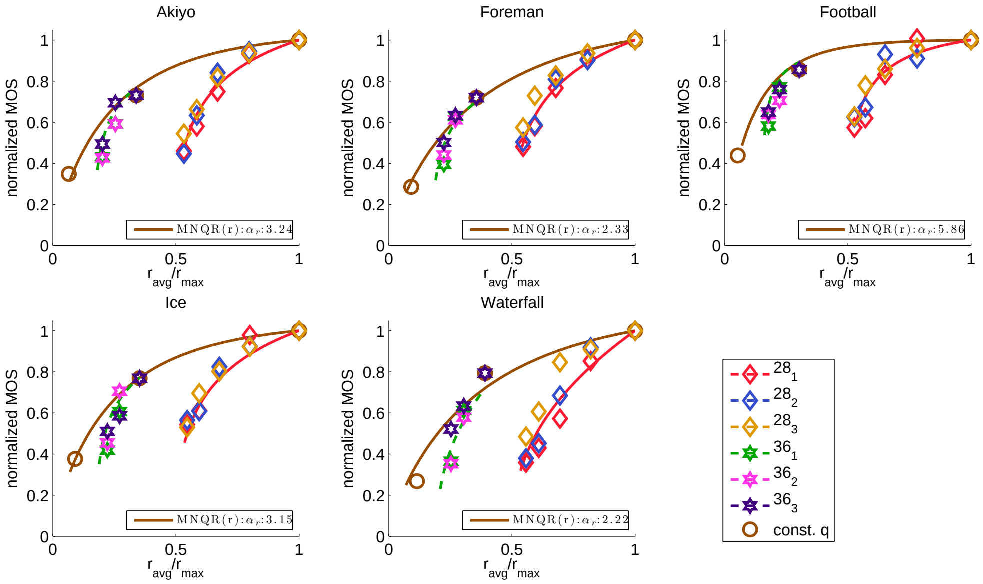


Figure 4.54: Q(r_{qh}, r_{ql}) vs. $r_{avg}/r_{\max}$ using Eq. (4.20) with PCC=0.956, RMSE = 0.053. Points are measured MOS and curves with same marker are the same q_l. The model uses a total of 3 model parameters for each sequence.

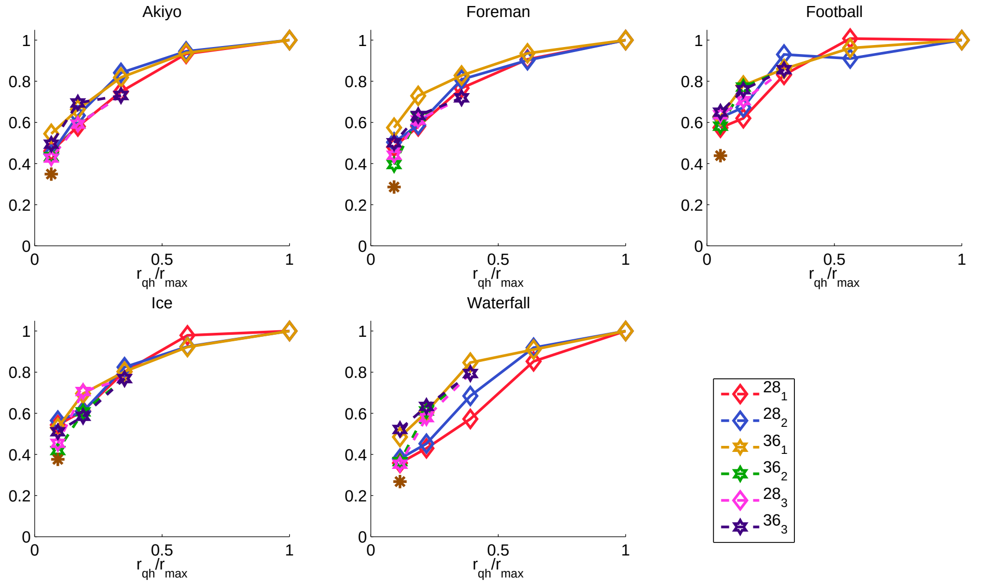


Figure 4.55: Q(r_{qh}, r_{ql}) vs. $r_{qh}/r_{\max}$.

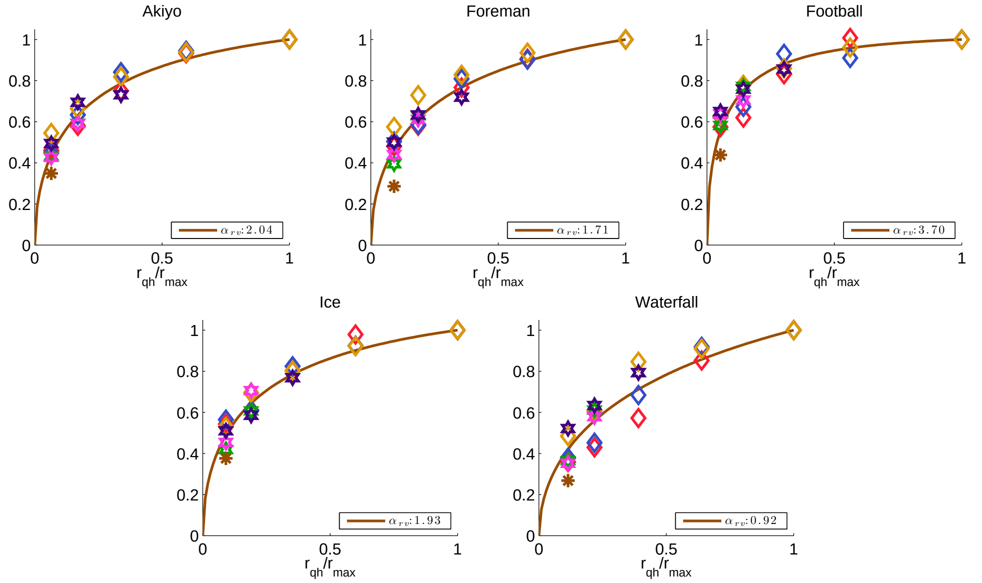


Figure 4.56: Q(r_{qh}, r_{ql}) vs. $r_{qh}/r_{\max}$ using Eq. (4.21) with PCC=0.961, RMSE=0.052. Points are measured data and model curves are obtained by least square fitting. The model uses one model curve for each sequence. β_{rv} is a constant value of 0.47 for all five source sequences.

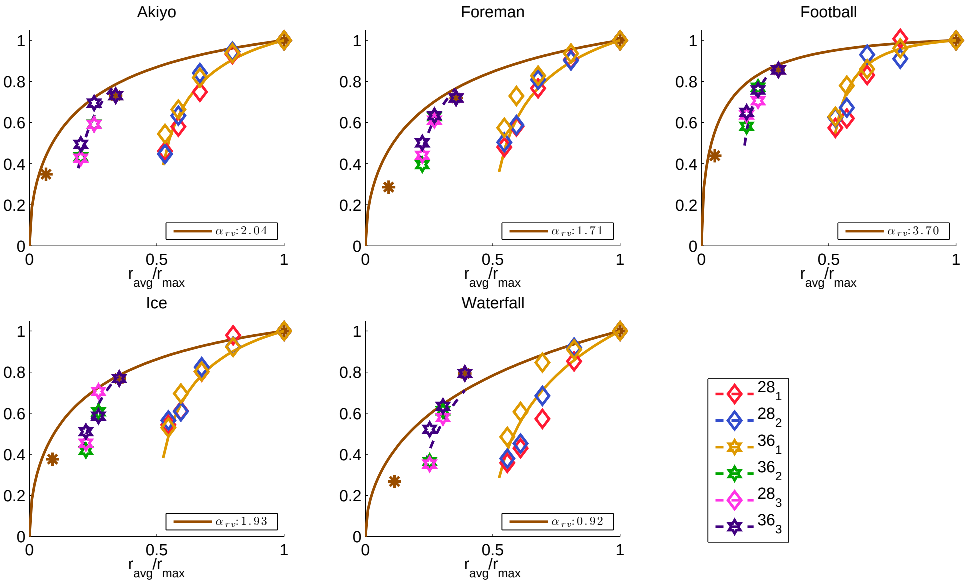


Figure 4.57: Q(r_{qh}, r_{ql}) vs. $r_{avg}/r_{\max}$ using Eq. (4.21) with PCC=0.961, RMSE=0.052.

4.4.2 Statistical Significance for QS Variation

Results given under P_1 in Tab. 4.10 are obtained by considering all test conditions together. We can see that QS variation, frequency, and video content all have significant impact on the subjective ratings. Furthermore, there is no significant interaction between each pair of QS variation, content and frequency and neither the three way interaction of them.

