PERCEPTUAL VIDEO QUALITY MODEL AND ITS APPLICATION IN WIRELESS MULTIMEDIA COMMUNICATIONS

A Dissertation
Submitted to the Graduate Faculty
of the
North Dakota State University
of Agriculture and Applied Science

By

Abdul Hameed

In Partial Fulfillment of the Requirements
for the Degree of
DOCTOR OF PHILOSOPHY

Major Department:
Electrical and Computer Engineering

May 2015

Fargo, North Dakota

North Dakota State University
Graduate School

Title

Perceptual Video Quality Model and Its Application in Wireless Multimedia Communications

By

Abdul Hameed

The Supervisory Committee certifies that this ***disquisition*** complies with North Dakota State University's regulations and meets the accepted standards for the degree of

DOCTOR OF PHILOSOPHY

SUPERVISORY COMMITTEE:

Prof. Sudarshan Srinivasan
Co-Chair

Prof. Rui Dai
Co-Chair

Prof. Benjamin Balas

Prof. Ivan Lima

Approved:

May 8, 2015
Date

Prof. Scott Smith
Department Chair

ABSTRACT

With the exponential growth of video traffic over wireless networked and embedded devices such as mobile phones and sensors, mechanisms are needed to predict the perceptual quality of video in real time and with low complexity, based on which networking protocols can control video quality and optimize network resources to meet the quality of experience requirements of users. This thesis is composed of three related pieces of work. In the first piece of work, an efficient and light-weight video quality prediction model through partial parsing of compressed from the H.264/AVC compressed bitstream is proposed. A set of features were introduced to reflect video content characteristics and distortions caused by compression and transmission and were obtained directly in parsing mode without decoding the pixel information in macro-blocks. Based on the features, an artificial neural network model was trained for perceptual quality prediction. In the second piece of work, a perceptual video quality prediction model is trained based on massive subjective test results. Prediction of perceptual quality is achieved through a decision tree using a set of easily calculated features from the compressed bitstream and the network. Moreover, based on the prediction model, a novel Forward Error Correction (FEC) scheme is introduced to protect video packets by taking into consideration video content characteristics, compression parameters, as well as network condition. Given a perceptual quality requirement, the error control scheme adjusts the level of protection for different components in a video stream such that the network overhead needed for transmission is minimized. In the third piece of work, a study was conducted to examine whether the previous prediction model could provide a good confidence measure in a different domain of judgments. The accuracy of judgements demonstrated the predictive validity of confidence measure with respect packet loss ratio traits. The results of this study were consistent with the previous one and

the experiments suggested that brief and evaluative thin slice judgments are made relatively intuitively. Present research represents a new entry into the domain of high level judgments, such as video confidence measure by the use of our existing perception quality model.

ACKNOWLEDGMENTS

Completing a Ph.D. has been an unforgettable journey in my life. It consists of extensive literature surveys, scientific exploration, and creative innovations. During this process, I received immeasurable help and inspiration from many people.

First of all, I would like to express my deepest gratitude to my advisor, Professor Rui Dai, for her excellent guidance, generous support, and helpful discussions during the studies toward my doctoral degree at North Dakota State University. I have learned a great deal from her expertise on the topic and sharp insight from different perspectives. I really appreciate her elaborate review of all my writings, timely advice, technical comments, many encouragements, and efforts for providing me with an excellent atmosphere for doing research. I would also like to thank Professor Benjamin Balas, who provided me a golden opportunity to work with his students for subjective quality tests and later helped me with experience the research of video confidence measure and address practical issues beyond the area of perceptual video quality. I would also like to thank Professor Sudarshan Srinivasan and Professor Ivan Lima, who kindly agreed to serve in my dissertation defense committee and for their precious time to attend the defense in the midst of their busy activities. Special thanks to Professor Rajendra Katti for his insightful comments, guiding my research, and helping me to develop my background in engineering, computer science, and psychology.

I would also like to thank my loving parents, caring siblings, inspiring laboratory members, and all amazing friends for making my life during the past four years a fun, challenging, and memorable one. They were always supporting me and encouraging me with their best wishes. I would never have been able to finish my dissertation without their constant support and encouragement.

DEDICATION

To my family and friends for their endless love, limitless encouragement, and unselfish sacrifice throughout my doctoral education.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

CHAPTER 1. INTRODUCTION

In the field of multimedia communications, video coding, and transmission over the unreliable network in real-time with the limited computational resources of transmission device is a challenging problem. During the process of acquisition, communication, and processing of videos, various types of distortions are introduced that may have major impact on video quality perceived by human observers. Moreover, the computational resources (or power consumption) on many embedded devices (such as smart phones and wireless sensors) for encoding are very restricted and the wireless communication channels can be very unreliable. We are also witnessing a change in paradigm towards integrating the end user at the terminals as the important factor in the perceptual video quality assessment. This shift of paradigm drives the creation of the Quality of Experience (QoE) concept, which is a customer service experience reflecting the blueprint of all human quality requirements and expectations [1]. QoE is purely subjective measure from the user's perspective measured at the end devices and can conceptually be seen as the perceived quality after the distortion through the network. The video researchers are facing a real challenge in fulfilling the requirements of the end user whose ultimate goal is to watch a perceptually good quality video sequence in real-time.

H.264/AVC (Advanced Video Coding) standard developed by ITU-T Video Coding Experts Group (VCEG) and ISO/IEC JTC1 Moving Picture Experts Group (MPEG) is one of the most widely used video codec for video compression and transmission [2]. It promises increased visual quality and reduced bandwidth and is based on an efficient hybrid video coding concept with the advantage of modern arithmetic algorithms that makes it a good choice for off-line video encoding. However, the packet losses in the wireless network cause a significant obstacle to utilize H.264/AVC for real-time applications. Due to the high computational complexity of the encoding

process and high sensitivity of the video bitstream to packet losses, various artifacts can happen that can degrade video quality [3]. To measure video quality, Peak Signal-to-Noise Ratio (PSNR) metric is utilized due to its simplicity but PSNR is a poor predictor of perceived visual quality and does not align well with human visual perception [4]. Therefore, developing better perceptual visual quality models to predict the evaluation of an observer to replace non-perceptual criteria is required in the field of video processing.

1.1. Factors related to video quality

The factors that affect perceptual video quality are either associated with source video or terminal or both. The relationship between different perceptual quality metrics are very complex and are independent of one another. The factors that can affect the source video are focus, brightness, contrast, and camera performance. The choice of codec and encoding parameters such as, bit depth, resolution, frame rate, and frame complexity in the form of spatial and temporal information also affects perceptual quality. The video quality may be degraded by compression artifacts (source distortion) and transmission errors (channel distortion) in the form of packet losses. Degradation can also occur at decoder end and during playback time. The viewing conditions such as display device, viewing distance, and ambient illumination also affect perceptual video quality [5]. In general, the video quality is affected during any processing stage of video acquisition, processing, compression, transmission, decoding, and display the coded video data that can introduce artefacts that reduce the perceptual quality [6].

During the transmission, the effect of packet loss is different based on the content, motion, and location of lost packet in videos. Even low rates of packet loss can cause severe degradation as the aforementioned factors have a different impact on perceptual quality. Moreover, a packet loss occurred in intra frames (I-Frame) and predictive inter frames (P-Frame) will propagate in the

whole Group of Pictures (GOP) while a packet loss occurred in bidirectional predictive frame (B-Frame) does not propagate in next frames in GOP [7]. Figure 1 explains how errors propagate through the different frame types in a GOP. An increasing number of multimedia applications are supported with the improvements in source coding, storage capacity, and network infrastructures. Most of lossy compression techniques operate on data that will be perceived by human consumers; therefore, the distortion measure should be modeled on human perception.

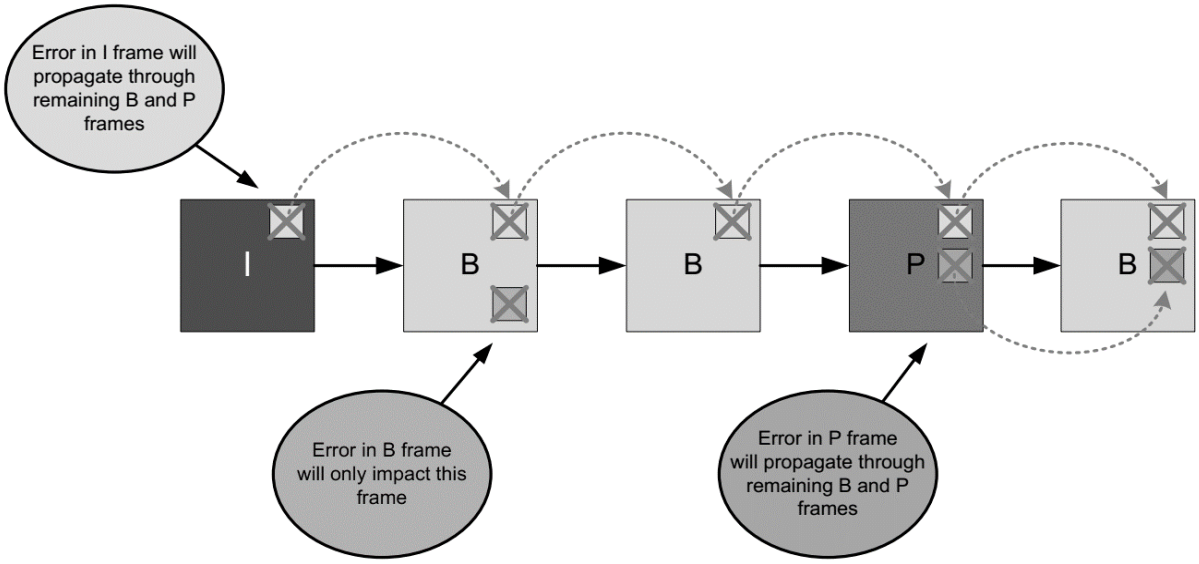

Figure 1. Impact of frame structure on error propagation

1.1.1. Video source coding

The source coding process maps information symbols to alphabetical symbols by eliminating redundancy from the source symbols. A signal *s* is produced at the source, which is mapped into bitstream *b* at the source encoder before the transmission of the bitstream over the error control channel. The source decoder processes the received bitstream *b'* and reconstructs the decoded signal *s'* before transporting it to the sink. Figure 2 represents a block structure for a typical transmission scenario. Due to the lossy nature and limited bandwidth of the communication systems, the video encoder compresses the video sequence and tries to make encoded sequence resilient to errors [4]. The source coding is an important part of communication system as it helps

efficient use of disk utilization and transmission bandwidth by reducing data size. It can either be lossy, where exact reconstruction of source symbols is not possible, or lossless, where error free reconstruction of source symbols is possible. The examples of coding parameters include the coded bit rate, the spatial resolution, and the temporal resolution [8].

Some encoding parameters may affect the resultant reconstructed perceptual video quality directly during video compression process. We briefly demonstrate how the parameters may affect the encoded/decoded video quality. The video quality increases upon increasing bitrate. The visible artifacts in pixel blocks and at block boundaries are visible in low bitrates are due to block transform coding. The lower the bit rate, the more coarsely the coefficients are represented and result in the discontinuities at the block boundaries. The spatial distortion results in block distortion whereas the temporal distortion effects can be seen as the reduction in frame rate or a frozen picture, resulting in a jerky picture and a loss of smoothness in moving objects [9]. There are also spatio-temporal distortion effects including disruption of the video signal.

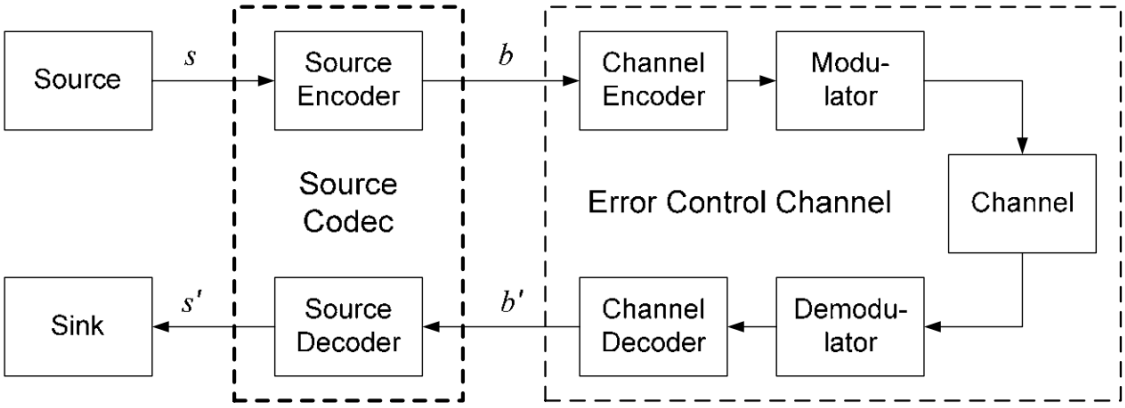

Figure 2. Typical structure of a transmission system

1.1.2. Network related video degradation and channel coding

In addition to compression artifacts (source distortion), video quality may be degraded by transmission errors (channel distortion), usually appearing in the form of packet losses. Moreover,

the wireless channels are noisy and the bitstreams delivered may be received imperfectly due to the impaired artifacts occurred by fading, multi-path, and shadowing resulting in a higher bit error rate (BER) and lower throughput [10]. The transmission delay can also fluctuate based on the network congestion when the video packets are transmitted over a network. The congestion process can result in packet loss, the effect of which is the degradation of perceptual video quality. The wireless data transmissions have higher variations in bandwidth and delay that causes high packet losses that results in frame loses in the output video.

The channel coding deals with error control techniques and allows the decoder to determine whether the received word is a valid code word or corrupted by noise. The process of error control coding detects and possibly correct errors by introducing redundancy to the stream of bits to be sent to the channel. The data transmission reliability is amplified by adding redundant information symbols with the reduction in information rate before the decoder identifies the code word sent after noise corruption. The process either allows an increased rate of information transfer at a fixed error rate, or a reduced error rate for a fixed transfer rate. In either scenarios, if the output data of a communications system has errors that are too frequent for the desired use, the errors can often be reduced. For wireless networks, the channel errors can broadly be categorized into three forms: (1) packet loss, (2) packet truncation, and (3) bit error. The former two are caused by network traffic and clock drift and latter is an effect of noisy channel [1] [3].

1.2. H.264/AVC

In H.264/AVC, a macroblock (MB) is a 16 x 16 displayed pixel area and is the basic unit of encoding and decoding process in data processing. To produce a compressed H.264/AVC bitstream, the first step during the encoding is the generation of a prediction MB from the edge pixels of neighboring previously-decoded MBs. The next process is the subtraction of prediction

MB from the current MB to generate a residual MB that results in less noticeable inconsistencies between neighboring blocks. After that, the transformation is done to de-correlate the data spatially following which is the process of the quantization of MBs using a quantization control parameter just before the encoding [11].

The process of prediction, transforming, and encoding is done with in video encoder. There is a mini-decoder with in H.264/AVC where the rescaling of the quantized data and inverse transformation process take place. The next step is to add the transformation to predict MB so that the reconstruction of the frame for later predictions is possible. Figure 3 and Figure 4 represent the whole H.264/AVC encoding and decoding process [2]. During the whole process, a number of coefficients, and parameters are generated. These syntax elements are related to quantization transformation, coefficients to recreate prediction, and information related to compressed data and compression tools before the conversion to an efficient and compact binary bitstream. Moreover, the process of variable length coding and arithmetic coding compression algorithms on these elements further compress the data before storing and transmission over the network.