In addition to conducting ANOVA over all data, we also looked at a few specific cases, which are also given in Tab. 4.10. P_2 and P_3 correspond to cases where q_l and q_h are similar and q_h is not too high. The ANOVA results show that the quality difference is insignificant, which is also as expected as the quality with q_l and that with q_h are very similar when q_l is close to q_h. This result confirms that the observed improvement when switching from q_h=40 to q_l=25, 16 in Fig. 4.29 is not statistically significant. P_4 and P_5 correspond to cases where q_l and q_h are very different. ANOVA results confirm that the difference in the quality ratings between the considered pairs is statistically significant, and Fz and video content both impact the observed difference. This is as expected, as a viewer can easily notice the difference in visual quality caused by switching between q_l and q_h, when q_l and q_h are

Table 4.9: Three-way ANOVA results for FR variation

Variables	F-value	p-value
P_1 : (30, 15) and (30, 30)		
ΔFR	1.74	0.187
Frequency	21.22	0
Content	10.58	0
Content*frequency	2.56	0.009
ΔFR*frequency	44.95	0
ΔFR*Content	1.65	0.15
P_2 : (15, 7.5) and (15, 15)		
ΔFR	45.02	0
Frequency	3.19	0.042
Content	11.34	0
Content*frequency	2.6	0.008
ΔFR*frequency	8.21	0
ΔFR*Content	0.52	0.72
P_3 : (30, 15) and (15, 15)		
ΔFR	44.52	0
Frequency	21.69	0
Content	14.24	0
Content*frequency	3.28	0.001
ΔFR*frequency	35.1	0
ΔFR*Content	0.47	0.75
P_4 : (15, 7.5) and (7.5, 7.5)		
ΔFR	31.36	0
Frequency	2.48	0.08
Content	7.59	0
Content*frequency	2.06	0.03
ΔFR*frequency	7.35	0.0007
ΔFR*Content	1.69	0.15

very different. This result confirms that the observed quality improvement when switching between 102 and 16 as QS_b=102 in Fig. 4.29 is not by chance. The effect of video content is significant for P_2-P_5. This suggests that the video content does affect viewer ratings for all the cases, especially it significantly influences viewer ratings for larger Δq.

Table 4.10: Three-way ANOVA results for QS Variation

Factors	F-value	p-value
P_1 : All Data		
Δq	26.8	1e-5
Content	5.25	9e-4
Frequency	4.28	0.01
Content*frequency	0.07	0.99
Δq*frequency	0.54	0.82
Δq*Content	0.45	0.96
Δq*Content*frequency	0.05	1
P_2 : Q(40, 40) and Q(40, 25)		
Δq	0.02	0.886
Content	8.16	6.3e-3
Frequency	2.89	0.11
Content*frequency	1.62	0.25
Δq*frequency	2.11	0.18
P_3 : Q(40, 40) and Q(40, 16)		
Δq	2.53	0.15
Content	11.1	0.002
Frequency	3.15	0.09
Content*frequency	1	0.5
Δq*frequency	3.15	0.09
P_4 : Q(16, 16) and Q(16, 102)		
Δq	3041	0
Content	20.41	3e-4
Frequency	5.52	0.03
Content*frequency	1	0.5
Δq*frequency	5.52	0.03
P_5 : Q(102, 102) and Q(16, 102)		
Δq	368.79	0
Content	243.21	0
Frequency	5.52	0.031
Content*frequency	1	0.5
Δq*frequency	5.52	0.03

4.5 Summary

In this chapter, we report the results of our subjective experiments to investigate the impact of periodic FR/QS variation on the perceived video quality. We observed following interesting trends. Regarding the FR variation, firstly, under the same t_{avg} the quality for a video with a constant frame rate is higher than that with FR variation, alternating between t_l and t_h; and secondly, the degradation due to FR change is more severe when Δt is higher; thirdly, points corresponding to t_l=t_h/2 are quite close to the operational quality-frame rate curves achievable by using constant FR for most of the sequences. But those with t_l lower than t_h/2 are much below the curve. Finally, when f_l is fixed, alternating between t_l and t_h is generally better than staying at t_l as $t_h/t_l \leq 2$. The slope of improvement reduces as t_h increases, and the degree of improvement is inconsistent. However, the quality improvement become saturated as $t_h/t_l > 2$.

Regarding the QS variation, firstly, under the same average QS, a video with a constant QS is perceptually more appealing than a video with variable QS; Secondly, the quality of a video with varying QS alternating between q_h and q_l is generally equal or better than a video with a constant high QS q_h , when q_h is high and the improvement is more significant when Fz is longer; and thirdly, the quality degradation due to QS variation between q_l and q_h, compared to the quality achievable with a constant low QS q_l follows an inverse exponential function of the ratio q_l/q_h, and the dropping rate is slower with a longer Fz.

Regarding the analytical modeling, the quality degradation due to FR/QS variation follows an inverse exponential function of FR/QS ratio t_l/t_h (or q_l/q_h). The overall quality for a given t_l, t_h (or q_l, q_h) can be modeled by either the product of two exponential functions or one exponential function. The former one consists of one sub-model for constant FR/QS and the other for FR/QS variation. The later one utilizes one simple function, which is dependent on only t_l (or q_h), because the perceived quality in the lower quality segments dominates the overall perceived quality. Although the model performance of the later one is not as good as the former one, it require less number of parameters (one content-dependent α_{tv}/α_{qv}) and acceptable prediction accuracy with PCC value of around 0.96-0.97. Regarding the bit rate fluctuation due to the FR/QS variation, for a given r_{tl}/r_{th} (or r_{qh}/r_{ql}), we propose two models to capture the quality effect. Similar to the proposed quality model in terms of t_l/t_h (or q_l/q_h), we use the same function form ,but, in terms of r_{tl}/r_{th} (or r_{qh}/r_{ql}) instead. It is also interesting to observe that the falling trend of $\mathrm{MNQT}_{cr}(r; s, t)$

is more exponential law, but the falling trend of $\mathrm{MNQQ}_{cr}(r; s, q)$ is more like power law or sigmoid function. We also conducted three-way ANOVA to evaluate the statistical significance of the impact of FR/QS variation, changing interval and video content on the perceived quality.

Chapter 5

Model Parameter Prediction Using Content Features

As shown in previous chapters, our model parameters are video content dependent. The model will be very useful if the model parameters can be predicted from some content features derived from the original or compressed video sequences in this section. In this chapter, we explore various features for parameter prediction, and then present the stepwise feature selection approach for selecting a subset of features that when linearly combined can minimize the cross-validation error for all test sequences.

This chapter is organized as followed. Sec. 5.1 gives some introduction and description of the content features. Sec. 5.2 and 5.2 present the parameter estimation for model parameters obtained by VQMTQ model (in Chapter 2) and QSTAR model (in chapter 3), respectively. We conclude this chapter in Sec. 5.4

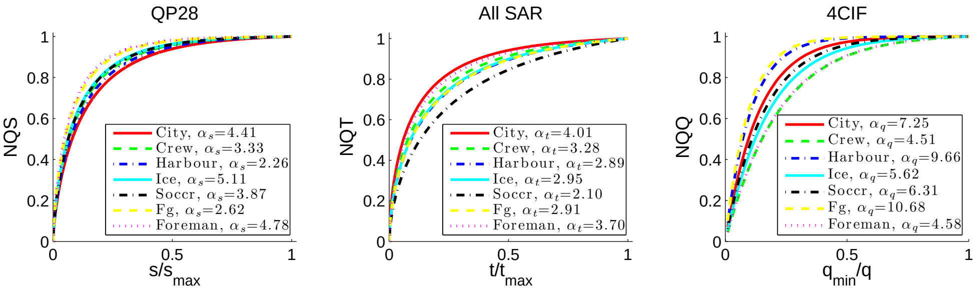


Figure 5.1: The content impact for NQS, NQT, NQQ.

5.1 Description of the Video Content Features

As shown in Fig. 5.1, the dropping rates of NQS, NQT, NQQ curves, and consequently the model parameters, $\hat{\alpha}_s$, α_t, α_q are sequence dependent. According this observation, we found that the dropping rate of the quality with TR, characterized by α_t, depends on the motion of the objects in the image. It becomes smaller (or model curves drops faster) for fast motion videos (e.g., *Ice*, *Soccer*) and vise verse. Based on the feedback from viewers, this is because that it's difficulty for the human eye in tracking moving objects in a video when the frame rate is reduced, and hence it should depend on features that reflect the temporal variation of the video. In addition, the dropping rate of the quality with SR and QS, characterized by α_s and α_q, respectively, are both depends on the spatial information of the images, such as texture details. For instance, the source sequences with smaller α_s and α_q tend to contain more details (e.g., *Foreman*, *Harbour* and *Soccer*), and hence are more susceptible to blurring introduced by up-sampling of high QS sequences. In the following, we define all the features that we have considered.

a) Frame Difference

A simple measure of the temporal variation of a video is the mean of the absolute difference between co-located pixels in successive frames, defined as FD.

b) Normalized Frame Difference

We note that a sequence with high contrast tends to have a large frame difference even with small motion, and vice verse. Therefore, we also define the normalized frame difference as

$$\mathrm{NFD} = \mathrm{FD}/\mathrm{STD}, \tag{5.1}$$

where STD stands for the average standard deviation of the pixel values in each frame and is used to measure the contrast of a video.

c) Motion Vector Magnitude (MVM)

Frame difference only measures the variances of co-located pixel values in successive frames, but it does not reflect the actual motion trajectory among successive frames. A more precise way to characterize the motion content is to evaluate motion vectors (MVs). In our study, MVs are extracted from bitstreams encoded by JSVM912 [60]. That is the faster searching method, which uses variable block sizes and maximum search range of 32 by 32 with quater-pel accuracy. In order to find the best matching blocks, we disable the rate-distortion optimization. We define the motion feature, MVM, as the mean of the motion vector magnitudes that are in the top 10% percentile.

d) Displaced Frame Difference (DFD)

DFD is the mean of the absolute difference between corresponding pixels in successive frames using estimated MV. When motion estimation is accurate, even when MVM is large, DFD could be small. Hence DFD could reveal whether the motion in the underlying sequence is complex and difficult to estimate.

e) Motion Activity Intensity (MAI)

In [47, 48], the authors proposed to use a motion activity intensity feature or MAI as the weighting parameter in their model (i.e. M_A in Eq.(5) proposed in [48]). This feature is defined as the standard deviation of motion vector magnitude, and is used as a MPEG-7 motion-descriptor [73].

f) MVM Normalized by the Contrast (MVM_STD)

Similar to normalized frame difference, we investigate several normalized motion vector features. The NMV_STD feature normalizes the MVM feature by the contrast, defined as

$$\mathrm{NMV\_STD} = \mathrm{MVM}/\mathrm{STD}, \tag{5.2}$$

g) MVM Normalized by Motion Activity Intensity (MVM_MAI)

Similar to the $\mathrm{NMV\_STD}$ in Eq. (5.2), we also normalize motion vector magnitude by motion activity intensity, i.e.,

$$\mathrm{NMV\_MAI} = \mathrm{MVM}/\mathrm{MAI}, \tag{5.3}$$

h) MVM Normalized by Variance of Motion Vector Direction (MVM_MDA)

Video sequences with higher motion vector magnitude as defined before do not necessarily have consistent large motion. It is possible that all the motion vectors are pointing to different directions. It is noted that human eye may be more sensitive to the motion with coherent direction. In other words, the human eye may observe the motion jitter more easily when the underlying video containing moving objects with consistent motion directions. We measure the motion direction incoherence by the variance of motion vector directions, where for a given MV with $\mathrm{MV_x}$ and $\mathrm{MV_y}$ as vertical and horizontal components, respectively, its direction is defined as

$$\theta_{MV} = \arctan(\mathrm{MV_x}/\mathrm{MV_y}),\ 0 \leq \theta_{MV} \leq 2\pi. \tag{5.4}$$

Here, we calculate the standard deviation of θ_{MV} and denote this feature as the motion direction activity or MDA. We further normalize the motion vector magnitude feature by MDA, yielding

$$\mathrm{NMV\_MDA} = \mathrm{MVM}/\mathrm{MDA}. \tag{5.5}$$

g) Gabor Texture (Gobar)

Judging from the results shown in Fig. 2.8 the parameter s is likely dependent on the contrast of the video and the amount of details present in a video. Notice that s is small for sequences with more details such as city, football, waterfall, and large for sequences with less details such as akiyo, ice, crew. To reflect the contrast of an image frame, we use the STD defined above, the standard deviation of gray level values. Figure 5.5 (a) shows that there is no consistent trend between s and STD.

In order to derive features that can reflect the amount of details in a video frame, we use the Log-Gabor filter [74]. It has been adopted to generate low level features for exploring visual attention, such as saliency map, foveate detection. The transfer function of Log-Gabor filter is constructed in term of two components, $F_m(w)$ and $F_n(\theta)$ with $scale$

$m = 1, 2, ...M$ and $orientation$ $n = 1, 2, ..., N$:

$$G_{mn}(w,\theta) = F_m(w) \cdot F_n(\theta) \tag{5.6}$$
$$F_m(w) = \exp(\frac{-\ln(w/w_{0,m})^2}{2(\ln(\sigma/w_{0,m}))^2})$$
$$F_n(\theta) = \exp(\frac{-(\theta-\phi_n)^2}{2\sigma_\theta^2}),$$

where $w_{0,m} = \frac{2}{m\ell}$, $\sigma_\theta = \frac{\pi}{N}$, and $\phi_n = \frac{(n-1)\pi}{N}$. In our study, we set $\ell = 3$ (pixels) $\sigma/w_{0,m} = 0.65$ following [75]. With $M = 2$, and $N = 2$, there are a total of four output images for each original image as shown in Figure 5.6. As can be seen, the Gabor filters with this parameter setting capture the horizontal and vertical edges in two different scales. We see that "City", "Football" and Waterfall" have much stronger responses than the other two images. We apply four Gabor filters (using the Matlab script from [75]) to 5 frames of each sequence that are uniformly sampled across the entire video, and find the mean and standard deviation of the absolute pixel values in each output image, and further average the resulting values from the four filters separately. These are denoted as $\mathrm{G_m}$ and $\mathrm{G_{std}}$. Finally we average $\mathrm{G_m}$ and $\mathrm{G_{std}}$ over all 5 frames to derive two Gabor features, $\hat{\mathrm{G}}_\mathrm{m}$ and $\hat{\mathrm{G}}_\mathrm{std}$, which measure the overall strength and variations of horizontal and vertical edges in a sequence. Figure 5.5(b) and (c) show the scatter plots of these two Gabor features with parameter s.

Table 5.1 lists all the content features included in the training procedure and their mathematical symbols. Note that each feature is derived from the original video signals at 4CIF and 30Hz.

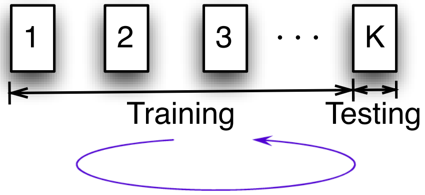


Figure 5.2: The Leave-One-Out Method

Table 5.1: The list of content features and the mathematical symbols

Features	Symbol Definition
FD/DFD	$\mu_{\mathrm{FD}}, \sigma_{\mathrm{FD}}, \mu_{\mathrm{DFD}}, \sigma_{\mathrm{DFD}}$
STD	σ
MVM/MDA/MAI	$\mu_{\mathrm{MVM}}, \sigma_{\mathrm{MDA}}, \mu_{\mathrm{MAI}}$
NFD	$\eta(\mu_{\mathrm{FD}}, \sigma) = \mu_{\mathrm{FD}}/\sigma$
NDFD	$\eta(\mu_{\mathrm{DFD}}, \sigma) = \mu_{\mathrm{DFD}}/\sigma$
NMV_STD	$\eta(\mu_{\mathrm{MVM}}, \sigma)= \mu_{\mathrm{MVM}}/\sigma$
NMV_MAI	$\eta(\mu_{\mathrm{MVM}}, \mu_{\mathrm{MAI}}) = \mu_{\mathrm{MVM}}/\mu_{\mathrm{MAI}}$
NMV_MDA	$\eta(\mu_{\mathrm{MVM}}, \mu_{\mathrm{MDA}}) = \mu_{\mathrm{MVM}}/\mu_{\mathrm{MDA}}$
Gabar	$\hat{\mathrm{G}}_{\mathrm{m}}, \hat{\mathrm{G}}_{\mathrm{std}}$

5.2 Parameter Estimation for VQMTQ Model

The scatter plots in Fig. 5.4 show that none of the features we considered correlate very well with parameter b, and several features had similar Pearson correlations. Therefore, we examined how to combine multiple features using the Generalized Linear Model [76]. Generally, the GLM using K features, f_k, $k = 1, 2, ..., K$, can be expressed as $\sum_k a_k f_k + a_0$. We use a stepwise feedforward approach to select the features. Specifically, we first choose one feature that minimizes a chosen error criterion. We then find the next feature, which, together with the first feature, has the largest reduction in the error. This process is repeated until the error does not reduce any more. In order for the solution to be generalizable to other sequences outside our test sequences, we use the leave-one-out cross-validation error (CVE) criterion. The main idea is to randomly pick one sequence for testing while the rest are for training as illustrated in Fig. 5.2. Assume the total number of sequences is M (In our case, $M = 7$). For a particular set of chosen features, we arbitrarily set one sequence as the test sequence and the remaining $M - 1$ sequences as the training sequences. We determine the weights a_k to minimize the mean square fitting error for the training sequences. We then evaluate the square fitting error for the test sequence. We repeat this process, each time using a different sequence as the test sequence. The average of the fitting errors for all the test sequences is the CVE associated with this feature set.

Recall the VQMTQ model proposed in chapter 2, i.e.,

$$\mathrm{VQMTQ}_1(\mathsf{PSNR}, f) = \hat{Q}_{\max}\left(1 - \frac{1}{1 + e^{p(\mathsf{PSNR}-s)}}\right)\frac{1 - e^{-b\frac{f}{f_{\max}}}}{1 - e^{-b}}. \tag{5.7}$$

$$\mathrm{VQMTQ}_2(q, f) = Q_{\max}\frac{e^{-c\frac{q}{q_{\min}}}}{e^{-c}}\frac{1 - e^{-d\frac{f}{f_{\max}}}}{1 - e^{-d}}. \tag{5.8}$$

Where b, s, c and d are model parameters. We will demonstrate the prediction results in the following sections.

5.2.1 Prediction of Model Parameter b

For parameter b, we consider the following features described earlier in Sec. 5.1Figure 5.4 shows the scatter plots of these features vs. parameter b, and also provides the Pearson correlation of each feature with the parameter.

Using the stepwise feature selection approach described in Sec. 5.2, we found that using the features MDA and DFD yields the lowest CVE. The final weighting coefficients are determined by minimizing the average square fitting error among all 7 sequences, which yields

$$\hat{b} = 10.72 - 0.6 \cdot \mu_{\mathrm{MDA}} - 0.13 \cdot \mu_{\mathrm{DFD}} \tag{5.9}$$

Table 5.2 lists the PCC and CVE of parameter b associated with each single feature and the predictor given in (5.9). Figure 5.3 shows the TCF curves obtained using predicted b values. We see that they fit with the measured normalized MOS quite well, only slightly worse than the results obtained when the parameter b is derived by fitting the TCF curve with the measured MOS data.