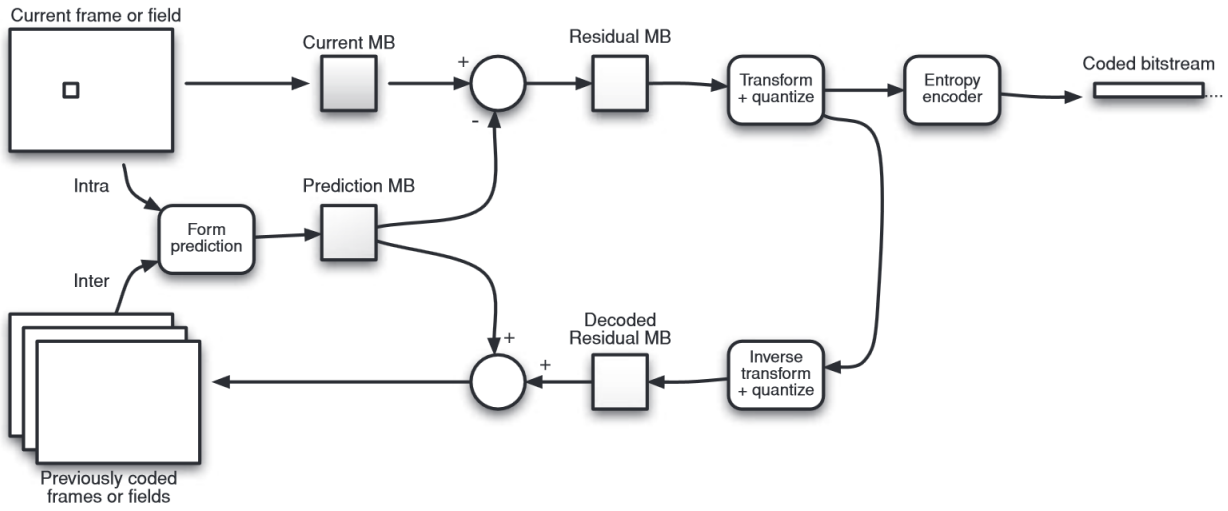

Figure 3. H.264/AVC encoding process

H.264/AVC provides a format and syntax that are powerful features for compressed video and offers real-time efficient decoding procedure for all profiles and levels supported by the standard. It enables the user by delivery of high-quality video at very low data rates and offers a set of mechanism for the delivery of effective and flexible video compression for a wide range of applications, from low delay and/or minimal storage memory applications to high-definition broadcast services [11]. An H.264 video decoder identifies the requirements to decode the bitstream and follows the decoding order of inverse transform and reconstruction processes. The process include the parsing and extraction of the data elements to produce a decoded video sequence.

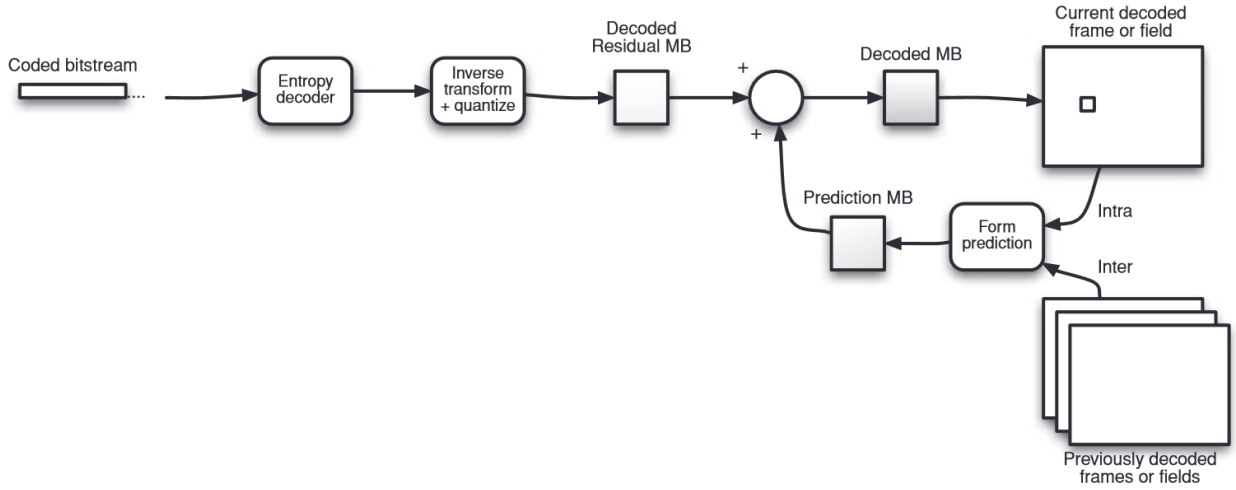

Figure 4. H.264/AVC decoding process

1.3. Video quality assessment methods

The significance of perceptual video quality has amplified with the growing interests with the increasing demand for multimedia services, visual processing algorithms, and video applications. The perceptual quality assessment aims at quantifying the quality of visual information in meeting the promised quality of service (QoS) and has gone through a wide range of developments and still growing. In the last decade, a number of reliable video quality assessment methods and metrics were proposed with different computational complexity and accuracy to

improve the end user's QoE [4]. The goal of designing assessment methods is to evaluate perceptual quality automatically during the different phases of design, implementation, optimization, and testing.

Any automatic visual quality assessment method plays an important role by offering the evaluations to compensate and can be used as an alternative for extensive subjective evaluations in the form of mean opinion score (MOS). However, due to the intrinsic nature of perceptual quality being subjective, it is hard to attain an accurate quality metric due to dependency on many features. All of these factors make it very difficult to accurately and quantitatively measure visual quality. Traditional video quality assessment approaches could be divided into two kinds: (1) objective video quality methods and (2) subjective video quality methods [6].

1.3.1. Objective video quality and its limitations

To analyze the compression and transport schemes and predict perceived video quality automatically, the objective quality metrics that are purely computational are proposed in the last few years. One of the goals of objective metrics is to adjust quality on the fly by monitoring the dynamic conditions of network. Moreover, such metrics can be used as a benchmark for future video processing systems and can be used to optimize the parameters for reconstruction and error concealment techniques. Depending on the type of input data used for evaluation, ITU standardization classified the objective quality measurement methods broadly into the following five main classes [12]: (1) Media-layer, (2) Parametric packet layer, (3) Parametric planning, (4) Bitstream-layer, and (5) Hybrid. Such methods are useful tools for video database systems and are also desired for a broad variety of video applications.

The objective video quality metrics do not provide a measure of the quality of video as perceived by the end user. The distortion caused by source and channel are of different types and

objective measures do not estimate reliably the relative qualities accurately when both types of distortions are present in different proportions [4]. Moreover, objective metrics do not take the temporal relation present in a video sequence. The industry wide standard Peak Signal to Noise Ratio (PSNR) also does not correlate with the subjective perception of quality. The limitations of crude metrics such as PSNR and the complexity of modern day objective methods is quite high due to the dependency on spatial and temporal features that accommodate spatial and temporal fidelity. However, the Video Quality Model (VQM) developed by ITS [3] accommodate perceptual effects of video impairments, still has lower computational cost and its results are comparable to the subjective evaluations. VQM considers the characteristics of the video applications, size formats, and the gaps in decoding process.

1.3.2. Subjective video quality assessment

Martínez et.al. suggested that subjective video quality evaluation in the form of MOS is the most consistent and reliable form of the video quality, which is a measured based on number of human assessors or test subjects [4]. The reliable approach to assess video quality perceived by usually non-expert human observers is to ask for the opinion for the evaluation of video quality. Subjective experiments involve a panel of participants and must be laid out rigorously. The factors that also affect any participant's opinion is the extent of interaction, viewing environment, viewing distance, test duration, room illumination, and the current state of mind. The subjective quality assessments are done in a controlled environment where evaluators' selection criteria need to be met before the start of experiments [12]. Moreover, test materials need to be selected carefully to have reliable and consistent results that can reflect the correlation between subjective and objective assessments. The aforementioned process is cumbersome, inconvenient, slow, and expensive for most applications but still used for ground truth data that can be used to evaluate and compare with

any video quality evaluation algorithm [12]. Though tedious, when conducted properly the subjective video quality assessment approaches are more accurate than the objective ones and attain the ultimate goal of matching human perception.

The subjective models are used for decades in telephony networks and serve as a benchmark for the performance evaluation of objective models. Subjective assessments are unable to provide instantaneous measurements but can provide valuable data to assess the performance of objective or automatic methods. The degradation of video quality may also impact other aspects of the viewer's subjective experience of a video, including their ability to make judgements about the content of the video. A systematic analysis of relationship between thin slice judgments and subjective quality has not been sufficiently explored. To investigate the significance of the perceptual quality affecting parameters and their relationships to evaluate the association of observers' estimate of speaker confidence and rating confidence measures is explored. The goal of this study is to understand whether the changes in perceptual quality directly impact how observers make thin-slice judgements based on the behaviors in short video sequences of 20 seconds duration.

1.4. Video confidence

The term video confidence is defined by measuring observed qualities that indicate situations where a perceiver senses and experience an instinct that a target is highly conscientious. The observed qualities are called justifiers that are perceivers' responses and provides a mindful justification of instinctive intuitions [13]. At the perceiver level, the potential source of confidence is person perception self-efficacy and can be accomplished by (a) cautious use of logic and evidence (b) the clear explanation for an implicit feeling or (c) link to explicit theory of conscientiousness. The research suggest that confidence measures fluctuate among people in the

same environment and that evaluators may have different ranks of confidence measures depending upon the context or situation. Therefore any individual who is very confident in a certain environment, may lose confidence in a new or unacquainted setting.

In literature self-efficacy and confidence are used interchangeably to define confidence but there are few differences between them in terms of definition and practical reasons. The scope of self-efficacy is limited to domain specific whereas the scope, predictions, and accuracy of confidence is not limited to a single or a specific domain and may be extended to extensive educational and social psychological reals [13]. Confidence measures the trust and expectation to accomplishment of any task and is abstracted by the differences in opinions to an extent to have firm trust and strong belief about the task accomplishment [14].

1.4.1. Thin slice

Thin-slicing refers to the process where very quick decisions are made with minimal amounts of available information. The process produces distinct biological and sociocultural effects and is an unconscious behavior. A visual thin slice is a brief silent selection of expressive behavior sampled from the behavioral stream. Visual thin slices offer an active and communicative action that permit relatively precise interpretations into the psychological and social life of target. Moreover, we expected to identify several factors that tend to increase confidence without raising accuracy, including both judgment-level factors (such as the target reminding the perceiver of a type of person) and judge-level factors (such as person perception self-efficacy). Person perception research has shown that consistent and accurate assessments of the traits can be made based on very brief observations, or “thin slices”. The end results would demonstrate levels of confidence in the impressions and vary considerably from judgment to judgment and from perceiver to perceiver.

1.4.2. Thin slice vision

Thin-slice vision is a viewer's ability to precisely determine the useful and accurate information about a target audience from very short-lived visual interaction. The research on thin-slice vision has largely focused on the detection of individual-level traits. Based on “thin slices” for consistent judgments, the reliable and precise behavior evaluations are possible in different problem areas. Thus, examining personal impressions based on thin slices offers an effective approach to how perceptions translate into real-world results. Personality is visible via thin slice and perceivers show note-worthy validity in gauging target personality based on the thin slices of evidence about a target. The psychological self depends on different factors that can enhance or limit thin-slice vision and the strangers might differ in inferring personal characteristics from very brief visual exposure.

1.4.3. High level tasks

The high level tasks such as personality traits, sexual orientation, relationships, and popularity or any nonverbal behavior are investigated in thin slice vision domain. Since, visual thin-slice impressions are different from auditory thin-slice impressions. Therefore, no prior knowledge of target, context, length, and verbal information to thin slices are required for thin slices. Moreover, the upper limit on the durations of thin slices are under 5 minutes and are normally closer to 30 seconds. In our case, a high level task is to find accuracy in measuring confidence measure of individuals by limited calibration between accuracy and confidence in first impressions.

1.5. Quality-of-experience (QoE) communication protocols

The QoE as defined by ITU means the overall acceptability of an application or service as *perceived subjectively* by the end user [5]. There has been considerable work published to evaluate

the QoE of video streaming and impact of packet loss in video transmission over IP networks. A plethora of research consider the effect on video quality with respect to the percentage of IP packet losses. In dynamic wireless networks the channel conditions can quickly change that may result in high error rates. In high bit-rate applications, the transmission of intra frames is required for resynchronization but the use of the intra-coding mode is constrained in low bit rate transmission systems as intra frames typically require several times more bits than inter frames. There are different possible patterns for packet losses both with a random distribution and with taking into account the effect of bursts.

The QoE-aware schemes are relatively less researched and one of the challenges is to adapt and optimize the distribution of real-time videos with QoE support for multimedia design and deliver in wireless/mobile multimedia applications. The traditional theoretical models are based on QoS parameters such as delay, jitter, and packet loss and have limitation to comprehend the quality affecting features and the relationships to accurately estimate QoE. To support QoE, a few models focused on encoder based distortions [15] [16] [17] [18] only, while others are based on a representative set of video content features [19] [20]. A very few models considered only network distortions [21] [22] and a very few considered both content and distortions caused by encoder [23] [24] [25]. The perceived video quality is affected by parameters associated with encoder and the network [26]. Traditional adaptation schemes do not take perceived quality into consideration because the inter-relationships between the application and network parameters and how the packet loss effects the perceptual quality are not well understood.

1.6. Error control for video communication

Real networks are heterogeneous and are unreliable due to changing bandwidth, transmission delays, and error characteristics due to various kind of noise, distortions, and

interferences. The increase in transmission bandwidth and computing power proliferate the interests in video communications in wireless networks. Due to variable bandwidth, video data retransmission is not a good choice for real-time systems as bitstream is sensitive to transmission errors [27]. The wireless channel error patterns differ from wired ones and may be unidirectional where receiver are passive devices that may result in no or few feedbacks. The wireless connections operate in environments that may change drastically and result in poor quality, lower bandwidth, less connection stability, and higher error rates [28][29]. The wireless channels can support high error rates and are vulnerable to corruption by noise.

In the reconstructed video quality, the bitrate can vary widely depending on the target application. Therefore, video bitstream is protected with different level based on the types of frame being encoded [30]. The error control mechanism are deemed necessary for keeping the Bit Error Rate (BER) low in robust multimedia communications. Moreover, packet loss conditions due to network congestion add further problems to have effective throughput and can disturb the transport of compressed video. Recently, the focus is more on simple and facile error control schemes on a secured transport level error control (channel coding) with error resilient approaches that involve error detection, prevention, and correction to help the decoder recover from transmission errors.by adding redundant packets in the bitstream. By choosing different options in architecture, we get different error control schemes, such as Forward Error Correction (FEC), Automatic Retransmission Request (ARQ), Error Resilient Packetization and multiplexing, and Unequal Error Protection (UEP) can be utilized in practice [31] [32] [33]. Many factors affect QoE during encoding, delivery, adaptation, decoding and playout stages and can be related to network parameters, video characteristics, bitrate, codec type, the length of the GoP, video content (the degree of spatial details and motion intensity) [34] [35].