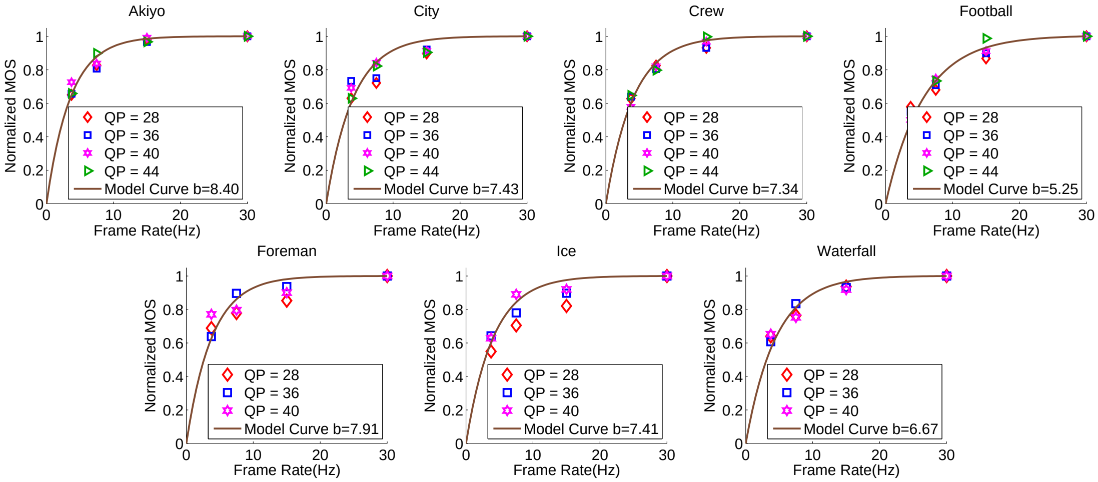


Figure 5.3: The measured normalized MOS and temporal correction factor (TCF) against frame rate using predicted b.

5.2.2 Prediction of Parameter s

For parameter s, we consider all the features introduced in Sec. 5.1. This yields a predictor involving two features, $\hat{\mathrm{G}}_{\mathrm{m}}$ and NFD:

$$\hat{s} = 34.8298 - 1.88 \cdot \hat{\mathrm{G}}_{\mathrm{m}} - 2.23 \cdot \eta(\mu_{\mathrm{DFD}}, \sigma) \tag{5.10}$$

Figure 5.5(d) shows that $\hat{s}$ is highly correlated with parameter s. Table 5.3 summarizes the PCC and CVE of parameter s.As can be seen the SQF matches with the measured MOS at the highest frame rate very well, almost as good as the SQF when the parameter s is obtained by fitting.

5.2.3 Model Verification Using Predicted b, s, c, and d

Figure 5.7 (left part) illustrates the predicted quality by the proposed model (Eq. 2.11) when the parameters b and s are predicted using Eq.(5.9) and Eq.(5.10) and measured MOS. Table 5.4 summarizes the model performance in terms of PCC and RMSE. Compared to previous Fig. 2.12, 2.21 and Table 2.2, we see that the predicted quality match with the measured MOS very well, only slightly worse than those obtained with parameters that are derived by fitting the model to the measured MOS directly.

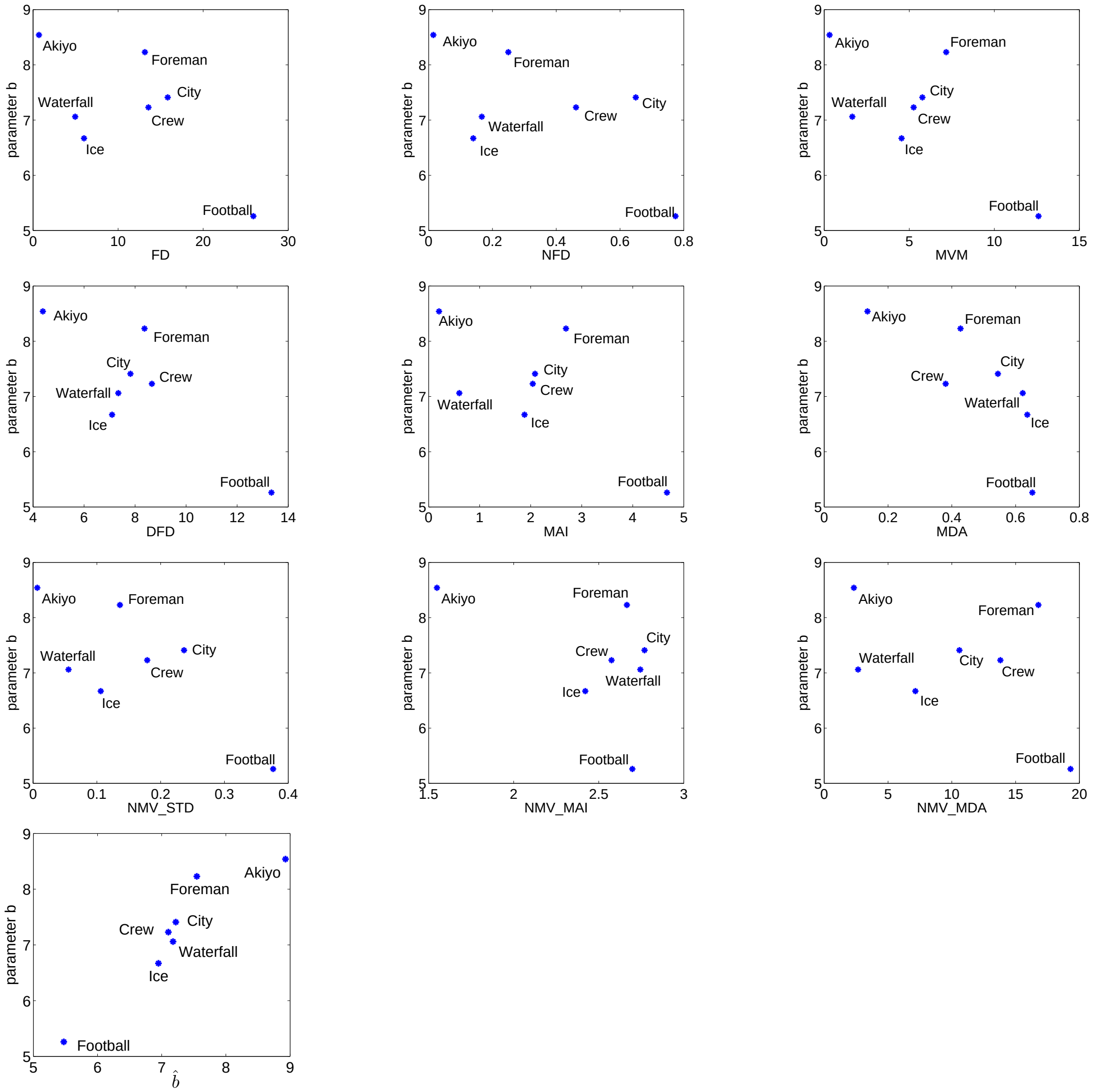


Figure 5.4: Relation between parameter b and the feature values for different sequences. The Pearson Correlation (PCC) coefficients between "b" and individual feature are (a) PCC=-0.66, (b) PCC=-0.64, (c) PCC=-0.69, (d) PCC=-0.78, (e) PCC=-0.7, (f) PCC=-0.78, (g) PCC=-0.74, (h) PCC=-0.51, (i) PCC=-0.43, (j) PCC=0.94

Table 5.2: Fitting Accuracy for Parameter b and different features and the proposed predictor

	Dataset #1	
	PCC	CVE
FD	-0.66	0.67
NFD	-0.64	0.72
MVM	-0.69	0.6
DFD	-0.78	0.47
MAI	-0.7	0.6
MDA	-0.78	0.48
NMV_STD	-0.74	0.53
NMV_MAI	-0.51	0.9
NMV_MDA	-0.43	1.5
$\hat{b}$	0.94	0.16

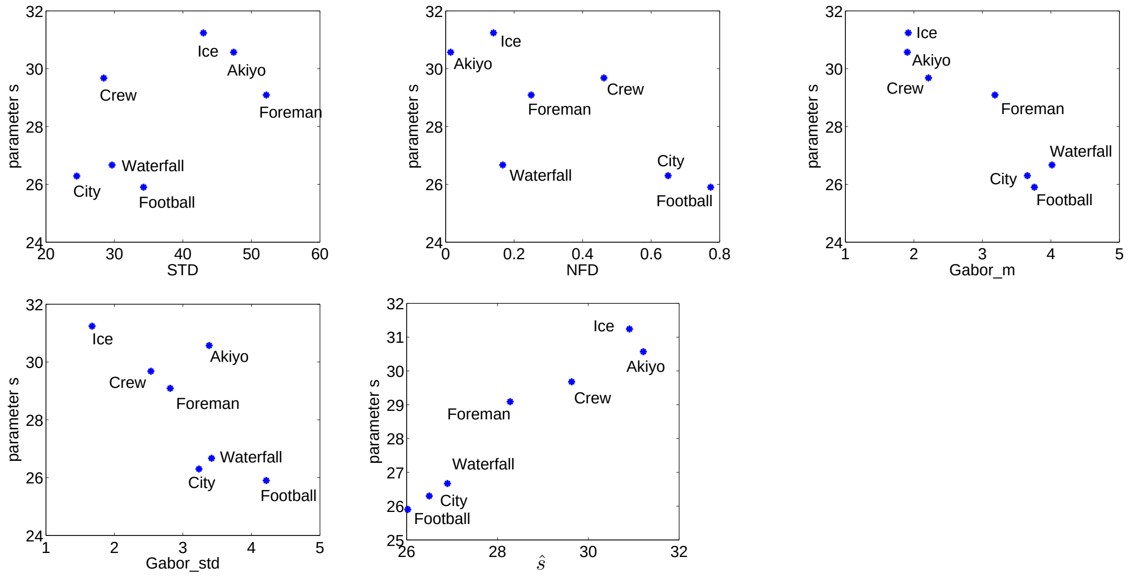


Figure 5.5: Relation between parameter s and the feature values for different sequences. The Pearson Correlation (PCC) coefficients between ”s” and individual feature are (a) PCC=0.63, (b) PCC=-0.7, (c) PCC=-0.94, (d) PCC=-0.76, (e) PCC=0.97

Table 5.3: Fitting Accuracy for Parameter s and different features and the proposed predictor

	Dataset #1	
	PCC	CVE
STD	0.63	3.25
NFD	-0.7	2.71
$\hat{\mathrm{G}}_{\mathrm{m}}$	-0.94	0.63
$\hat{\mathrm{G}}_{\mathrm{std}}$	-0.76	1.67
$\hat{s}$	0.97	0.21

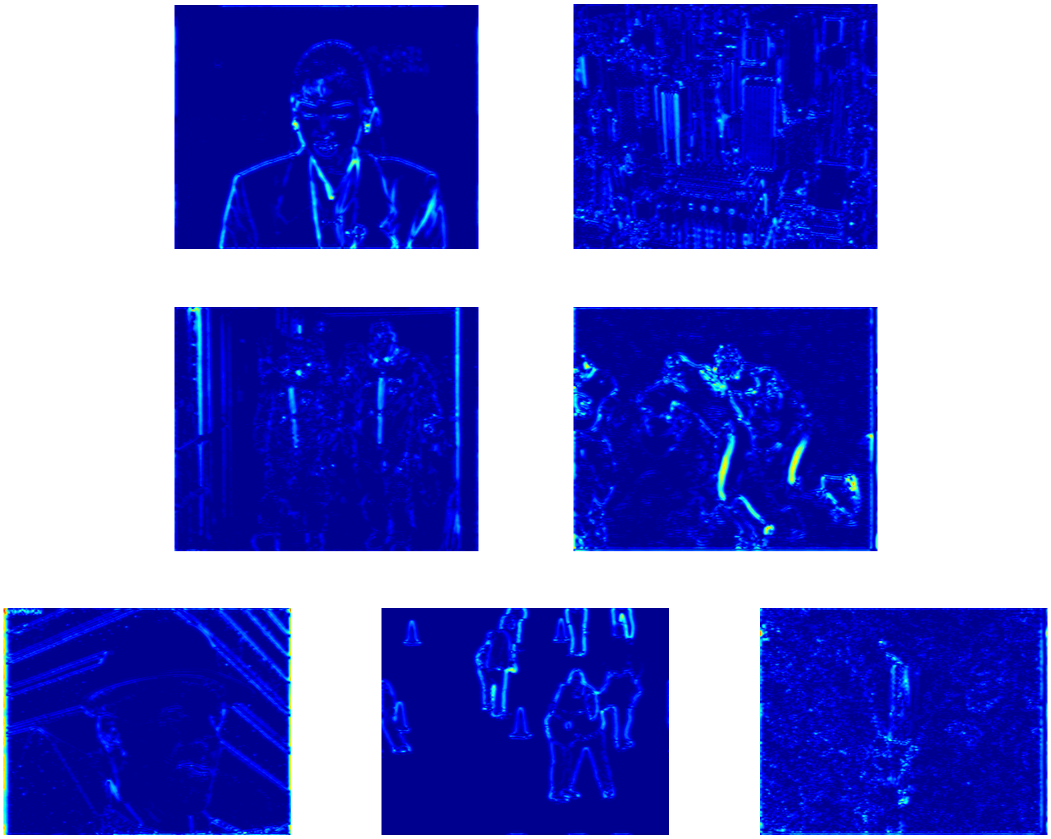


Figure 5.6: The Gabor texture map of each sequence, where $m = 2, n = 2$

Table 5.4: Goodness of fit by VQMTQ model using predicted parameters

	akiyo	city	crew	football	foreman	ice	waterfall	Ave
	Estimated parameters of $\mathrm{VQMTQ}_1(\mathsf{PSNR}, f)$							
$\hat{s}$	31.21	26.49	29.63	26.02	28.28	30.90	26.89	
$\hat{b}$	8.40	7.43	7.34	5.25	7.90	7.40	6.66	
RMSE	3.92%	5.31%	2.3%	3.95%	8.61%	7.1%	3.23%	4.94%
PCC	0.99	0.97	0.99	0.98	0.90	0.95	0.98	0.97
	Estimated parameters of $\mathrm{VQMTQ}_2(\mathrm{q}, \mathrm{f})$							
$\hat{c}$	0.14	0.13	0.12	0.09	0.12	0.12	0.13	
$\hat{d}$	8.01	7.33	6.87	5.59	7.01	6.84	7.60	
RMSE	5.10%	6.19%	7.02%	4.24%	6.80%	7.35%	4.49%	6.00%
PCC	0.98	0.94	0.96	0.97	0.93	0.95	0.97	0.95

Similar to the prediction of parameter b and s, we also find the optimum predicted c and d by minimizing CVE. Table 5.4 summarizes the model performance in terms of PCC and RMSE. By utilizing parameter predictor, we can predict the perceptual quality of a video when coded using a chosen (t, q) combination automatically. In practical encoding applications, since we can access the original source signal at encoder side, it is easy to estimate the model parameter from its corresponding content features.

5.3 Parameter Estimation for QSTAR Model

Recall that the proposed QSTAR model in chapter 3 is defined as

$$\begin{aligned}\mathrm{QSTAR}(s, t, q) &= \mathrm{MNQQ}(q; s_{\max})\mathrm{MNQS}(s; q)\mathrm{MNQT}(t)\\ &= \frac{1 - e^{-\alpha_q(\frac{q_{\min}}{q})}}{1 - e^{-\alpha_q}} \frac{1 - e^{-\hat{\alpha}_s L((\mathrm{QP}(q))(\frac{s}{s_{\max}})^{\beta_s}}}{1 - e^{-\hat{\alpha}_s L(\mathrm{QP}(q))}} \frac{1 - e^{-\alpha_t(\frac{t}{t_{\max}})^{\beta_t}}}{1 - e^{-\alpha_t}},\end{aligned} \tag{5.11}$$

and we use a subset of features introduced in Sec. 5.1 to build predictors for the model parameters (e.g., $\hat{\alpha}_s$, α_t, α_q). As for parameter prediction for the VQMTQ model described in Sec. 5.2, we use a GLM [76] to predict each parameter from multiple features. There, the features to be included and the predictor coefficients for different parameters are determined separately. However, in this section, we choose to find a minimal feature set that can predict all parameters simultaneously and accurately.

Let $p_{l,m}$, $m = 1, 2, ..., M$, $l = 1, 2, ..., L$, denote the l^{th} parameter of m^{th} sequence and $f_{k,m}$, $k = 1, 2, ..., K$, the k^{th} feature of m^{th} sequence, $p_{l,m}$ is predicted using a generalized linear predictor, $g_{l,0} + \sum_{k=1}^{K} g_{l,k} f_{k,m}$. This predictor can be written in a vector form, i.e.,

$$\hat{\mathbf{P}}_m = \mathbf{G}\mathbf{F_m}, \tag{5.12}$$

where $\hat{\mathbf{P}}_m = [\hat{p}_{1,m}, \hat{p}_{2,m}, ..., \hat{p}_{L,m}]^{\mathrm{T}}$ contains the predicted parameters for sequence m, $\mathbf{G}$ is a $L{\times}(K{+}1)$ matrix containing coefficients $g_{l,k}$ and $\mathbf{F_m} = [\mathbf{1}, \mathbf{f_{1,m}}, \mathbf{f_{2,m}}, ..., \mathbf{f_{K,m}}]^{\mathrm{T}}$ contains the features of sequence m.