1.6.1. Forward error correction (FEC)

FEC is one of the channel coding techniques where redundant bits called parities are added to the compressed video bitstream to enable error detection and correction to deliver video data to end users. The studies have shown that adding n-k packets to k source packets will results in n packets to be sent on the network and can detect t errors and correct up to t/2 errors. FEC uses additional parity data added to the actual information to correct varying error conditions experienced by receivers. FEC codes transform some number of equal length k symbols into n symbols, where n > k, by adding (n-k) additional symbols, called parity symbols. An FEC code can reconstruct any k lost or corrupted symbols of the n symbols as any subset of k encoded blocks suffices at the receiver to reconstruct the source data.

Analysis over a range of network conditions indicate that adjusting FEC with quality scaling to protect crucial part of information provides significant performance improvement for the worst-case channel characteristics[36]. Adding FEC to allow reconstruction of corrupted packets reduces the effective transmission rate of the original video content especially if a channel has a very low PSNR, the overall performance of wireless link may not improve with the addition of parity bits [37]. Therefore, selecting the optimum amount of FEC scheme in presence of packet loss and the corresponding quantization level improves the video quality to reconstruct the original video for the target users.

Typical FEC techniques are static in nature and may assure a certain QoS requirement given the same amount of error-control resources but adaptive FEC schemes with QoE guarantees are needed to provide error recovery in dynamic and high error rate networks. In wireless networks, traditional FEC schemes cannot counter the problem of the effects of random and burst packet losses due to the fact that the redundancy is based on averaged packet loss rate. Modern day FEC approaches add redundant information to the certain bit-stream areas based on the channel burst

packet loss that provide some degree of additional reliability [9]. Moreover, the research has shown that packet level FEC implementation is more practical than bit level implementation and the amount of redundancy added to intra frame has a different affect from that of inter frame [38].

1.6.2. Unequal error protection for video packets

In any network, depending on the frame type not all the bits are equally important during the transmission of video data. Therefore, to decrease packet errors, different protection levels are provided to different bit planes due to sensitivity to errors. Unequal Error Protection (UEP) add redundancy to the subsets of information depending upon the importance of bits on the basis of a loss pattern feedback to adapt the changes in network condition. In wireless networks, interactive application, and control systems, providing more (less) protection to the most (least) important subset information to recover the lost packets is resourceful. Traditional FEC approaches do not take into account the relative perceptual importance of packets and may occupy a large bandwidth that can lead to additional packet loss due to network congestion with the change in PLRs and network traffic load.

The optimized UEP with non-uniform error correction capabilities is one of the promising techniques that outperforms the traditional schemes by providing better perceptual quality with the efficient usage of network resources [39]. In the network, the packets that contain the important portions of the bitstream, such as the transitions between two different video scenes, the loss of which cannot be easily concealed, are perceptually significant than the packets representing a static video scene. The effectiveness and reliability of UEP communication technique will proliferate and will result in performance gain by taking into consideration the error sensitivity of protected packets in the network [40] [41]. UEP is used for protecting more vulnerable source bits for visual

coding. A video stream is comprised of various frame types that contribute differently to the perceptual video quality.

Any video stream not coded with UEP requires a large overhead for all blocks of coded packets, which would result in an ineffective data transmission scheme [42]. Therefore, UEP uses prior knowledge of the media to differentially protect data on frames with higher influence on the quality, as the binary bits in a compressed video bit-stream are not equally important. The most widespread UEP techniques contain spatial, temporal, and quality scalability in which more (less) important data bits are protected with more (less) redundancy. Without UEP, the video bit-stream is very sensitive to packet losses and even 1 percent packet loss rate is enough for significant perceptual quality degradation. The UEP technique can be implemented at the application layer using frame level RS codes without any modification of other network layers. Generally, the intra-frames are protected using RS codes with a high redundancy level while the inter-frames are protected with a lower redundancy level.

1.6.3. Energy efficiency considerations of FEC

In networked wireless multimedia systems, the growth of video traffic over wirelessly connected and embedded devices, e.g., smart phones and sensors have increased exponentially in the past few years [29]. The increase in network traffic has posed certain challenges. The three main problems that need attention are to: (1) minimize the transmission energy required, (2) provide an acceptable level of video quality, and (3) the content based protection of the important packets with varying channel conditions [43]. The current and future design of wireless networks demand low power consumption and implementation of energy efficient communication protocols for allocation of network resources that have some tolerable degradation to different packet losses that exist in the network. In wireless networks, the energy efficiency has the highest priority for

the design of the video transmission systems due to the limitation of computing and energy resource.

A quality-driven approach in which adding parities by the UEP technique to the tiny portions based on the importance of packets in the whole compressed bitstream that can result in limited energy resource of the devices installed in the network is desired. This is important because severe degradation is possible due to packet losses and the effect of which may be seen during decoding of semantic information and video content of video frames in the compressed bitstream [44]. Traditional UEP matches does not consider the power consumption but energy constraints are also the driving factors to optimize overall performance. The goal of energy efficient communication strategies is to have a holistic approach to improve the performance of video quality should utilize the extra content dependency with selective encryption to improve the video quality especially in channels with a relatively high bit error rate.

1.7. Organization of the thesis

In this thesis, we identified the features that affect the perceptual video quality through light-weight bitstream analysis. The relationships of the features were identified through decision tree model, which were used for content aware FEC implementation. We also estimated the judgments made about high level tasks by estimating the confidence measure of speaker and how confident the evaluators are by using both neural network and decision tree models.

This thesis is organized as follows. In Chapter 2, we present a light weight and reference free perceptual video quality model based on light weight bitstream analysis. We introduce the description of subjective tests first, followed by the identification of the features that affect perceptual quality based on video content and encoding distortions. In the last part of this Chapter, we introduce the neural network quality prediction model and compared the performance

evaluations in terms of correlation coefficients and number of instructions needed to reconstruct pixels, extract features, and compute motion vector our prediction model with standard full reference and reference free models.

In Chapter 3, we proposed the decision-tree-based quality prediction model to depict the explicit relationships among the features identified in Chapter 2. We designed an energy-efficient and content-aware FEC algorithm for QoE provisioning based on the relationship between FEC and QoE and updated the features that describe the impact of packet loss. In the last part of this Chapter, we compared the average network overhead, average SSIM, and average VQM of our algorithm with the two standard algorithms present in literature. We also compared the percentages of videos exceeding different thresholds classified based on MOS.

In Chapter 4, we identified that degradations in video quality also impact other aspects of viewer's experience, including the ability to make judgments about the content of video. The thin slices of non-verbal behavior provide information to make reliable high-level social inferences to different problem domains. We investigated how the variations in perceptual quality affect the judgments and how the impact of degradations in network transmission can affect the confidence measures of presenters within videos based on the proposed models described in Chapter 2 and Chapter 3. Finally, in Chapter 5, we draw the main conclusions and outline future research directions.

CHAPTER 2. PREDICTING PERCEPTUAL VIDEO QUALITY THROUGH LIGHT-WEIGHT BITSTREAM ANALYSIS

2.1. Motivation and related work

In recent years we have witnessed exponential growth of various video applications over wireless networked and embedded devices such as mobile phones and sensors. Maintaining good visual quality for these applications is a focal concern of service providers and network designers for satisfying the quality of experience (QoE) requirements of end users. Moreover, for many applications, it is essential to guarantee good visual quality since users make critical decisions based on their visual observations, e.g., identifying intruders based on videos from a wireless surveillance network.

The majority of multimedia networking protocols aim to satisfy quality of service (QoS) requirements, which are usually given in terms of bandwidth, delay, and packet loss ratio. However, these networking parameters alone do not necessarily reflect a user's experience of viewing the received videos. According to many subjective test results, the mean opinion scores (MOS) given by viewers on distorted videos cannot be merely determined by bit rate and packet loss ratio [45]. Given a video compression algorithm, the perceptual quality of compressed videos under the same bit rate can vary with different video content characteristics, such as the level of spatial details (brightness, edges, texture complexity, etc.) and temporal details (e.g., the extent of motion). The distortion caused by transmission is related to the locations of lost packets in a bitstream, and the visibility of packet loss also significantly depends on the content of the video [46].

To achieve more effective control of QoE for various video applications, there is a demand for mechanisms that can predict perceptual video quality accurately and in real time. More

specifically, the MOS of networked videos, which are collected from time-consuming subjective tests, should be predicted as functions of observable parameters from the video stream or the network. In particular, many inter-nodes in wireless networks, such as mobile phones and sensors, are embedded devices with limited processing power; therefore, quality prediction is expected to be conducted in a computationally efficient way. Networking protocols can leverage the prediction to control video quality and optimize network resources to meet the QoE requirements of users.

Perceptual video quality can be measured using reference-based or reference-free methods. Reference-based methods require access to the original source video (or quality features derived from the source video) to assess the quality of a compressed video [47], while reference-free methods assess video quality based on information from the compressed video without referring to the source video [48]. Since reference-based methods are complicated for implementation and cannot be used in cases where source videos are absent, reference-free methods are preferred for real-time monitoring and control of video quality in a network.

Several reference-free quality prediction models have been introduced in the literature. The ITU-T recommendation G.1070 [49] provides a parametric model that estimates video quality using bit rate, frame rate, and percentage of packet loss. One major drawback of the model is that video content is not taken into consideration. In [50] and [51], video quality estimation models were developed using regression techniques, and both models made use of detailed content information such as motion vectors (MV). The models in [52] and [26] estimated quality based on content features such as blockiness, blurriness, and extent of motion, which have to be extracted from a reconstructed (fully decoded) video. Extracting very detailed content information, such as motion vectors or even reconstructed pixels, makes the estimation process complex but not necessarily leads to better performance.

Recently, QoE estimation and control solutions have been proposed for different networking scenarios. A video quality estimator for UMTS networks was proposed in [26]. This model clustered video sequences into several groups based on content type, and video quality was estimated by a nonlinear function of content type, sender bit rate, block error rate, and mean burst length. Then a fuzzy logic based control algorithm was developed to adapt the rate of video streams to network dynamics to satisfy QoE requirements. This rate adaptation scheme focused on adjusting source coding rate but did not address FEC for protecting video streams against packet losses. In [33], a QoE-aware FEC mechanism was proposed for multimedia sensor networks, in which redundant FEC packets were created based on the impact of the frame on the user experience. An adaptive video-aware FEC mechanism with unequal error protection was proposed in [31]. This work also clustered video sequences into several groups with different content characteristics and made adaptive FEC choices based on the content of a video. Both algorithms in [31] and [33] have shown better performances than basic FEC or UEP algorithms; however, the decisions of FEC for different types of frames were not based on any quantitative models of perceptual quality. Our work is different from [31] [33] in that we design our FEC scheme based on a perceptual quality model, so that it is more effective to satisfy user experience. Our work also demonstrates how to bridge the research in perceptual video quality prediction and FEC with UEP.

In this chapter, in view of the need for accurate video quality prediction on wireless embedded devices, we propose a new reference-free and light-weight perceptual video quality estimation model. We introduce a set of features to depict video content characteristics, distortion caused by lossy compression, and distortion caused by network transmission. All the features can be obtained from light-weight parsing of the compressed bitstream: this enables the extraction of enough information to depict content related effects without introducing too much computation

for decoding the MBs and reconstructing the pixels. We have trained an artificial neural network (ANN) model for perceptual quality prediction, using a distorted video database consisting of videos with varying content characteristics, encoding parameters, and packet loss levels. Since the proposed model achieves efficient perceptual quality prediction through low computational costs, it serves as an appropriate tool for real time control of video quality in wireless networks. The rest of this chapter is organized as follows. The next section will describe our subjective test and the data set we used. In the next section, we introduce the details of the video quality estimation model and performance evaluation of results.

2.2. Description of the subjective test

In this section, we describe the generation of data sets used for developing the model and the steps of our subjective tests. To conduct subjective tests, we first generated a database of distorted videos taking into account a variety of factors that contribute to video quality. As perceptual quality is closely related to video content, 12 source videos with varying spatial and temporal details were selected from the Xiph collection of videos [54]. Description of these videos are presented in Table 1, and snapshots of them are shown in Figure 5. H.264/AVC, currently the most commonly used video coding standard, was applied to encode the videos. Each source video was encoded under different bit rates, frame rates, and GOP structures. When delivering video over a network, factors such as transmission errors, congestion, and excessive delays will result in packet losses that degrade video quality.

Figure 5. Snapshots of training and test video sequences

We simulated packet loss ratios (PLR) of 1, 5, and 10 percent in our test environment. A summary of the test conditions are given in Table 2. Combining different test conditions, 24 distorted videos were generated for each source video, which results in a total number of 288 videos in our data set. Our subjective tests followed the ITU-T recommendations, and we used the single-stimulus absolute category rating method with a quality scale of 1-9 [55]. A total number of 100 subjects participated in this experiment. The subjects were all students from NDSU with normal vision. To keep each subjective test within a manageable period, the test videos were divided into 4 groups, each containing 72 videos. Each group of videos was evaluated by 25 subjects.

Table 1. Test videos

	Low Motion	Medium Motion	High Motion
Low Spatial	Grandma Student	Mad Stock	City Deadline
High Spatial	Paris Dinner	Bigbuck bunny Sintel	Elephants dream Factory

We adopted the screening method recommended by BT.500-11 [56] to screen our collected data. We identified and neglected all those viewers whose ratings were not consistent with the ratings of other viewers, and after filtering these unreliable results, we obtained the MOS for each video sequence. Figure 6 shows a portion of our test results, where the MOS of six videos are categorized under combinations of bit rates and PLRs (with 30 fps frame rate and IBBP GOP structure). It can be found that there is substantial variation of the MOS for different videos. Therefore, content characteristics have to be taken into account to achieve accurate estimation of perceptual quality.

2.3. Neural network based prediction model

The first step to neural network based prediction model is to extract all the necessary features from the bitstream.

Table 2. Test conditions

Encoder	JM18.5 High Profile
Resolution	CIF(352x288)
GOP format	IPPP, IBBP
Frame rate	15,30
PLR (%)	1,5,10
Duration	20sec
GOP Size	16
Rate control	Enabled
Bitrate(Mbps)	1.5,4

2.3.1. Feature extraction

To build a reference-free quality estimation model, our primary goal is to identify and extract useful features from a compressed bitstream. Figure 7 illustrates the levels of access to an H.264/AVC bitstream. The bitstream consists of a sequence of Network Abstraction Layer Units (NALU). The payload in a NALU can be accessed at frame layer, slice layer, and MB layer, with increased details in the video. After accessing all the syntax elements at the MB layer (MV, prediction type, etc.), the decoder can reconstruct the pixels in the MB. Quality estimation models in [48] [50] [51] [52] use either motion vector information or reconstructed pixels to characterize video content. That is, they require accessing the bitstream at the MB level. Our work aims to efficiently predict the perceptual quality of video with low computational cost. Unlike these works, we parse the information at the slice level without fully decoding each MB. We have found that the information contained in the slice level can be leveraged to depict video content characteristics, encoder distortions, and packet loss patterns, which are essential for quality estimation. Therefore, we do not need to waste a lot of computational resources on decoding MBs and reconstructing pixels. In the following, we present a set of features that can be obtained by parsing the bitstream at the slice level.

2.3.1.1. Average quantization parameter (QP_{avg})

Video encoders adjust the bitrate by changing the quantization parameters (QP). The value of QP directly reflects the degree of blockiness and blurriness in a video. In the H.264 bitstream, the QP of a slice is included in the slice header at the slice layer in figure 7. We extract the QP of each slice from the slice headers, and compute the average QP of an entire video sequence.

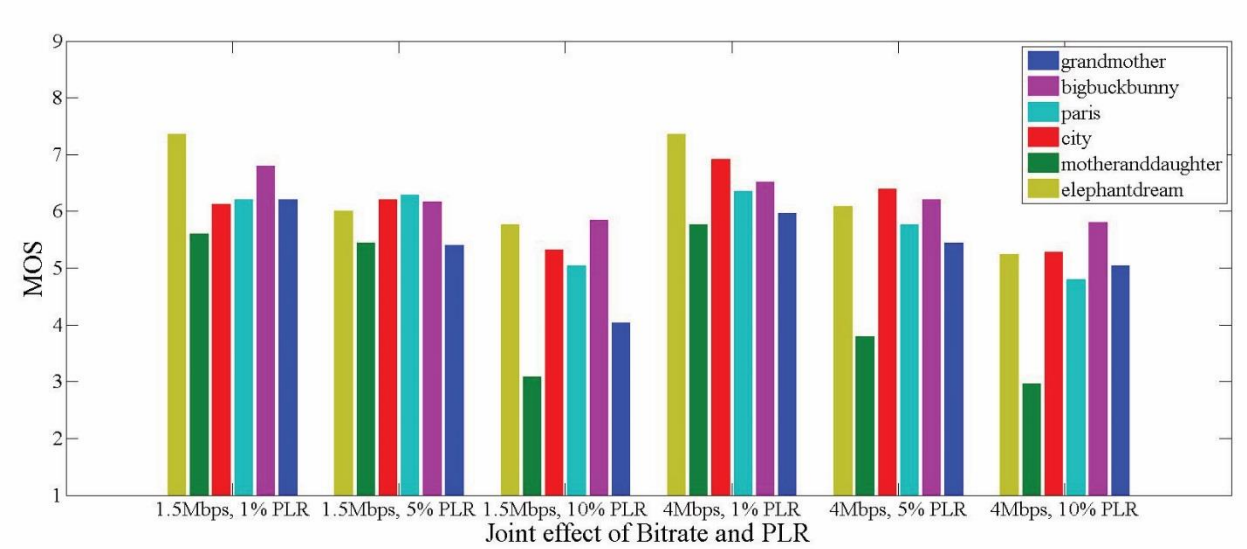

Figure 6. Mean opinion scores for different bit rates and PLRs

2.3.1.2. Average number of bits per intra frame ($BIntra_{avg}$) per pixel

The spatial complexity of a video can be reflected by the number of bits used to encode its intra frames. If given the same QP for intra frames, a complex scene needs a lot of bits to represent, while a simple scene requires significantly fewer number of bits. The overall spatial complexity of a video can be estimated by measuring $BIntra_{avg}$ and QP_{avg} together.

2.3.1.3. Average number of bits per inter frame ($BInter_{avg}$) per pixel

This feature reflects the temporal complexity within a video. Low (high) motion videos require fewer (more) number of bits per inter frame. For precise measurements of temporal complexity, the decoding of exact MVs at the MB layer is required, but counting the number of bits per inter frame at the slice layer addresses the same purpose.

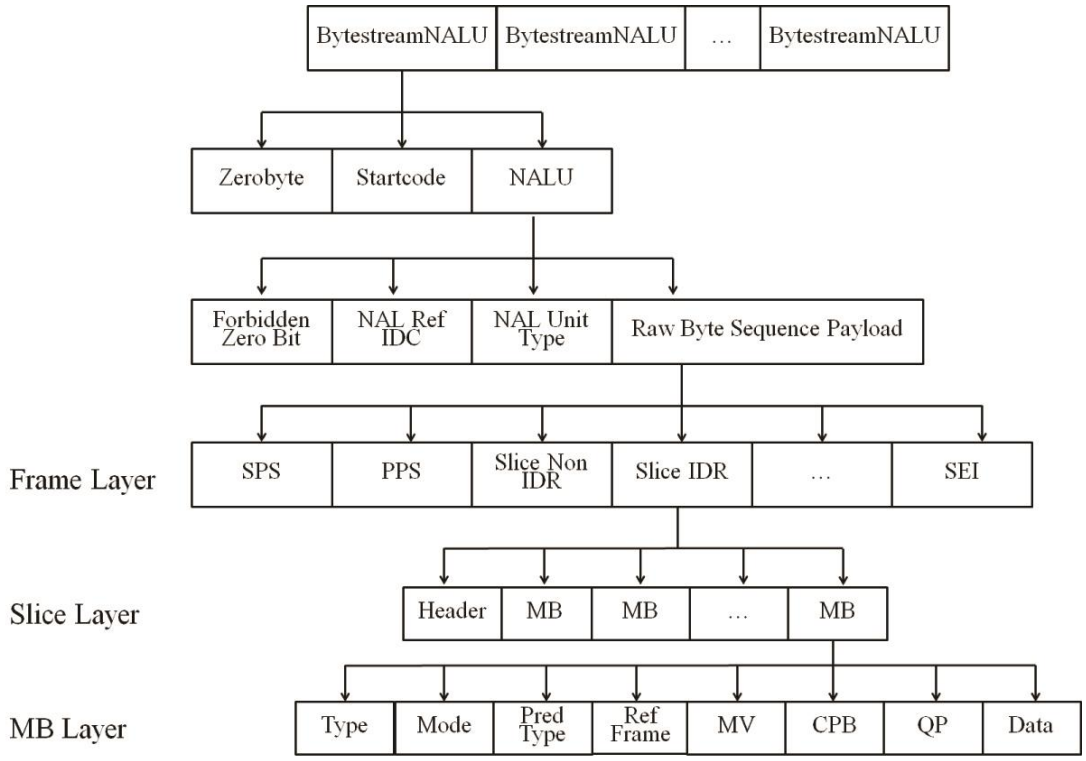

Figure 7. Level of access of H.264 bitstream

2.3.1.4. Average ratio of number of bits per inter frame to number of bits per intra frame in the same GOP ($R_{inter/intra}$)

$R_{inter/intra}$ is also used to depict temporal (motion) complexity. While $BInter_{avg}$ depends on the quantization level, this ratio is introduced as a metric independent of quantization.

2.3.1.5. Percentages of intra, inter, and skip MBs in inter frames ($\%_{intra}$, $\%_{inter}$, $\%_{skip}$)

The percentages of intra, inter and skip MBs in inter frames are related to the texture and motion characteristics in the video. For homogeneous regions, such as large objects, large partition sizes can result in a lot of intra MBs. For detailed regions, small partitions sizes can lead to more inter MBs. The skip mode is usually used for background or big size objects in local scene statistics.

2.3.1.6. Frame rate (FR)

The frame rate of 15Hz is the threshold for human satisfaction level [14]. In general, the perceptual quality increases with the frame rate. However, the effect of frame rate is also content dependent: adjusting frame rate can cause different effects on low motion videos and high motion videos [30]. In our model, this feature will work together with the other features that depict spatial and temporal characteristics of a video.

Due to the complex organization of video syntax elements in a compressed bitstream, using packet loss ratio (PLR) alone is not sufficient to describe the effects of packet loss. For example, the duration of visible artifact depends on how many frames an error propagates, which is different for intra and inter frames. An intra-frame packet loss can result in visual quality degradation in the duration of a GOP. As another example, the error visibility caused by packet loss depends mostly on the temporal or motion complexity of the video. If large motion exists in the packet or the temporally adjacent packets, it can severely impact an inter frame. Accessing a compressed bitstream at the slice level actually allows the extraction of more detailed packet loss information than a simple PLR. We have identified the following features to depict packet loss effects.

2.3.1.7. Maximum and average burst length (BL_{max}, BL_{avg})

The effect of consecutive slice losses in a frame is different from that of scattered slice losses in separate frames. We introduce BL_{max}, the maximum number of consecutive slices lost in the bitstream, because a long burst length may cause significant degradation in user experience. The average length of consecutive slice losses is introduced to count for the average effect of consecutive slice losses in the entire video sequence.

2.3.1.8. Percentage of slices that can be successfully decoded (DSlice%)

Packets lost at different positions can result in different lengths of error propagation. This feature shows how many slices are affected by packet losses after error propagation

2.3.1.9. Percentage of frames that can be correctly decoded (DFrame%)

Similar as DSlice%, this feature depicts the impact of packet losses at the frame level after error propagation. The aforementioned features are chosen to reveal all the factors contributing to perceptual quality: video content characteristics, distortion caused by lossy encoding, as well as degradation caused by transmission. All the features are either readily available in the bitstream or can be computed easily from the information at the slice level.

2.3.2. Prediction model

It has been well-known that perceptual quality is related to video content characteristics, encoding distortions, and packet losses in a non-linear fashion. The challenging problem is how to reveal this kind of nonlinear relationship to predict perceptual quality. We propose to build an ANN based prediction model [43], as ANN is capable of implicitly detecting complex nonlinear relationships and detecting all possible interactions between predictor variables.

Figure 8 depicts our ANN model. ANN is a sorted triple (N,V,w) with two sets N,V and a function w, where N is the set of neurons and V is a set of connection $\{(i,j)\mid i,j\in \mathrm{N}\}$ whose elements are connections between neuron i and neuron j. The function $w(i,j)$ is the weight of connection between neuron i and j. During the learning process the weights are updated for many iterations until a convergence criterion is satisfied, e.g., the sum-squared error is less than an error goal.

As in most applications of ANNs, we use a three-layer network structure where the set of neurons are partitioned into three subsets: input nodes, hidden nodes, and output nodes. Each input

neuron takes the input of one of the aforementioned features extracted from the bitstream. The hidden layer implements the nonlinear transformation of input variables to a function for exact interpolation on a set of data points in multidimensional space. The number of neurons in the hidden layer needs to be determined properly based on the inputs, outputs, training set, and the precision model. In our model, we set the number of neurons in the hidden layer as 11, after we experimented with different candidate numbers. The output layer consists of one neuron that combines the hidden layer output and it presents the predicted MOS as the final output. The radial basis function, which is commonly used to model nonlinear systems, is used as the kernel function for all the neurons.

2.4. Performance evaluation

In this section, we evaluate the proposed model in terms of prediction accuracy and computation complexity. From our video sequences listed in Table 1, we used the videos titled Student, Stock, Deadline, Dinner, Sintel, and Factory to train the ANN model described in the last section, and then we evaluated the prediction model by testing on the rest six videos. Videos in both the training and the testing sets cover a wide variety of content characteristics.

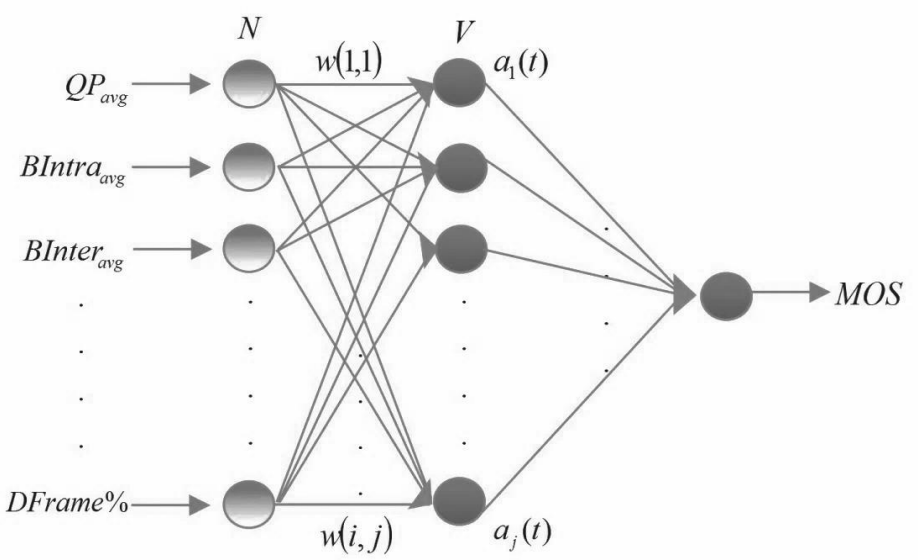

Figure 8. ANN model for video quality assessment

The proposed model was compared with VQM, SSIM, and the model proposed by Aguiar et al. [53]. Both VQM and SSIM are popular full reference video quality evaluation models. VQM is a standardized method that estimates video quality by measuring perceptual impairments including blurring, jerky motion, global noise, block distortion, and color distortion [47]. SSIM measures the similarity between two videos and considers video degradation as perceived change in structural information [44]. Both VQM and SSIM require complete decoding of pixels at the MB layer. The model proposed by Aguiar et al. [53] is a reference-free method. It divides videos into three content types (low, medium, and high motion), and builds separate neural network models for each type. The features used in this model were GOP size, overall frame loss rate, and loss rates of I, P, and B frames. For a fair comparison, we used the same features to train separate ANN models on the low, medium, and high motion videos in Table 1, where the parameters of the ANN were the same as the proposed model, so that we can compare the effect of different sets of features.

Figure 9 shows the prediction of MOS of these models on a set of the test videos. The performance of quality prediction was evaluated by the Pearson correlation coefficient, which depicts the linear dependence between the predicted quality and the corresponding subjective test results. The models were also compared in terms of the Spearman correlation that assesses how well the relationship between two variables can be described using a monotonic function. Table 3 shows values of these correlation coefficients along with the root mean square errors (RMSE) of prediction. The proposed model achieves a correlation accuracy of 87.09% (Pearson) for predicting MOS, which is better than the correlation values of 59.36% for VQM and 52.84% for SSIM. Moreover, Table 4 shows a comparison of the proposed model and the model by Aguiar et al. [53] for different video content types. It can be seen that our proposed model achieves better prediction performance than Aguiar et al. [53] for all the content categories.

Table 3. Prediction performance

Metric	Pearson	Spearman	RMSE
Proposed	0.8709	0.8689	0.1153
VQM	0.5936	0.5839	0.1869
SSIM	0.5284	0.5197	0.1978

After our ANN model is trained, estimating MOS is simple once the proposed features are acquired. Therefore, we focused on investigating the computations needed to parse the bitstream to acquire the proposed features. We computed the average total number of instructions needed to acquire the proposed features per frame, and compared it with the average number of instructions required to reconstruct pixels and compute MVs per frame. The results are as shown in Table 5.

Table 4. Prediction performance for different types of videos

Category	Pearson		RMSE	
	Proposed	Aguiar	Proposed	Aguiar
Low Motion	0.9154	0.4752	0.0658	0.0987
Mid Motion	0.7283	0.4811	0.07921	0.1162
High Motion	0.5589	0.4693	0.1078	0.1754

For the test videos under investigation, the time required to compute MV at the MB layer took an average of 80% of the total time for full decoding (reconstructing pixels), whereas acquiring the proposed features only requires 31% of the total time for full decoding. This demonstrates the computational efficiency of the proposed model.