In order to find the optimum solution for the feature set and corresponding $\mathbf{G}$ that can minimize the prediction error and be generalizable to other sequences outside our test sequences, we use the leave-one-out cross-validation error (CVE) criterion. Let $\mathcal{F}^{\mathcal{K}}$ denote all the possible feature sets with a number of K features. For a particular set of chosen K features $\gamma_K \in \mathcal{F}^{\mathcal{K}}$, we arbitrarily set one source sequence as a test sequence (i.e., m_t) and remaining (M-1) sequences as training sequences (i.e., Γ). We determine the optimal predictor matrix $\mathbf{G}$ by minimizing the fitting error $\mathcal{E}_\Gamma$ for (M-1) training sequences, defined as $\mathcal{E}_\Gamma = \frac{1}{M-1} \sum_{m \in \Gamma} \left\|\hat{\mathbf{P}}_m - \mathbf{P_m}\right\|^{\mathbf{2}}$. We then find the predicted model parameter $\hat{\mathbf{P}}$ using the previously determined $\mathbf{G}$ for the test sequence, and evaluate the fitting error $\mathcal{E}_{m_t}$ which is the sum of the quality difference absolute value under all STAR combinations for this sequence. We repeat this process, each time using a different sequence as the test sequence, and find the average of all fitting errors, $\mathcal{E}_{\gamma_\mathcal{K}} = \frac{1}{M} \sum_{m_t=1}^{M} \mathcal{E}_{m_t}$, associated with this feature set γ_K. For a given K, the set of features that leads to the least CVE $\mathcal{E}_{\gamma_\mathcal{K}}$ is chosen. We evaluate the CVE starting with $K = 1$ and increase K until the minimal CVE does not reduce significantly. The resulting K features are the final feature set chosen. We then re-compute the predictor matrix $\mathbf{G}$ to minimize the average parameter fitting error over all the sequences, i.e., $\frac{1}{M} \sum_{m=1}^{M} \left\|\hat{\mathbf{P}}_m - \mathbf{P_m}\right\|^{\mathbf{2}}$.

Using this procedure, we found that four features, σ_{DFD}, σ, μ_{MDA}, $\eta(\mu_{\mathrm{MVM}}, \mu_{\mathrm{MAI}})$,

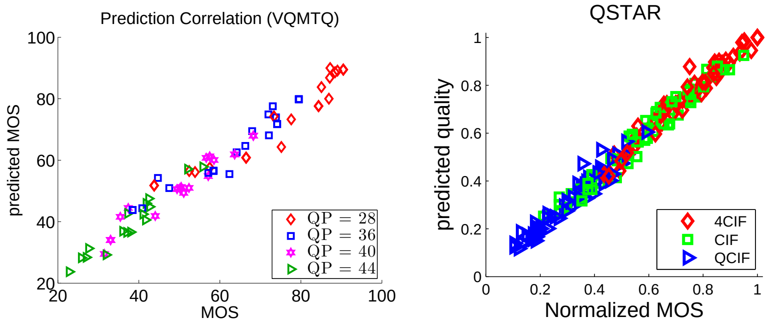


Figure 5.7: Predicted quality against measured MOS. Left: Model parameters b and s are predicted using Eq.(5.9) and Eq.(5.10) with PCC=0.97 for VQMTQ model.; Right: Model parameters, $\hat{\alpha}_s, \alpha_t, \alpha_q$ are predicted from content features with PCC=0.98 for QSTAR model.

can accurately predict 3 model parameters $\hat{\alpha}_s, \alpha_t, \alpha_q$, with the following predictor matrix:

$$\mathbf{G} = \begin{bmatrix} -1.1586 & -0.1161 & 0.0817 & 1.4706 & 1.0795 \\ 2.3749 & -0.3763 & 0.0142 & 0.3601 & 1.0728 \\ 23.8838 & 0.4797 & -0.1039 & -10.1363 & -4.2012 \end{bmatrix} \tag{5.13}$$

Table 5.5: The Predicted parameters and performance of QSTAR model.

	city	crew	harbour	ice	soccer	fg	foreman	Avg.
	Parameters obtained by least square fitting with MOS data							
RMSE	0.018	0.025	0.038	0.033	0.032	0.058	0.038	0.035
PCC	0.998	0.996	0.992	0.993	0.992	0.979	0.991	0.991
	Parameters predicted from video content features							
α_q	7.45	4.29	9.17	6.15	6.37	10.76	4.38	
$\hat{\alpha}_s$	3.68	4.07	3.85	4.33	4.55	5.11	5.56	
α_t	3.99	3.35	2.76	2.92	2.15	2.95	3.69	
RMSE	0.017	0.023	0.039	0.054	0.030	0.060	0.032	0.036
PCC	0.997	0.995	0.988	0.989	0.991	0.977	0.992	0.987

We plot the scatter plot of predicted quality using the model parameters estimated from the video content with measured data in Fig. 5.7 (right part). Table 5.5 (lower half) summarizes the model performance and shows that the four-feature prediction provides a very high PCC of 0.988 and a small RMSE of 0.036, very close to the performance of the model whose parameters are from least square fitting (see Tab. 5.5). Recall that in Sec. 5.2, each model parameter requires at least two content features and a total of five features are need for estimating two parameters. Here, we only need four features to predict three parameters by finding the optimal features and the predictor matrix for all parameters jointly.

5.4 Summary

In practice, our proposed models will be more useful if we can estimate the model parameters via the underlying video contents. Also, based on our observations from previous chapters, we have found that the model parameters are indeed content dependent. In order to automatically predict the model parameters from original video signals, we first abstract useful content features, and develop a lightweight pre-processor to obtain those features. In general, we have considered the features related to the residual signal such as frame difference, displace frame difference etc, motion fields, such as motion vector magnitude, motion direction activity, etc and original video signal, such as video contrast. Different feature combinations are examined in our study to show the parameter prediction accuracy. By applying the GLM with CVE criteria on combinations of all content features, the simulation results show that stepwise selection of two-feature combined prediction for each parameter provides the accurate estimation for CIF video while four-feature exhausted-search prediction for all parameters is appropriate for QSTAR model. For different video resolution, different feature combinations and weighted functions will be applied. We also compare the predicted quality using estimated model parameters with the measured data and show that both the VQMTQ and QSTAR models still have high accuracy using predicted parameters.

Chapter 6

Conclusions

In this dissertation, our goal was to investigate how to estimate perceptual video quality for video transmission under heterogeneous network environment using a simple quality metric considering temporal, spatial and quantization artifacts. In this chapter, we first summarize our major contributions, and address some possible future works.

6.1 Summary and Conclusions

While perceptual video quality assessment has attracted significant attention in recent years, objective quality assessment is in its early phase. The research discussed and models proposed in this dissertation are important supplements to the existing research work in literature. The major contributions of this thesis are listed below.

Chapter 2: This work is concerned with the impact of quantization and frame rate on the perceptual quality of a video. We conducted subjective ratings of a total of 100 video sequences coded at different frame rates and quantization stepsizes from seven source videos of different characteristics. The videos are displayed on a native size (CIF resolution) in laptop monitor. We demonstrate that the degradation of the perceptual quality due to the increase of QS and reduction of FR can be accurately captured by the product of two functions, a spatial quality factor that reflects the impact of quantization, and a temporal correction factor that reveals the impact of frame rate.

For the temporal correction factor, an inverse exponential function of the normalized frame rate can accurately reflect the impact of frame rate reduction on the perceived quality. For the spatial quality factor, we established three possible models. The first one employs a sigmoidal function of the PSNR, the second one uses the exponential function of the normalized QS, and the third one adopts the inverse exponential function of the inverse of the normalized QS (which is equivalent to the normalized amplitude resolution). The model using PSNR is useful for evaluating the quality of coded video, whereas the models using QS is useful for rate control in video encoding and for adaptation of pre-coded scalable video. Each model function has a single parameter that is video-content dependent. The proposed model is shown to be highly accurate, compared to the subjective ratings from our own subjective tests as well as test results reported in several other papers. Although for the subjective test data we have, the two proposed spatial quality models using QS have similar accuracy, the inverted exponential model should be more applicable in applications involving a large range of QS.

Chapter 3: The model developed in Chapter two considering only the impact of TR and QS is extended to consider the individual and joint effect of SR, TR, and QS. We conducted subjective ratings over 189 videos coded at different combinations of SR, TR, and QS, created from 7 source videos with wide range of video contents. The videos are displayed on mobile display platforms so that our model can be more applicable in mobile video applications. Subjective tests are conducted on mobile display platforms so that our model can be more applicable in mobile video applications In the proposed model, we use a one-parameter inverse exponential function to capture the quality decay v.s. SR, TR and QS individually. The parameter in each function is sequence dependent. The overall model is the product of these three functions (i.e., MNQS, MNQT, MNQQ). We further found that the model parameter of MNQT is statistically independent SR and QS, but the parameters of MNQS and MQNQ depend on both SR and QS. The model with the content-derived features has a high PCC (=0.988) with subjective ratings. In addition to MNQQ model, we further propose MNQP and MNQR when QS is not available. They hold the same merit of MNQQ that the falling trend is independent of TR but SR. Both MNQP and MNQR are derived using inverted exponential function, which approximate the measured data very well with PCC=0.98. The proposed models are further validated on subjective ratings reported in eight other databases.

Chapter 4: We conducted subjective experiments to investigate the impact of periodic FR/QS variation on the perceived video quality. We observed several interesting trends, such as under the same average FR/QS, a video with a constant FR/QS is perceptually more appealing than a video with FR/QS variation; the quality of a video with alternating FR (t_h,t_l) is generally equal or better than a video with a constant FR t_l for fast motion content, and the lower the t_l, the larger the improvement; while the quality of a video with alternating QS (q_h, q_l) is usually equal or better than a video with constant QS q_h, and the higher the q_h, the larger the improvement. We further found that and inverse exponential function of FR ratio t_l/t_h can accurately approximate the quality degradation from a video with constant FR=t_h. Similarly, an inverse exponential function of the QS ratio q_l/q_h can accurately capture the quality degradation from a video with constant QS=q_l. Each model has a PCC with the subjective ratings around 0.97. We further examined the relation between the quality degradation and the bit rate fluctuation due to FR/QS variation, and found that the quality degradation can also be accurately modeled by an inverse exponential function of the bit rate ratio (r_{tl}/r_{th} or r_{ql}/r_{qh}). Finally, we conducted three-way ANOVA to evaluate the statistical significance of the impact of FR/QS variation, changing interval and video content on the perceived quality.

Chapter 5: The parameters of the models presented in the previous chapters depend on the characteristics of the video content. While developing these models, the parameters for each source video are determined by least squares fitting with the subjective ratings for all the processed versions of this source. For these models to be applicable to other videos for which no subjective data are available, we must be able to predict the model parameters from content features computed from original or processed video. In this chapter, we consider how to predict the parameters for the VQMTQ model and the QSTAR model from content features computed from original sources. We considered a variety of content features that can reveal the motion and texture characteristics of a video. By an exhaustive search among all possible combinations of these features under a cross validation error (CVE) criterion, we found that the two parameters of the VQMTQ model can be accurately predicted from a generalized linear model using two features only, whereas the three parameters of the QSTAR model can be accurately predicted using four features. Both models still have high correlation with subjective ratings of our test videos when using predicted model parameters.

6.2 Future Work

In this section, we present some ideas for future extensions of the work developed in this dissertation.

First, although the proposed models, i.e., QSTAR, is developed for videos generated by the H.264/SVC codec, we expect that the same function form is applicable to scalable videos coded using other codecs and to non-scalable videos coded at different (s,t,q) combinations. However, the model parameters for the same video content may differ, depending on the encoder configurations. This hypothesis needs to be validated in future studies.

Second, The proposed quality models, together with the rate model, also as a function of STAR in [71], can be used to determine the optimal STAR that maximizes the quality given a rate constraint, both for video encoding/transcoding and for scalable video adaptation. Our prior work [54, 72] has investigated a subset of this problem, where SR is fixed, and only TR and QS are adapted, based on quality and rate models as functions of TR and QS only. Extension of this work to include the SR dimension, using the newly developed quality and rate models, both as functions of SR, TR, and QS, is another interesting direction for future research.

Third, the work in Chapter 4 investigates the impact of periodic frame rate/quantization changes on the perceptual quality individually under several changing intervals. Future studies will examine the impact of other factors including wider range of changing interval, variation of spatial resolution, and joint impact of FR, QS and spatial resolution variation. One challenge issues in modeling the impact of temporal/quantization variation of the STAR on the quality is how to design the subjective test to help understand the effect of likely variation patterns. Our subjective study that involves periodic FR/QS variations is only the first step towards this direction. Based on findings from these as well as other subjective tests, we hope to model the video quality over a certain duration with STAR variation, as a function of the average STAR as well as some statistics characterizing their temporal variation.

Fourth, based on our simulations in Chapter 5, the content feature set is quite stable for various videos. however, the weighting coefficients for the chosen features vary largely, which brings the large variation when we introduce new test video. It might be helpful to do the feature normalization before conducting the generalized linear regression.

Bibliography

[1] Zhongkang Lu, Weis Lin, Boon Choong Seng, S. Kato, Susu Yao, Eeping Ong, and X. K. Yang, "Measuring the Negative Impact of Frame Dropping on Perceptual Visual Quality", in *Proc. SPIE Human Vision and Electronic Imaging*, Jan. 2005, vol. 5666, pp. 554–562.

[2] H.-T. Quan et al., "Temporal Aspect of Perceived Quality of Mobile Video Broadcasting", *IEEE Trans. on Broadcasting*, vol. 54, no. 3, pp. 641–651, Sept. 2008.

[3] R. Feghali et al., "Video quality metric for bit rate control via joint adjustment of quantization and frame rate", *IEEE Trans. on Broadcasting*, vol. 53, no. 1, pp. 441–446, Mar. 2007.

[4] D. Wang, F. Speranza, A. Vincent, T. Martin, and P. Blanchfield, "Towards Optimal Rate Control: A Study of the Impact of Spatial Resolution, Frame Rate, and Quantization on Subjective Video Quality and Bit Rate", in *SPIE Proc. on Visual Communication and Image*, 2003, vol. 5150, pp. 198–209.

[5] G. Zhai, J. Cai, W. Lin, X. Yang, W. Zhang, and M. Etoh, "Cross-Dimensional Perceptual Quality Assessment for Low Bit-Rate Videos", *IEEE Trans. on Multimedia*, vol. 10, no. 7, pp. 1316–1324, Nov. 2008.

[6] Cheon Seoe Kim, Suh Dongjun, Tae Meon Bae, and Yong Man Ro, "Measuring Video Quality on Full Scalability of H.264/AVC Scalable Video Coding", *IEICE Tran. on Communications*, vol. E91-B, no. 5, pp. 1269–1275, May 2008.

[7] Video Quality Experts Group (VQEG), "VQEG Final Report on the Validation of Video Quality Models for High Definition Video Content", 2010.

[8] J.-S. Lee, F. De Simone, and T. Ebrahimi, "Subjective quality evaluation via paired comparison: application to scalable video coding", *IEEE Transactions on Multimedia*, vol. 13, no. 5, pp. 882–893, 2011.

[9] C. J. van den Branden Lambrecht and M. Kunt, "Characterization of Human Visual Sensitivity For Video Imaging Applications", *Signal Processing*, vol. 67, no. 3, pp. 255–269, 1998.

[10] S. Daly, "The Visible Differences Predictor: an Algorithm for the Assessment of Image Fidelity", in *Human Vision, Visual Processing, and Digital Display III*, 1992, pp. 179–206.

[11] J. Lubin, "A Human Vision System Model for Objective Picture Quality Measurements", in *IEEE Proc. on Broadcasting*, 1997, pp. 498–503.

[12] E. M. Yeh, A. C. Kokaram, and N. G. Kingsbury, "Psychovisual Measurement and Distortion Metics for for Image Sequences", in *Proc. of European Signal Processing Conference*, 1998, vol. 2, pp. 1061–1064.

[13] Andrew B. Watson, James Hu, and John F Mcgowan Iii, "Dvq: A digital video quality metric based on human vision", *Journal of Electronic Imaging*, vol. 10, pp. 20–29, 2001.

[14] Video Quality Experts Group, "VQEG Subjective Test Plan", Tech. Report, Available [online]: <http://ftp.crc.ca/test/pub/crc/vqeg/>.

[15] S. Winkler, "Issues In Vision Modeling For Perceptual Video Quality Assessment", *Signal Processing*, vol. 78, no. 2, pp. 231–252, 1999.

[16] Rec. ITU-R BT.500-11, "Methodology for the Subjective Assessment of the Quality of Television Pictures", 2002.

[17] Rec. ITU-T P.910, "Subjective Video Quality Assessment Methods for Multimedia Applications", 2001.

[18] M. Pinson and S. Wolf, "Psychovisual Measurement and Distortion Metics for for Image Sequences", in *SPIE Proc. of Visual Communications and Image Processing*, 2003, vol. 5150, pp. 573–582.

[19] A. Webster, "Final Report From the Video Quality Experts Group on the Validation of Objective Models of Multimedia Quality Assessment, Phase 1", Tech. Report, 2008, Available [online]: <ftp://vqeg.its.bldrdoc.gov/Documents/Projects/multimedia/MM_Final_Report/>.

[20] R. Hamberg and H. Ridder, "Time-varying Image Quality: Modeling the Relation between Instantaneous and Overall Quality", *SMPTE Journal*, vol. 108, pp. 802–811, Nov. 1999.

[21] D. Hands, "Temporal Characterization Of Forgiveness Effect", *Electronic Letter*, vol. 37, pp. 752–754, 2002.

[22] M. H. Pinson and S. Wolf, "Techniques for Evaluating Objective Video Quality Models using Overlapping Subjective Data Sets", Tech. Rep. TR-09-457, NITA, Nov. 2008.

[23] M.P. Eckert and A.P. Bradley, "Perceptual quality metrics applied to still image compression", *Signal Processing*, vol. 70, pp. 177–200, 1998.

[24] B. Girod, "What's Wrong With Mean-Squared Error", *Digital images and human vision*, pp. 207–220, 1993, MIT Press, Cambridge, Massachusetts.

[25] Z. Wang, L. Lu, and A. C. Bovik, "Video quality assessment using structural distortion measurement", in *Proc. of ICIP*, Jun. 2002, pp. III–65–68.

[26] M. H. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality", *IEEE Trans. on Broadcasting*, vol. 50, no. 3, pp. 312–322, Sep. 2004.

[27] VQEG, "Final report on the validation of models of video quality assessment, FRTV Phase II", Tech. Report, 2003.

[28] ITU-T Recommendation J.144, "Objective Perceptual Video Quality Measurement Techniques For Digital Cable Television In The Presence Of A Full Reference", 2004.

[29] ANSI T1.801.03, "American National Standard for Telecommunications−Digital transport of oneway video signals−Parameters for objective performance assessment", 2003, American National Standard Institute Report.

[30] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack, "Study of Subjective and Objective Quality Assessment of Video", *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1427–1441, 2010.

[31] H. R. Sheikh and A. C. Bovik, "A Visual Information Fidelity Approach to Video Quality Assessment", *The First International Workshop on Video Processing and Quality Metrics for Consumer Electronics*, 2005.

[32] K. Seshadrinathan and A. C. Bovik, "Motion Tuned Spatio-temporal Quality Assessment of Natural Videos", *IEEE Transactions on Image Processing*, vol. 19, no. 2, pp. 335–350, 2010.