Table 5. Number of instructions needed for accessing the bitstream

Test Video Sequences	Average number of instruction needed to		
	Reconstructed Pixels	Extract Features	Compute MV
Grandma	807	259	632
Mad	773	206	626
City	819	318	631
Paris	835	207	688
Bigbuck Bunny	776	259	620
Elephants Dream	865	275	718

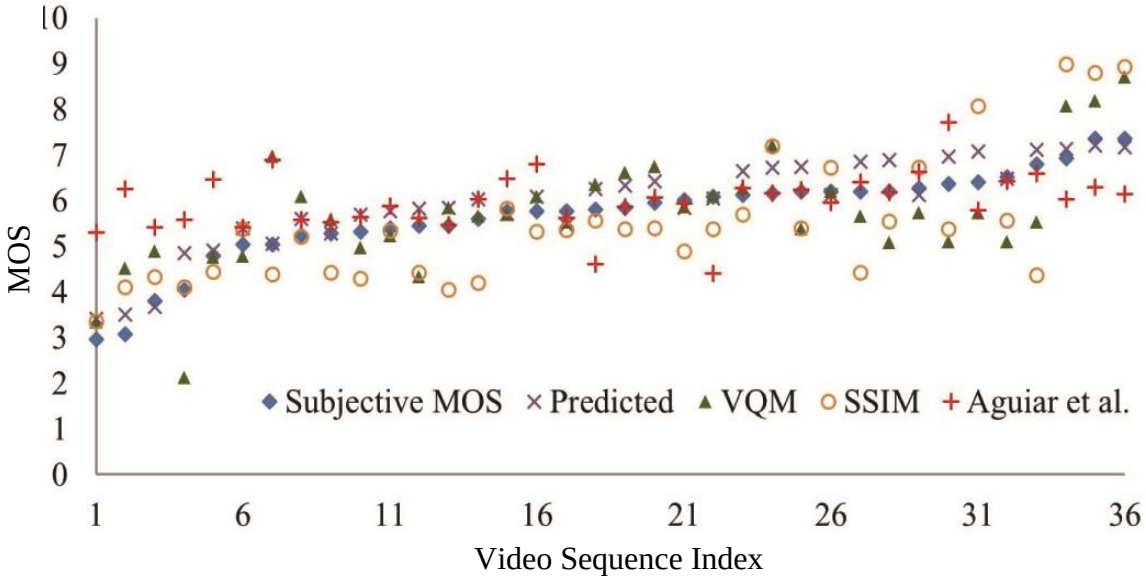

Figure 9. MOS prediction results on different test videos

2.5. Conclusion

This chapter proposes an efficient and light-weight video quality prediction model through partial parsing of compressed video bitstreams. A set of features were introduced to reflect video content characteristics and distortions caused by compression and transmission. All the features can be obtained directly from the H.264/AVC compressed bitstream in parsing mode without decoding the pixel information in macro-blocks. Based on these features, an ANN model was trained for perceptual quality prediction. Evaluation results show that the proposed prediction model can achieve accurate prediction of perceptual video quality through low computation costs. Therefore, it is well-suited for real time networked video applications on embedded devices.

CHAPTER 3. ENERGY EFFICIENT AND CONTENT - AWARE FEC FOR WIRELESS MULTIMEDIA COMMUNICATIONS

3.1. Introduction

It is generally known that perceptual quality is related to video content characteristics, distortion caused by lossy compression, as well as impairment introduced during transmission [26]. Yet the challenge lies in how to efficiently model these factors using parameters which can be easily obtained from video bitstream and the underlying network, so as to facilitate the design of QoE control mechanisms. A lot of parametric models have been proposed which take into account video content characteristics (brightness, texture, motion, etc.) and the effect of source coding [57]. These models work well for predicting video quality without degradations introduced by transmission loss. A more difficult task is to model the effect of network impairments. For this purpose, learning-based data analysis has been introduced and shown effective. There are learning algorithms for classifying packet loss effects [46], [48]. The ANN based analysis approach was applied for quality prediction in several studies and have shown good prediction results [53] [58] [59].

Even though, ANN has the capability of capturing non-linear and complex underlying characteristics and can approximate any function. But, the approximation specified by ANN cannot interpret and extrapolate any relationships between input parameters and output variable. An ANN is a non-parametric, data-driven, self-adaptive, and computationally flexible tool but is unable to extracting the knowledge (weights in ANN). An ANN cannot deal with uncertainties and do not specify or provide any insights on the type or structure of approximate function used internally. However, because of the black box nature of neural network models, there is no explicit

relationship between the input features and the resulting quality, and therefore, it is hard to apply a neural network model to control video quality by adjusting the input features.

On the other hand, the application level error control mechanisms provide reliability and error correction coding in wireless channels. For source and channel coding, certain measures are taken to ensure that the coding stream has error restoration capability. In this chapter, an algorithm is designed to minimize the transmission error of bit-streams under a total bit rate constraint while satisfying perceptual quality constraints. We save energy while maintaining acceptable video quality through a controlled reduction in the number of parity symbols in the RS code.

We propose a way of selecting an energy-efficient level of packet redundancy by the intelligent use of RS code based on the traversal of a decision tree that takes into consideration the source and channel parameters. Our algorithm is optimized to reduce the overall energy consumption of mobile devices running a video application, while incurring only an increase in the performance of error recovery. The error control in mobile and embedded devices is QoE optimized against energy consumption under bitrate constraint to meet the user's preferences. This motivates us to develop an energy management scheme, which can guarantee a certain acceptable perceptual quality performance by reducing processing time while extending the battery life. In this chapter, in view of the need for perceptual video quality prediction and control for wireless embedded devices, we propose a reference-free quality model that can predict perceptual video quality in real-time and with low complexity. Based on the model, we introduce an application layer content-aware FEC scheme to ensure the quality of wireless transmitted video in an energy-efficient way. Our contributions are summarized as follows:

- Our perceptual video quality prediction model takes into consideration video content characteristics, distortion caused by lossy compression, and distortion caused by network

transmission. We extract a set of simple features from the compressed bitstream and the network, and build a decision tree model to predict quality based on these features. The model allows prediction of perceptual video quality in real-time and with low complexity. More importantly, the decision tree approach reveals explicit relationship between perceptual quality and the set of features, making the control of perceptual quality a much easier task.

- We establish quantitative relationships between FEC- parameters and perceptual video quality. Based on such relationships, we introduce an algorithm for adjusting FEC parameters by jointly considering perceptual quality requirement, video content characteristics, source coding parameters, and network conditions. To the best of our knowledge, this is the first quantitative study that reveals the nonlinear relationship between FEC parameters and perceptual video quality.
- We give a solution on how to achieve energy-efficient QoE control for video applications. Our FEC algorithm is designed to minimize the transmission energy consumption of embedded devices running a video application, while maintaining a perceptual quality requirement under a bitrate constraint in the network. Through the operation of the FEC algorithm, we show that there is potential for saving energy by leveraging the characteristics of perceptual video quality.

3.2. Background and related work

The background and related work will be summarized in the following categories:

3.2.1. Perceptual video quality model

A plethora of research is conducted to estimate video quality based on different features by using reference-free quality prediction models. More details about different perceptual quality

metrics that fall under reference-based and reference-free methods can be seen in Section 2.1 of Chapter 2. Earlier, we implemented the ANN based quality prediction model that is trained on massive subjective test results. Even though, the ANN prediction model reflects the QoE by extracting the features, that depict video content characteristics and encoding distortions, in parsing mode from the bitstream. However, due to black box nature of ANN, the structure of approximate function used internally and the hidden relationships among the features are not fully extrapolated, which makes ANN models not suitable for QoE control.

3.2.2. QoE support for video communications

Existing QoS metrics are typically used to indicate the impact on the video quality level from the network's point of view, but do not reflect the user's perception. The traditional encoding related measures of quality like Peak Signal-to-Noise Ratio (PSNR)[60], Structural Similarity (SSIM)[61], and Video Quality Metric (VQM)[62] require the presence of the original videos to compare and assess the user perceived QoE. Normally, when working on real-time QoE assessment platforms, the original transmitted data is not available. There is the lack of a standardized approach to measure QoE, and there is a fundamental requirement for mechanisms capable of assessing user satisfaction and the effective usage of network resources. Any QoE assessment mechanism has to decide the most suitable features for the quality assessment. The QoE prediction model requires the definition of a suite of quality metrics that capture several effects introduced by the delivery mechanisms, different types of data, delay variations of packets, and losses related to network performance.

Earlier, QoE prediction models have been proposed for different types of wireless networks. A video quality estimator for UMTS networks was proposed in [26]. This model clustered video sequences into several groups based on content type, and video quality was

estimated by a nonlinear function of content type, sender bit rate, block error rate, and mean burst length. Khan et. al. [26] performed 4-way repeated ANOVA to understand the interactions of the four variables in the regression modeling. The layered encoding was used for adapting the video streams to the network dynamics and the fuzzy logic algorithm processed the feedback information. The video quality estimator for wireless mesh networks in [53] also adopted a clustering technique to separate videos into several categories based on the degree of movements in a video, and a neural network model was built to estimate video quality for each category of videos. These two models used mostly network-related parameters to estimate quality, but did not explore much of video content characteristics.

The degradation of the perceptual quality due to quantization and frame-rate reduction as model parameters can be predicted accurately from some content features were identified in [63]. The authors proposed an accurate perceptual quality estimation based on temporal correction and spatial quality factors as functions of frame rate and PSNR respectively.

3.2.3. Error control for video communications

The wireless network channels are non-stationary and are affected by the rapidly changing channel conditions. Since the channel bit error rate varies over time, perceptual video quality in applications that require video transmissions cannot afford higher error rates. In literature, two basic error control strategies titled error detection and error correction mechanisms that bring the improvement of end-to-end video quality and combat bit errors against channel conditions [64] [65][66]. In terms of perceptual quality control in wireless environments, we focus our study on application layer error control, which is an essential mechanism for guaranteeing video quality by protecting video packets against packet losses. The bit errors are usually corrected at the lower network layers and only packet losses (or erasures) can happen at the application layer. In wireless

networks, the schemes, such as FEC with Reed-Solomon (RS) codes [57], interleaving [65], and Unequal Error Protection (UEP) [66] for different frame types are used to increase the resilience of the bit-stream to transmission errors.

3.2.4. Energy efficiency for wireless video communications

The wireless connections operate in environments that may change drastically and result in poor quality, lower bandwidth, less connection stability, and higher error rates. In networked wireless multimedia systems, the two main problems are to minimize the required transmission energy and to provide an acceptable level of video quality over time with varying channel conditions [57]. The perceptual quality can be improved by using an RS code with more parity, at the cost of increased energy consumption. Both energy consumption and video quality are affected by the use of RS codes. The H.264/AVC video transmission system will result in a high computational complexity with distortion estimation caused by packet losses, erasure-correction coding, and UEP optimization [66]. To determine the energy-efficient RS code and QoE requirement is a challenge; therefore, our goal is to find an optimal solution for performance and the energy consumption used by the RS coding-based error recovery.

The energy is a critical resource in battery-operated wireless networked and embedded devices such as mobile phones and sensors. An end user prefers to watch the video in real time at a lower perceptual quality, rather than experience the sudden termination of high-quality video because of a flat battery. Various attempts have been made to reduce the energy consumption while maintaining an adequate performance [65]. These studies have provided a good understanding of the trade-off between performance and the energy used by the RS coding based error recovery. However, there has not been much work on the problem of reducing the energy consumed while maintaining end user acceptable video quality requirement with a controlled reduction in the

number of parity symbols, which results in energy saved by transmitting less data. An optimal configuration that adapts to the packet loss for wireless video networks was needed. Therefore, effective implementation of FEC that reduces the amount of data by a considerable factor by maintaining the threshold on perceptual quality is desired. Although current UEP mechanisms assigned different priorities for video packets based on the structure of video bitstream, there is not quantitative relationship between UEP and perceptual video quality, and thus they cannot provide guarantee to a certain level of perceptual quality.

3.3. Decision-tree-based quality prediction

Decision trees are simple, run fast with lots of observations, and do not require a lot of effort from users for data preparation. The features used in a decision tree are emphasized during the process of construction phase when all nonlinear relationships between parameters are squeezed without affecting tree performance. Due to non-parametric nature of decision tree, the outcome results are fully explored without missing any detailed information. Decision trees implicitly perform variable screening and perform well with large data by analyzing comprehensively the consequences of each possible decision. The focus of this work is to find efficient ways to tune video coding and transmission parameters to control the QoE.

In our previous work [67] and other related works such as [53], neural network models were trained to estimate perceptual video quality. Although good prediction results were obtained, the black box nature of neural network prevents us from analyzing the direct relationship between input features and the predicted output. For this reason, we propose to build a decision tree model which provides explicit relationships between the input features and the predictions. Decision tree is also advantageous in that it is simple, runs fast with lots of observations, and does not require a lot of effort from users for data preparation.

We use category judgement scales to build our model. Based on the distribution of MOS in our subjective test, six different categories of subjective quality were introduced: poor, bad, fair, good, very good, and excellent. We have obtained the mean (μ) and standard deviation (σ) values of all valid values of MOS from our subjective test. Any value less than (μ-2σ) was treated as poor. The values in the range of (μ-2σ) to (μ-σ) were bad, (μ-σ) to μ were fair, μ to (μ+σ) were good, (μ+σ) to (μ+2σ) were very good and any values above (μ+2σ) were in excellent category. The MOS ranges and corresponding categories are presented in Table 6.

A decision tree model was trained using the aforementioned features, and based on videos entitled student, stock, deadline, dinner, sintel, and factory. The complete model is shown in Figure 10. The most important factor is DSlice% that represents the percentage of successfully received frames is on the top of the tree. Other factors such as $B_{Intraavg}$, $B_{Interavg}$, and $R_{inter/intra}$ are present in level 1 and level 2 of decision tree. There are many test conditions that distinguish different quality criteria from poor to excellent. If the user wants to perceive excellent quality, we should have DSlice% ≥ 0.9177 and $B_{Intraavg} \geq 0.5249$ and $B_{Interavg} \leq 2.3241$.

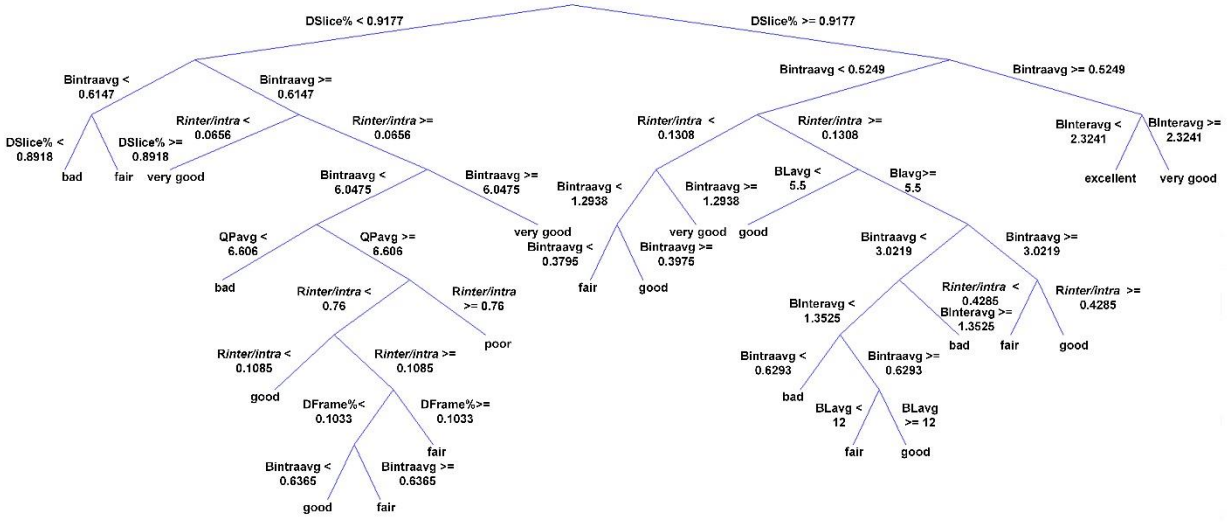

Figure 10. Decision tree generated from the training data

If $B_{Interavg} \geq 2.3241$, then the resulting quality will be very good instead. Other factors, such as QP_{avg}, BL_{avg}, and DFrame% are also important and can be seen in lower half of the tree. The decision tree prediction model was tested on the 6 types of videos in Table 1 which were not used during the training, and the testing results were summarized in Table 7. For all the 144 videos for testing, the prediction model resulted in 16 mismatches of classes i.e., 128 out 144 vides were correctly identified, and that corresponds to 88.9% classification accuracy.