[33] VQEG, "Test Sequences of Video Quality Experts Group validation of models of video quality assessment, Phase 1", 2000, Available [online]: <http://www.its.bldrdoc.gov/vqeg/projects/frtv_phaseI/>.

[34] A. A. Webster, C. T. Jones, M. H. Pinson, S. D. Voran, and S. Wolf, "An Objective Video Quality Assessment System Based On Human Perception", in *SPIE Proc. of Human Vision, Visual Processing, and Digital Display IV*, 1993, vol. 1913, pp. 15–26.

[35] P. Bretillon, N. Montard, J. Baina, and G. Goudezeune, "Quality Meter And Digital Television Applications", in *SPIE Proc. of Visual Communications and Image Processing*, 2000, vol. 4067, pp. 780–790.

[36] M. Carnec, P. Le Callet, and D. Barba, "New Perceptual Quality Assessment Method With Reduced Reference For Compressed Images", in *SPIE Proc. of Visual Communications and Image Processing*, 2003, vol. 5150, pp. 1582–1593.

[37] L. Itti, C. Koch, and E. Niebur, "A Model of Saliency-Based Visual Attention for Rapid Scene Analysis", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, Nov. 1998.

[38] L. Itti and C. Koch, "Computational Modeling of Visual Attention", *Nature Reviews Neuroscience*, vol. 2, no. 3, pp. 194–203, Mar. 2001.

[39] D. Parkhurst, K. Law, and E. Niebur, "Modeling the role of salience in the allocation of overt visual attention", *Vision Research*, vol. 42, no. 1, pp. 107–123, Jan. 2002.

[40] W. Osberger and A. Maeder, "Automatic identification of perceptually important regions in an image using a model of the human visual system", in *International Conference on Pattern Recognition*, Aug. 1998, vol. 1, pp. 701–704.

[41] X. Feng, T. Liu, D. Yang, and Y. Wang, "Saliency Inspired Full-Reference Quality Metrics for Packet-Loss-Impaired Video", *IEEE Transactions on Broadcasting*, vol. 51, no. 1, pp. 81–88, Mar. 2011.

[42] W. Lin, E. Ong, S. Yang, and X. K.Yang, "Perceptual-quality significance map (PQSM) and its application on video quality distortion metrics", *2003 IEEE International Conference on ICASSP*, vol. 3, pp. III– 617–20, Apr. 2003.

[43] J. Y. C Chen and J. E Thropp, "Review of Low Frame Rate Effects on Human Performance", *IEEE Trans. on Systems, Man and Cybernetics*, vol. 37, pp. 1063–1076, Nov. 2007.

[44] Y. Wang, S-F. Chang, and A. Loui, "Subjective Preference of Spatio-Temporal Rate in Video Adaptation Using Multi-Dimensional Scalable Coding", in *Proc. of ICME'04*, Jun. 2004, vol. 3, pp. 1719–1722.

[45] G. Yadavalli, M. Masry, and S. S. Hemami, "Frame Rate Preference in Low Bit Rate Video", in *Proc. of ICIP*, Nov. 2003, vol. 1, pp. I–441–4.

[46] J. McCarthy, M. A. Sasse, and D. Miras, "Sharp or smooth ?: Comparing the effects of quantization vs. frame rate for streamed video", in *Proc. of ACM CHI on Human Factors in Computing Systems*, Apr. 2004, pp. 535–542.

[47] K.-C. Yang, C. C. Guest, K. El-Maleh, and P. K Das, "Perceptual Temporal Quality Metric for Compressed Video", *IEEE Trans. on Multimedia*, vol. 9, pp. 1528–1535, Nov. 2007.

[48] Sung Ho Jin, Cheon Seog Kim, Dong Jun Seo, and Yong Man Ro, “Quality Measurement Modeling on Scalable Video Applications”, in *Proc. of IEEE Workshop on Multimedia Signal Processing*, Otc. 2007, pp. 131 – 134.

[49] I-Hsien Lee, Sheng-Chieh Huang, Chung-Jr Lian, and Liang-Gee Chen, “A Quality-of-Experience Video Adaptor for Serving Scalable Video Applications”, *IEEE Trans. on Consumer Electronics*, vol. 53, pp. 1130–1137, Aug. 2007.

[50] Gert Hauske, Thomas Stockhammer, and Rolf Hofmaier, “Subjective Image Quality of Low-Rate and Low-Resolution Video Sequence”, in *in Proc. of International Workshop on Mobile Multimedia Communications*, 2003, pp. 5–8.

[51] H. Wu, M. Claypool, and R. Kinicki, “ARMOR - A System for Adjusting Repair and Media Scaling for Video Streaming”, *Elsevier Journal of Visual Communication and Image Representation (JVCIR)*, vol. 19, no. 8, pp. 489–499, Dec. 2008.

[52] E. Ong, X. Yang, W. Lin, Z. Lu, and S. Yao, “Perceptual Quality Metric For Compressed Videos”, in *Proc. of ICASSP*, Mar. 2005, vol. 2, pp. 581 – 584.

[53] Yen-Fu Ou, Zhan Ma, and Yao Wang, “Perceptual Quality Assessment of Video Considering both Frame Rate and Quantization Artifacts”, *IEEE Transaction on CSVT*, vol. 21, pp. 286–298, 2011.

[54] Y. Wang, Z. Ma, and Y.-F. Ou, “Modeling rate and perceptual quality of scalable video as functions of quantization and frame rate and its application in scalable video adaptation”, in *IEEE 17th Packet Video Workshop*, 2009, pp. 1–9.

[55] Yuanyi Xue, “Perceptual Quality Assessment of H.264/SVC on Spatial Resolution And Quantization”, Master Thesis, 2010.

[56] Yen-Fu Ou, Yuanyi Xue, Zhan Ma, and Yao Wang, “A Perceptual Video Quality Model for Mobile Platform Considering Impact of Spatial, Temporal, and Amplitude Resolutions”, in *Image, Video, and Multidimensional Signal Processing (IVMSP) on Perception and Visual Signal Analysis*, Jun. 2011.

[57] Y.-F Ou et al., “A novel quality metric for compressed video considering both frame rate and quantization artifacts”, in *Proc. of Intl. Workshop Video Processing and Quality Metrics for Consumer (VPQM)*, Scottsdale, AZ, Jan. 2009.

[58] V. Baroncini, D. Ciavatta, G Gaudino, R. Felice, G. Iacovoni, and F. Ubaldi, “Variable Frame Rate Video For Mobile Devices”, in *Proceedings of the 2nd international conference on Mobile multimedia communications*. 2006, pp. 35:1–35:6, ACM.

[59] Joint Video Team, ”, Available [online]: http://wftp3.itu.int/av-arch/jvt-site/.

[60] *Joint Scalable Video Model*, "Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, Doc. JVT-X202", Jul. 2007.

[61] Y.-F. Ou, T. Liu, Z. Zhao, Z. Ma, and Y. Wang, "Modeling The Impact of Frame Rate on Perceptual Quality of Video", in *Proc. of ICIP*, San Diego, Oct. 2008, pp. 689 – 692.

[62] S. Wolf and M. Pinson, "Video Quality Measurement Techniques", Tech. Rep. 02-392, NTIA, Jun. 2002.

[63] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity", *IEEE Trans. on Image Processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[64] TI Zoom2 MDP, ", Available [online]: http://www.omapzoom.org/platform.html/.

[65] Google Android, ", Available [online]: http://www.android.com/.

[66] G. Sullivan and S. Sun, "AHG Report on Spatial Scalability Resampling", Joint Video Team of, ISO/IEC MPEG & ITU-T VCEG, Document: JVT-Q007, Oct. 2005.

[67] E. Francois and et al., "Generic Extended Spatial Scalability", Oct. 2004.

[68] A. M. van Dijk, J. B. Martens, and A. B. Watson, "Quality Assessment of Coded Images using Numerical Category Scaling", in *Proc. SPIE*, 1995, vol. 2451, p. 90101.

[69] G. W. Snedecor et al., *Statistical Methods, 8th ed Ames*, IA: Iowa State Univ. Press, 1989.

[70] *Text of ISO/IEC 14496-10:2005/FDAM 3 Scalable Video Coding*, "Joint Video Team (JVT) of ISO-IEC MPEG & ITU-T VCEG", N9197, Lausanne, Sep. 2007.

[71] Zhan Ma, Meng Xu, and Yao Wang, "Modeling Video Rate as a Function of Frame Size, Frame Rate and Quantization Stepsize", in *Proc. of IEEE MMSP*, Oct. 2011.

[72] Zhan Ma, Meng Xu, Kyeong Yang, and Yao Wang, "Modeling of Rate And Perceptual Quality of Video And Its Application To Frame Rate Adaptive Rate Control", in *Proc. of IEEE ICIP*, 2011.

[73] S. Jeannin and A. Divarakan, "MPEG-7 visiual motion descriptors", *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 11, no. 6, pp. 720 – 724, Jun 2001.

[74] B. S. Manjunath and W. Y Ma, "Texture features for browsing and retrieval of image data", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 18, no. 8, Aug. 1996.

[75] P. D. Kovesi, "MATLAB and Octave functions for computer vision and image processing", School of Computer Science & Software Engineering, The University of Western Australia, Available from: <http://www.csse.uwa.edu.au/∼pk/research/matlabfns/>.

[76] P. McCullagh and J. A. Nelder., *Generalized Linear Models*, New York: Chapman & Hall, 1990.