Table 6. Classification of different scales based on MOS

Mean Opinion Score	Categorization
<3.3770	Poor
3.3770-4.3695	Bad
4.3695-5.3621	Fair
5.3621-6.3547	Good
6.3547-7.3473	Very Good
>7.3473	Excellent

Table 7. Classification accuracy of results

Categorization	Poor	Bad	Fair	Good	Very Good	Excellent	Total
Poor	18	1	1	0	0	0	20
Bad	1	16	1	0	0	0	18
Fair	0	2	23	1	0	0	26
Good	0	0	2	40	2	0	44
Very Good	0	0	0	2	24	1	27
Excellent	0	0	0	1	1	7	9

3.4. Energy-efficient and content-aware FEC

For common wireless imaging applications, video clips are captured by embedded devices, compressed to reduce redundancy, and then divided into packets for transmission. Error control has to be applied to reduce the degradation of video quality caused by packet losses in the network [57]. We focus on FEC at the application layer, in which additional channel coding protection bits will be added to the compressed video. Our goal here is to find the most energy efficient solution for FEC under a certain perceptual quality requirement for an application.

The decision tree model have revealed the relationship between perceptual quality and features related video content, encoder parameters, and packet loss. After the compression process, the features related to video content and encoding can be determined, and by tuning FEC options (i.e., how much redundancy to add to a certain type of video packets), the features for packet loss could be adjusted to meet perceptual quality requirements. In the rest of this section, we first analyze the relationships between FEC options and the features for packet loss in our quality model, and then introduce a content-aware FEC algorithm with the objective to minimize the energy needed for transmitting redundant packets.

3.5. Relationship between FEC and perceptual quality

We apply frame-level FEC to encoded video frames, i.e., upon completion of the compression of each frame, we add FEC packets to protect the frame. We make the following assumptions about network transmission:

- Each video packet contains one slice, which is an independently encoded unit.
- During transmission, the transmitted packets can be dropped, delayed, or corrupted. The channel is modeled as a packet erasure channel, which assumes that a transmitted packet is either correctly received or lost.

- From the underlying network, we can obtain two parameters: average packet loss rate (p_B) and average packet burst length (the average number of consecutive packet losses, denoted by l_B).

Many FEC codes, such as the Reed-Solomon (RS) code [68] [69] and RCPC [70] [71] [72] can provide a series of channel coding rates, denoted by $\{R_C^1, \ldots, R_C^N\}$. A proper channel coding rate should be selected depending on the network condition and quality requirement. We apply the Reed-Solomon (RS) coding technique which has been widely used for protecting video packets in packet erasure channels. For a RS (N, K) code, the coding rate RC is given by K/N. For a video frame consisting of K packets after source coding, we add N - K redundant packets to it and send the N packets. The RS code can correct up to errors.

$$t = \left\lfloor \frac{N-K}{2} \right\rfloor \quad (1)$$

The loss of packets is modelled as a series of independent Bernoulli trials. (Although for common memory channels there exists correlation of loss probabilities among adjacent packets, channel memory could be eliminated by the interleaving technique.) The probability that a frame can be decoded without any errors is the probability that at least N-t out of N trials are successful, which is given by

$$q(N, K, p_B) = \sum_{i=N-t}^{N} C_N^i (1-p_B)^i {p_B}^{N-i} \quad (2)$$

There are three features describing the impact of packet loss to quality in our decision tree model: percentage of slices that can be successfully received (DSlice%), percentage of frames that can be correctly decoded (DFrame%), and average burst length (BL_{avg}). We will estimate these features given the parameters for network conditions and the FEC parameters.

1) DSlice%: This feature is the percentage of slices that can be successfully received at the decoder application. If no FEC packets are applied, because each slice is sent out as a packet, DSlice% is equivalent to the probability that a packet is successfully received, which is given by 1-p_B. When FEC packets are applied at the frame level, there are two cases that a slice can be successfully received: 1) the slice is not lost during transmission, which occurs with probability 1-p_B; 2) the slice itself is lost, but a sufficient number of slices in the frame are received such that the lost slice can be recovered. Combining these two cases, DSlice% is estimated by

$$\widehat{DSlice}\% = (1 - p_B) + p_B . q(N-1, K, p_B) \quad (3)$$

In this equation, $q(N-1, K, p_B) = \sum_{i=N-t}^{N-1} C_{N-1}^{i} (1-p_B)^i {p_B}^{N-1-i}$ is the probability that at least K packets are received out of the rest N - 1 packets for this frame, and $p_B . q(N-1, K, p_B)$ is the probability that the slice can be recovered from other slices.

2) DFrame%: There are three types of frames in the H.264/AVC bitstream: intra-coded (I), predictive (P), and bidirectional (B) frames. For each type of frame, based on its size, a certain amount of FEC packets are added. Given the packet loss rate p_B, the probability for successfully transmitting a frame can be derived from the above Bernoulli trial equation. The probabilities of successful transmission of each type of frame are given by

$$q_I = q(S_I + S_{IF}, S_I, p_B) \quad (4)$$

$$q_P = q(S_P + S_{PF}, S_P, p_B) \quad (5)$$

$$q_B = q(S_B + S_{BF}, S_B, p_B) \quad (6)$$

where S_I, S_P, and S_B are frame sizes (in packets) for I, P, and B frames, and S_{IF}, S_{PF}, and S_{BF} are the FEC packets added for these frames.

A GOP is typically structured as IB...BPB...BPB...B, where one I frame is in the beginning and it is followed by P and B frames. Denote the number of P frames in the GOP by N_P, the number

of B frames in the GOP by N_B, and the number of B frames in between an I and a P or two P frames by N_{BP}. The total number of frames in a GOP is N_G and it is equal to $1 + N_P + N_B$. According to the analytical study on playable frame rate in lossy networks in [57], the expected percentage of decodable frames at the receiver can be estimated as

$$\widehat{DFrame}\% = \frac{q_I}{N_G}\left[1 + \frac{q_P - q_P^{N_P+1}}{1 - q_P} + N_{BP}.q_B.(\frac{q_P - q_P^{N_P+1}}{1 - q_P} + q_I.q_P^{N_P})\right] \quad (7)$$

3) Average burst length (BL_{avg}): In our perceptual quality model, burst length is defined as the number of consecutive slice losses as seen by the video decoder at the receiver. If the video packets are transmitted in the network without any FEC protection, the burst length in our model BL_{avg} can be directly estimated by the average packet burst length (l_B) from the network. It becomes more complicated to estimate BL_{avg} when different levels of FEC are applied to all I, P, and B frames. Applying FEC helps to recover from small burst errors, but it is possible that it cannot recover from a large burst error. Statistically, we have $\widehat{BL_{avg}\%} \leq l_B$. For simplicity, we use the upper bound l_B to estimate the average burst length at the video decoder.

3.6. Energy-efficient and content-aware FEC for QoE provisioning

Based on the distribution of MOS values collected from our test videos, user-perceived quality of a networked video could be classified into six categories: poor (P), bad (B), fair (F), good (G), very good (V), and excellent (E). A user could specify a quality requirement as "the quality of received video should at least be good". We denote this threshold quality by Q_T (e.g., Q_T = G). After the source coding process, there are many options to many options to apply different FEC coding rates on different types of frames. From a series of candidate FEC parameters, we seek a set of coding options that can minimize the resulting rate R_t with the objective to save energy

for wireless transmission, and meanwhile can result in a perceptual quality (Q) equal or higher than Q_T.

We propose to find FEC solutions on the GoP level. In a typical video coding scenario, a video clip is organized into several GoPs, which is usually structured as IB...BPB...BPB...B, where one I frame is in the beginning and it is followed by P and B frames. Each frame consists of several slices, and one slice is treated as one packet before transmission. Previously, we derived that the total number of frames in a GoP, N_G, is equal to$1+N_P+N_B$. We denote the average number of packets/slices generated by I, P, and B frames by S_I, S_P, and S_B, and denote the average size (number of bits) for I, P, and B packets by l_I, l_P, and l_B. The frame rate is given by FR. Suppose S_{IF}, S_{PF}, and S_{BF} are the number of redundant FEC packets added to I, P, and B frames, respectively. The overall transmission rate can be computed by adding up all the of bits generated for frames with a duration of one second, which is given by

$$R_t = \frac{FR}{(1+N_P+N_B)}\{(S_I+S_{IF}).l_I + N_P.(S_P+S_{PF}).l_P + N_B.(S_B+S_{BF}).l_B\} \quad (8)$$

We formulate and FEC rate assignment problem as follows:

Given: $\{S_{IF}, S_{PF}, S_{BF}\}, p_B$

Find: $(S_{IF}^*, S_{PF}^*, S_{BF}^*)$ (9)

Minimize: R_t (10)

Subject to: $Q \geq Q_T$ (11)

$R_t \leq MSR$ (12)

The objective is to minimize the overall transmission rate in (8) while satisfying the quality requirement and the maximum sending rate (*MSR*) constraint given by the network. By solving this problem, the best FEC parameters $(S_{IF}^*, S_{PF}^*, S_{BF}^*)$ are determined for every GoP to minimize the overhead in the network. The problem could be solved as follows. First, for every GoP that has been encoded, we compute the features related to video content and encoding parameters. Second, we find out all the possible paths in the decision tree that can yield a satisfactory quality. Then we can further specify the ranges of network related features *DSlice%*, *DFrame%*, and BL_{avg}. Lastly, from these possible ranges, we will find out the best combination of FEC parameters that minimizes the overall transmission rate. Detailed steps for solving the FEC problem are shown in Algorithm 1.

Algorithm1. Energy-efficient and content-aware FEC algorithm

1: Given the perceptual quality threshold Q_T and target bitrate MSR
2: Measure parameters related to video content and source coding: QP_{avg}, $B_{intraavg}$, $B_{interavg}$, and $r_{inter/intra}$
3: Compute the video frame sizes (S_I, S_P, and S_B)
4: From the decision tree, find all candidate paths {*P(i)*} such that QP_{avg}, $B_{intraavg}$, $B_{interavg}$, and $r_{inter/intra}$ align with the ranges in the paths
5: **for each** candidate path P(i)
6: **if** $Q(P(i)) \geq Q_T$ **then**
7: Obtain valid ranges of *DSlice%, DFrame%, and* BL_{avg}
8: Based on FEC structure, compute different combinations of S_{IF}, S_{PF}, and S_{BF}
9: **for each** combination of S_{IF}, S_{PF}, and S_{BF}
10: Estimate *DSlice%(eq. (3)), DFrame% (eq. (7)), and* BL_{avg}
11: Set this combination as valid if can result in
12: i) $R \leq MSR$
13: ii) *DSlice%(eq. (3)), DFrame% (eq. (7)), and* BL_{avg} *are within the valid ranges indicated by this path*
14: **end for**
15: **end if**
16: **end for**
17: From all valid combinations of (S_{IF}, S_{PF}, S_{BF}), find the one that minimizes rate R
18: **if** no candidate path meet Q_T irrespective of any redundancy is added **then**
19: Select the one level lower than Q_T and update Q_T
20: Repeat the steps 5 through 18
21: **end if**

3.7. Performance evaluation

To evaluate the proposed FEC algorithm, experiments were carried out by a component-based network simulator using Wireless Simulation Environment for Multimedia Networks (WiSE-MNet) [73]. Based on OMNet++ [74] [75] and Castalia [76], WiSE-MNet is a discrete event simulator engine for wireless multimedia sensor networks. The modules in WiSE-MNet were written in C++ with high level network description (NED) language with the use of configuration file to assemble modules. The network simulation is carried out over point-to point (single-hop) 802.11 channels in distributed coordination function where there is a direct connection between both nodes. IEEE 802.11 standard support the frequency of 2.4 GHz with the communication range under 100 m and the maximum data rate support of 100 Mbps for Wireless Local Area Networks (WLANs). All the videos in Table 1 were used for this evaluation. Each test case was simulated 20 times and the averages values were recorded.

FEC schemes from two other studies were implemented for comparison. The first study, QoE-aware FEC [33], was designed for providing QoE support for multimedia sensor networks using FEC. QoE-aware FEC creates redundant packets for I-frames and first 50% of P-frames. B-frames and the last 50% of P-frames are transmitted without any redundancy. While several scenarios were studied in [33], we simulated scenarios 5 (QoE-aware FEC (80, 0)) and 7 (QoE-aware FEC (100, 0)) which showed the best performance, with RS coding of 80% and 100% of redundancy, respectively. Both scenarios sent the last 50% of P-Frames without redundancy as the loss in these frames cause lower perceptual video distortion. The second scheme is an adaptive video-aware FEC approach (Video-awareness [31]) that divides videos into several content types and adds a fix amount of redundancy to both I-frame and P-frames depending on the content types.

We have performed a set of simulations to discover: i) how much network overhead is generated from the proposed algorithm, and ii) how well the proposed algorithm can help to

improve perceptual video quality at the receiver end. The average number of redundant packets is used to evaluate network overhead. Figure 11 shows the average network overhead for the testing videos under different network conditions (1%, 5%, and 10% PLRs) under the source coding bit rate of 1.5 Mbps. Figure 12 shows the average network overhead for 5% PLR for two different source coding rates: 1.5 Mbps and 4 Mbps. The average value of network overhead of QoE-aware FEC (100, 0) was 39.45% and that of QoE-aware FEC (80, 0) was 33.99% under the 1.5 Mbps source coding rate. For the video-aware FEC scheme, the average network overhead was 32.30%. Our proposed algorithm resulted in 19.15% overhead, considerably lower than the other three schemes. Similarly, for a source coding rate of 4.0 Mbps, the network overhead was 51.94% for QoE-aware FEC (100, 0), 43.89% QoE-aware FEC (80, 0), and 37.68% for video-aware FEC. In comparison, our proposed algorithm produced a better result of 22.64% overhead.

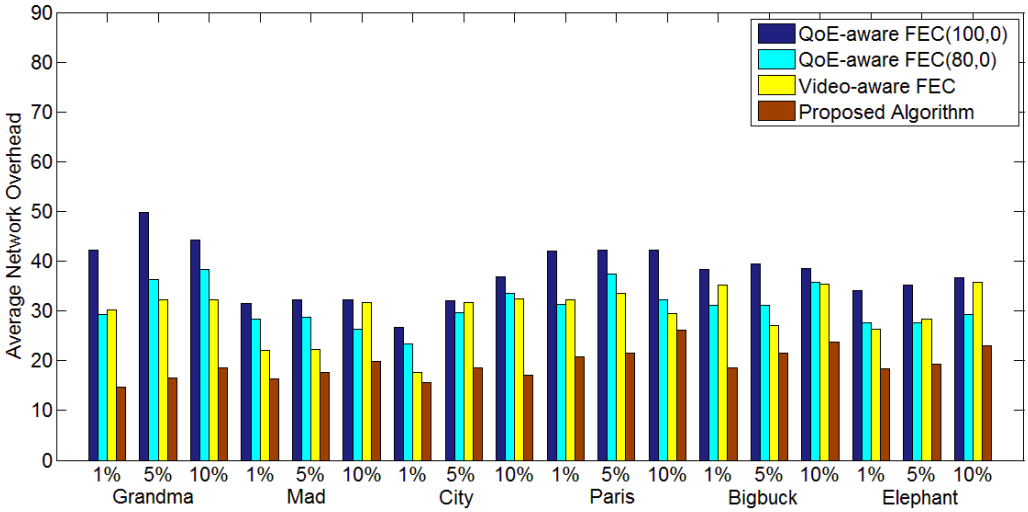

Figure 11. Average network overhead for different testing videos for 1.5 Mbps bitrate

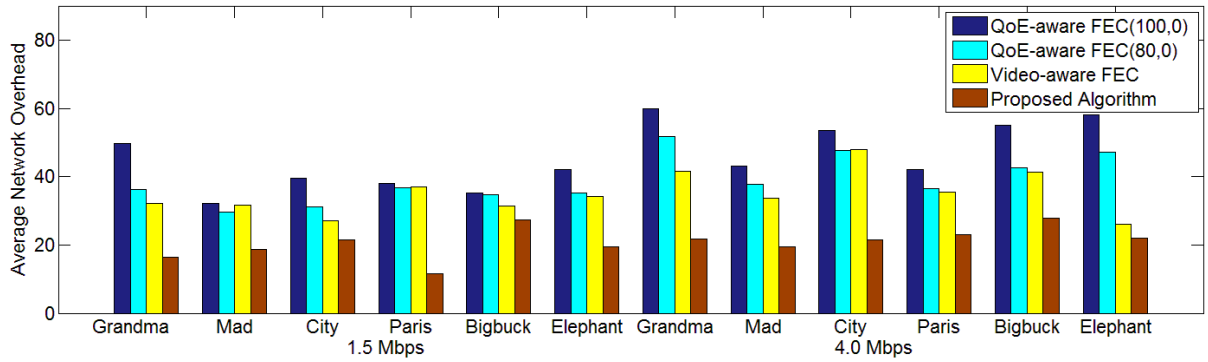

Figure 12. Average network overhead for 5% PLR for both 1.5Mbps and 4 Mbps

The perceptual video quality at the receiver end was evaluated using SSIM and VQM. Figure 13 and Figure 14 show the results of average SSIM and average VQM of all the transmitted testing videos with network packet loss scenarios. For SSIM, the value can range from 0 to 1, and a value closer to 1 indicate a better video quality. The average SSIM values achieved with QoE-aware FEC (100, 0) for different videos was 0.8404, which outnumbered QoE-aware FEC (80, 0) with 0.7373, video-aware FEC with 0.7411. Our algorithm resulted in an average SSIM of 0.8161, which is close to the best performance generated by QoE-aware FEC (100, 0). As for VQM values, videos with better quality score results in less magnitude. The average VQM values were 2.45 for QoEaware FEC (100, 0), 2.77 for QoE-aware FEC (80, 0), 2.91 for video-aware FEC, and 1.39 for the proposed algorithm. The proposed algorithm outperformed the other three algorithms in perceptual quality evaluated by VQM.

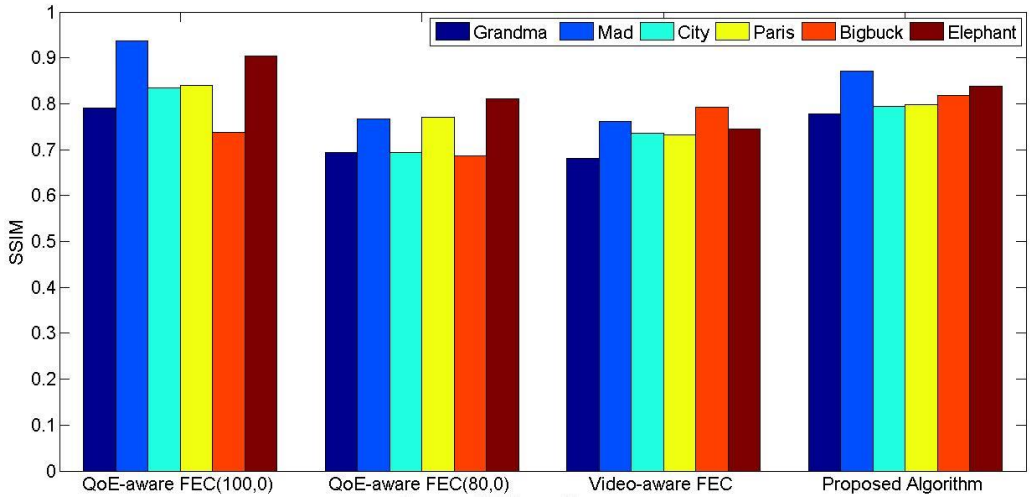

Figure 13. Comparison of average SSIM for different video sequences

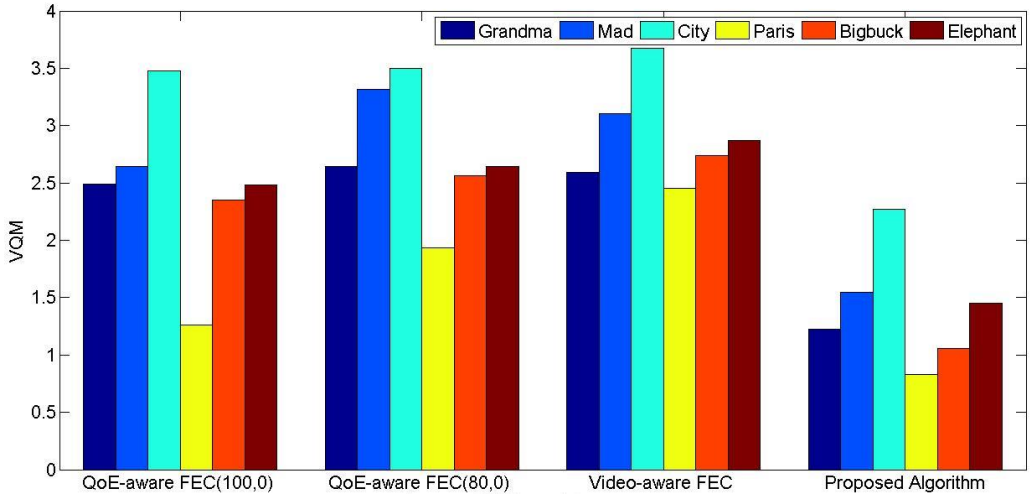

Figure 14. Comparison of average VQM for different video sequences

Next, the proposed algorithm was evaluated under different user requirements. The perceptual quality threshold was set to “good” and “very good”, and then results generated by the proposed algorithm were collected. Figure 15 represents the percentage of videos exceeding threshold of perceptual quality of “good” and "very good" under source coding rate of 1.5Mbps and 4Mbps for a PLR scenario of 5%. On average for all the videos used for testing, the criteria of “good” was met 89.44% for 1.5 Mbps vs 91.94% for 4 Mbps, whereas the threshold of “very good” was met 88.89% for 1.5 Mbps vs 90.28% for 4 Mbps bitrate for same PLR. In summary, the

simulation results demonstrate that the proposed FEC algorithm improves perceptual video quality and at the same time reduces network overhead. It is a promising solution for providing QoE support for video over wireless embedded devices.

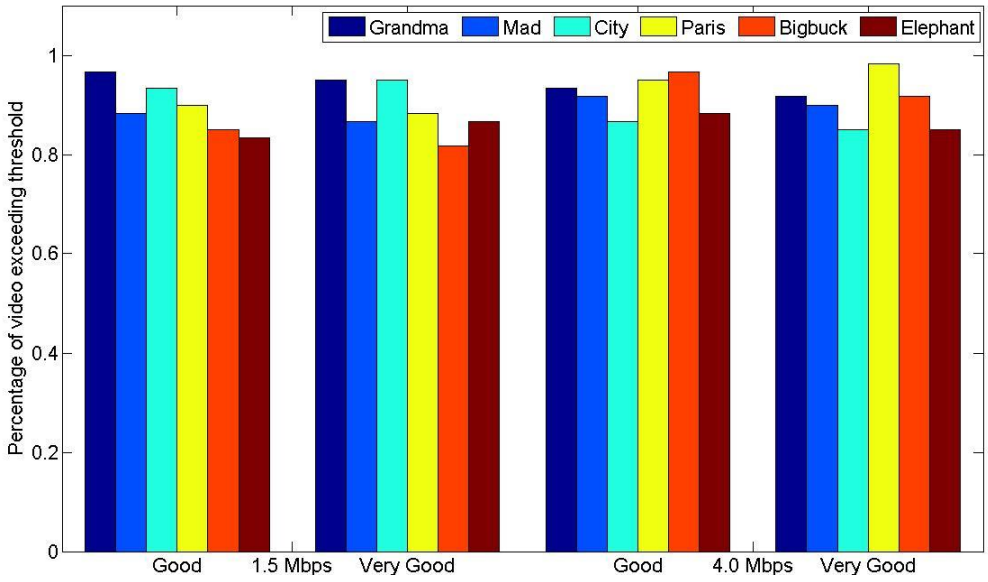

Figure 15. Comparison of percentage of videos exceeding good and very good thresholds

3.8 Conclusion

In this chapter, we have introduced a reference-free and lightweight model for perceptual video quality prediction, and based on the model, we have developed an energy-efficient and content-aware FEC scheme. The quality prediction model reveals the nonlinear relationships between perceptual quality and features related to video content, source coding parameters, and network conditions. Experimental results have shown that it achieve good prediction performance on videos with different characteristics. The proposed FEC scheme is the first study that leverages the quantitative relationship between FEC parameters and perceptual quality. It can reduce network overhead while maintaining satisfactory perceptual quality for end users.

CHAPTER 4. THIN SLICE PERCEPTION: INFERENCE OF VIDEO CONFIDENCE MEASURE

4.1. Introduction

Many assessments and standardized efforts have been made to design models and algorithms to predict the perceptual video quality for different bitrate constraints and network conditions [50]. The transmission channel bandwidth and unreliable network conditions are major causes of degradation in video quality, as the loss of even a few packets can affect both spatial and temporal perceptual quality for video sequences. Changes in source coding parameters and losses in network packets will result in the degradation of perceived video quality. Recent research has addressed the improvement of perceptual video quality by addressing the artifacts caused by source coding, channel coding, and the combination of the two.

The impact of network conditions, etc. on video quality is typically assessed by measuring *perceived* video quality by estimating a mean opinion score gathered from subjective tests using human observers. Ratings in these tasks may vary from bad to excellent quality. However, degradations in video quality may also impact other aspects of the viewer's subjective experience of a video, including their ability to make judgments about the content of the video. For example, "thin slices," of non-verbal behavior (brief instances of facial gestures and body language), typically provide sufficient information for observers to make reliable high-level social inferences. Given short videos depicting natural behavior, observers can typically provide consistent ratings of personality characteristics including (but not limited to): teaching effectiveness [77] [78], sales performance [79], and sexual orientation [80] [81]. Thin-slice judgments of personality traits correlate well with subjective and objective evaluations of the traits under consideration, and in some instances observers can provide robust thin-slice judgments from only a few seconds of

video. Given that observers can make high-level judgments like these from short videos of human behavior, how do variations in perceptual quality affect these judgments? Do changes in perceptual quality directly impact how observers make thin-slice judgments based on the behaviors in short video sequences? In some cases, severely degraded faces [82] [83] and bodies [84] carry enough information to support a range of recognition tasks, suggesting that perceptual quality may not predict thin-slice judgments. However, the impact of the specific degradations that commonly affect videos subject to network transmission have not been examined.

4.2. Motivation

The perceptual quality score can be predicted by perception quality model, which is used to account decisions of participants who classify whether the video quality is bad or good. We examined whether or not a model designed to account for subjective quality ratings could also account for the variation in thin-slice judgments in videos that were subject to transmission-related artifacts. Though we are using our existing perceptual quality model [67], we are not focusing on measuring perceptual video quality in this paper. Rather, we are focusing on people's judgments of high-level social variables – in this instance, the confidence of a speaker depicted in a short video. It stands to reason that high level personality characteristics, such as the perception of a speaker's behavioral traits, may also be changed with the bandwidth and packet loss effects. In prior work, we have developed an efficient and light-weight video quality prediction model [67] through partial parsing of compressed video bitstreams.

There has been considerable research on the use of thin slice judgments, but no study has demonstrated the predictive validity of the assessor's confidence measure in evaluating the speaker's confidence with respect to change in packet loss ratios and bandwidth. We measured two parameters in our study: (1) the confidence of a single individual present in all the videos

called speaker's confidence (SC), and (2) how confident the participants are in evaluating a speaker's confidence (PC). The present research represents a new entry into the domain of high level judgments, i.e. PC and SC by the use of our existing perception quality model. We examined initial impressions of people formed by participants based on very brief observations, or thin slices, after watching only short video clips. Our experimental results indicate that there is significant consensus and high correlation among participants when asked to evaluate the PC in short video clips.

In this paper, we tested whether this existing model for video quality estimation can be used to predict people's subjective estimates of SC, as well as their own PC in the rating they assigned. We find that our model can predict how observers' judgments changed as a function of perceptual quality and how the change in bandwidth and network packet loss variations affect high-level and non-verbal behaviors. We continue by describing the features our model uses to predict perceptual quality (in previous work) and "thin-slice" judgments of video content (the current work). Earlier, the model was trained on a distorted video database consisting of videos with varying content characteristics, encoding parameters, and packet loss levels.

The motivation of this piece of work is as follows:

- We utilized the same set of features that were earlier used to depict video content characteristics, distortion caused by lossy compression, and distortion caused by network transmission to estimate SC measure.
- We derive quantitative relationships between source and channel coding parameters and how they are related with the help of a trained ANN algorithm to see the correlation of judgments about the content of the video.

- We also utilized a decision tree model to identify the relationships between the features identified earlier. Unfortunately, the relationships were not visible in ANN due to the black-box nature of ANN.

4.3. Method

4.3.1. Subjects

As a part of the experiments, a total of 25 adult subjects were recruited to evaluate SC measure. In prior work, to provide sufficient data for training the video quality model [67], 100 undergraduate students from the Psychology department at North Dakota State University (NDSU) were enrolled. All enrolled students were in the age range from 18 to 26. All participants reported normal or corrected to normal vision and were unaware of the design and purpose of the study. We obtained written informed consent from all participants.

4.3.2. Stimuli

In our experiments, the video clips showed naturally occurring behavior, were not artificially created, and were originally downloaded from the Xiph collection of videos [54]. Earlier, the perceptual quality model was trained by 6 different video clips that covered a wide variety of combinations of low, medium, and high motion with low and high spatial details as shown in Table 8. We have generated a database of distorted videos taking into account different test conditions. As the impact of encoder and network distortions are very dependent on video content, the training videos were chosen in a way to represent different spatial and temporal complexity scenarios. Each video was encoded with different bitrates, frame rates, GOP structures, and with different packet error rates for network impairments. These tests follow ITU-T recommendations, using the single-stimulus absolute category rating method with a quality scale of 1-9 [47].

4.3.3. Procedure

A total of 288 videos were generated, taking into consideration the combinations of GOP structure, frame rate, bitrate, and packet loss ratio (PLR). The whole video set was divided into 4 groups. Every group of videos contained 72 videos each. Each piece of video was evaluated by 25 subjects.

Table 8. Video clustering based on spatial and temporal complexity

	Low Motion	Medium Motion	High Motion
Low Spatial	Student	Stock	Deadline
High Spatial	Dinner	Sintel	Factory

4.4. Results

We adopted the screening method recommended by BT.500-11 [56] to screen our collected data. The first step of the screening process was the calculation of individual mean score μ_{ijkr} and overall mean score μ_{jkr} based on the score of observer *i* on test condition *j* for sequence *k* and for repetition *r*. The second step was to assess the reliability of experiments and was done by the computation of 95% confidence intervals $\left[\mu_{jkr} - \delta_{jkr}, \mu_{jkr} + \delta_{jkr}\right]$ based on standard deviation δ_{jkr} and size of each sample N from raw scores. The kurtosis coefficient β_{2jkr} was computed to determine whether the distribution is normal or not. Following that, the next step was to compute P_i and Q_i and to determine whether to reject the observer *i* record based on the ratios of $\frac{\mathrm{P}_i+\mathrm{Q}_i}{J.K.R} > 0.05$ and $\left|\frac{\mathrm{P}_i-\mathrm{Q}_i}{\mathrm{P}_i+\mathrm{Q}_i}\right| < 0.3$. Equations (13) to (22) explain the necessary steps taken for the rejection of observers during the filtration process from the raw data.

$$\mu_{jkr} = \frac{1}{N}\sum_{i=1}^{N} \mu_{ijkr} \quad (13)$$

$$\delta_{jkr} = 1.96\frac{S_{jkr}}{\sqrt{N}} \quad (14)$$

$$S_{jkr} = \sqrt{\sum_{i=1}^{N}\frac{\left(\mu_{jkr} - \mu_{ijkr}\right)^2}{(N-1)}} \quad (15)$$

$$B_{2jkr} = \frac{m_4}{(m_2)^2} \quad (16)$$

$$where\ m_x = \frac{\sum_{i=1}^{N}\left(\mu_{jkr} - \mu_{ijkr}\right)^2}{N} \quad (17)$$

$if\ 2 \le B_{2jkr} \le 4, then$:

$$if \mu_{ijkr} \ge \mu_{jkr} + 2S_{jkr}\ then\ \mathrm{P}_i = \mathrm{P}_i + 1 \quad (18)$$

$$if \mu_{ijkr} \ge \mu_{jkr} - 2S_{jkr}\ then\ \mathrm{Q}_i = \mathrm{Q}_i + 1 \quad (19)$$

$else$

$$if \mu_{ijkr} \ge \mu_{jkr} + \sqrt{20}S_{jkr}\ then\ \mathrm{P}_i = \mathrm{P}_i + 1 \quad (20)$$

$$if \mu_{ijkr} \ge \mu_{jkr} - \sqrt{20}S_{jkr}\ then\ \mathrm{Q}_i = \mathrm{Q}_i + 1 \quad (21)$$

$$if\ \frac{\mathrm{P}_i + \mathrm{Q}_i}{J.K.R} > 0.05\ and\ \left|\frac{\mathrm{P}_i - \mathrm{Q}_i}{\mathrm{P}_i + \mathrm{Q}_i}\right| < 0.3\ then\ reject\ observer\ i \quad (22)$$

where N: number of observers, j: number of test conditions, k: number of test clips, r: number of repetitions

We identified and neglected all those viewers whose ratings were not consistent with the ratings of other viewers, and after filtering these unreliable results, we obtained SC for each video sequence. The subjective video quality prediction model evaluated the new set of videos to estimate predicted MOS and identified the correlation of SC with respect to MOS. We identified and neglected 2 viewers during the training and 1 during the testing phase as their ratings were not consistent with the ratings of other viewers. We filtered the unreliable results and obtained estimated MOS for each video sequence by using our perceptual quality model that identified all the features that can be obtained from light-weight parsing of the compressed bitstream. This

procedure enabled the extraction of required information to depict content related effects without introducing too much computation for decoding the MBs and reconstructing the pixels.

For the testing phase, 25 more participants were recruited, none of whom participated in the prior task evaluation. The testing videos were also chosen in a way to have one presenter in each video and were also chosen from the Xiph collection of videos [54]. All the participants were told that they were about to see several short videos of 20 seconds length and would be asked to tell how confident was the presenter in the video on a scale of [1-9]. The evaluator's confidence in their ability to evaluate was also measured on a scale of [1-9]. The experimenter described the physical set-up of the recordings and made clear that there would be no soundtrack and that the participants would see each presenter in the video repeated multiple times during the experiment. During the task, the test clips were presented in a randomized order using custom software written in MATLAB. Each video clip played for 20 seconds, after which a response screen with the scale [1-9] appeared ((1) asking how confident was the presenter and (2) how confident was the participant in evaluating the presenter based on the scene). The participants responded by clicking on a dropdown menu, and the response times were not recorded.

The training data of the decision tree were the same dataset of six different video clips used in Table 1. Moreover, the features that depict the video content characteristics and encoding distortions as mentioned in Section 2.3 were used in the decision tree. Figure 16 illustrates the decision tree for the training videos and all the conditions that need to be fulfilled for decision. The decision data are evaluated one by one by traversing the path specified from top to bottom or otherwise. We grouped the confidence metrics into more coarse-grained bins and selected the granularity, during which the model predicted engagement by appropriately setting the number of classes to six (poor, bad, satisfactory, good, very good, and excellent). After building a suitable

tree, the traversal was done by visiting different branches that indicated the maximum video confidence specified by the excellent measure. Naturally, the prediction accuracy would diminish if the model were forecasting at a higher granularity. Therefore, we used similar domain-specific discrete classes for the SC measure to bin the different perceptual quality metrics.

The correlation measures for SC were 0.9636 and 0.9811 respectively, based on the choice of method (ANN or decision tree). The results suggested that the consistent judgments about one's personality confidence could be made based on the brief observations of participants. After brief observation of only 20 seconds of each test clip, a few high-level judgments about strangers were molded very rapidly. The participant's evaluated confidence measured were estimated with the use of the perceptual video quality model.

4.5. Discussion

Due to the black box nature of ANN, we are unable to find the relationships among the factors. For this reason, we utilize the decision tree model that identifies the most important factors and the relationships among them. Unlike the relationships identified to measure the perceptual video quality mentioned in Section 3.3, the most important factor for SC measure is $R_{inter/intra}$, which is on top of the decision tree as shown in Figure 16. As a test case, to traverse the path of excellent in the decision tree, besides $R_{inter/intra} \geq 0.5247$, we have to make sure that other source coding parameters and the test conditions, such as $MB_{inter} < 0.9228$, $QP_{avg} \geq 7.5359$, and $B_{inter} \geq 5.2941$, are maintained. The aforementioned case is particularly true when the loss in packets is very low. To achieve a confidence scale of very good, network parameters, such as DSlice% ≥ 12 and source coding parameter $B_{intra} \geq 0.4744$ and $B_{inter} < 5.2941$, are maintained. Note that the maximum PLR is 10% in all the cases, so the presence of an upper bound DSlice% and DFrame% is required as well. All other conditions of QP_{avg}, MB_{inter}, and $R_{intra/inter}$ should also be validated as previously mentioned.

Figure 16 also shows 4 different paths for good and fair categories and can be achieved by following the test conditions by traversing the path in the decision tree. There is one path for each bad and poor confidence measures, but the network condition DFrame% ≥11.5 is common in both paths. For poor confidence measure the ranges of $BL_{avg} < 12.8948$, $7.7976 \leq QP_{avg} < 26.5117$, and $R_{inter/intra}<0.52471$ must be maintained. For bad confidence measure both source and channel coding parameters $B_{intra} \geq 0.32822$ and $BL_{avg} \geq 12.8948$ are required. Again there is an upper limit on BL_{avg} values, as the maximum PLR value used in the experiments is 10%. The criteria used to divide the confidence into different classes is presented in Table 9.

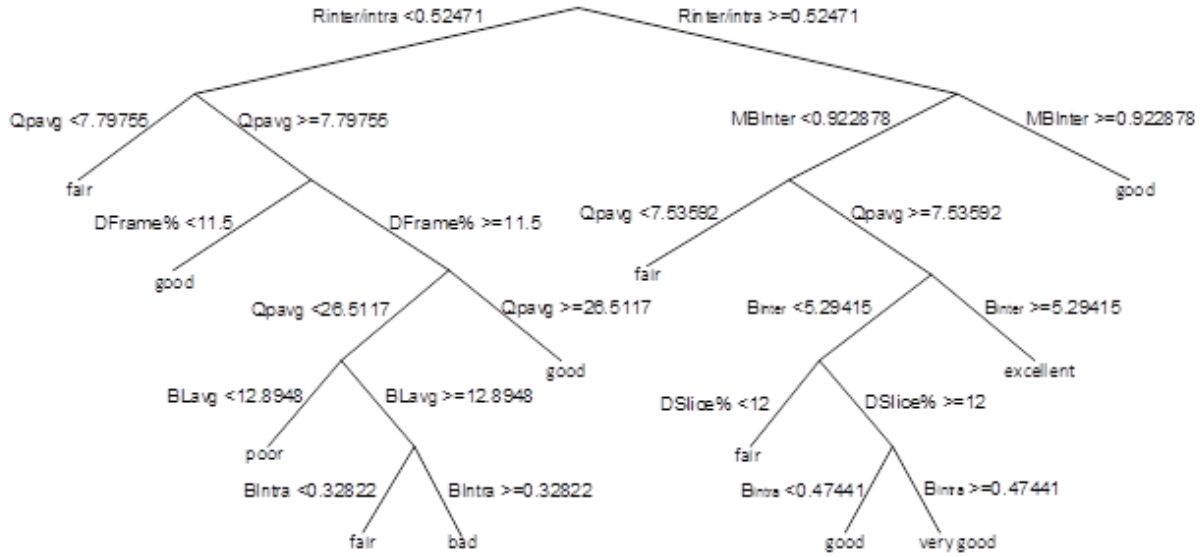

Figure 16. The decision tree to estimate SC

The visual examination of the decision tree and network of strong relationships between input variables can clearly be seen as the branches in the decision tree help in predicting, explaining, and classifying the target outcome. The decision tree supports the collapse and combination of a set of parameters into ranges that are aligned with the values of the selected target variable. In the context of multiple influences like the features mentioned above and based on the relationships of source and channel coding parameters, the object of analysis is $R_{inter/intra}$. $R_{inter/intra}$ serves as a target field and is present as the root node in the decision tree. Other source coding

parameters, such as QP_{avg} and MB_{inter}, are present on the first level of the decision tree. The presence of a source coding parameter at the top of the tree and the majority of other source coding parameters in the leaf nodes suggest that strong influence of the source coding process in the evaluation of SC. It is also evident from Figure 16 that source coding parameters control the criteria of excellent category mentioned. The factors related to network coding process can be seen in the middle portion of the decision tree also influence the estimates of SC and can equally influence the SC measures in the tree.

Table 9. Criteria used to categorize SC measures

Class	Range of confidence values
Poor	< 3.3098
Bad	≥ 3.3098 and < 4.3094
Satisfactory	≥ 4.3094 and < 5.5589
Good	≥ 5.5589 and < 6.0587
Very good	≥ 6.0587 and < 7.3083
Excellent	≥ 7.3083

Our task reveals several intriguing aspects of non-verbal high level judgments. First, our model is offering a good estimate of SC that correlates with perceptual quality results in estimating how confident a person is if participants evaluate that person's behavior. We identified exactly what features are used to make high level judgments in thin-slice tasks where participants perform the task by relying solely on confidence measures of the presenter in the video clips and how confident they are after watching each video clip. Figure 17 shows a portion of our test results, where the MOS of six videos are categorized under combinations of bit rates and PLRs (with 30 fps frame rate and IBBP GOP structure). The evidence shows that there is substantial variation of

the MOS for different videos. Therefore, content characteristics play an important role to achieve accurate estimations of video confidence measures.

Since the individual ratings are grouped together, and for the ease of interpretation, Figure 18 depicts view-counts as the function of the mean ratings of each test clip aggregated across bitrate and PLR. We averaged the mean ratings of each video across all the participants. To address the potential concern that these ratings merely reflect the perceived confidence of individual participants, we showed the clip to 25 participants and asked them to rate each video clip in the experiment.

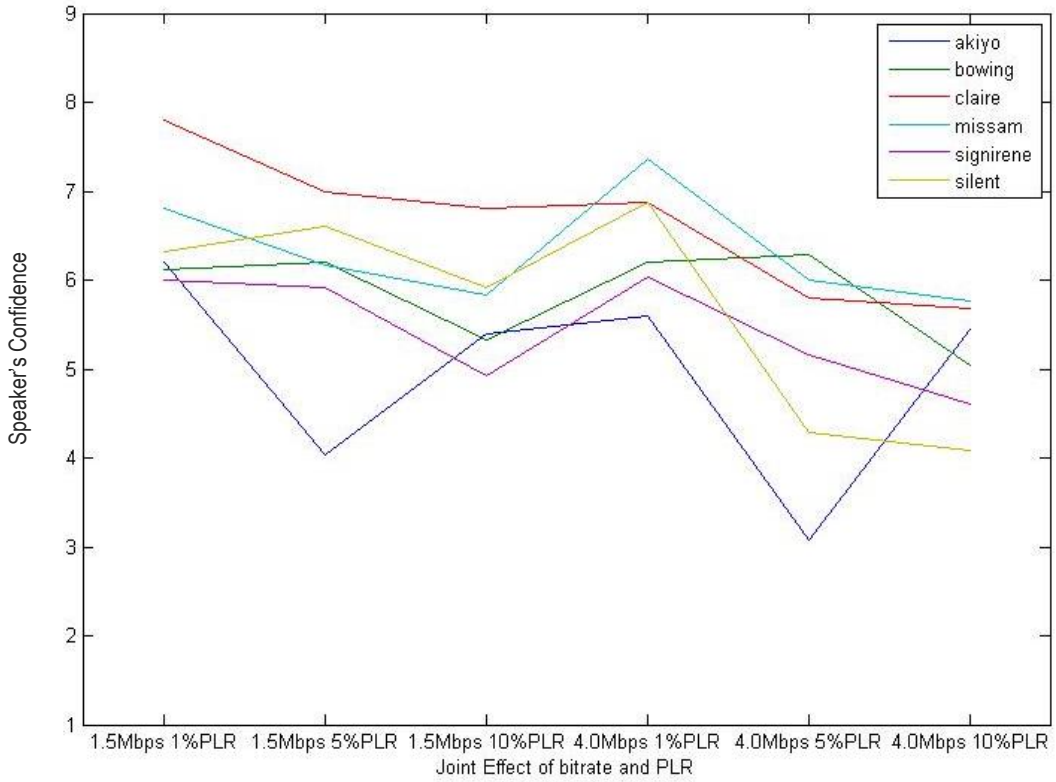

Figure 17. Speaker's confidence measure

The results of this study demonstrate that thin slices can be used to assess confidence measures as computed by participants' evaluations that correlate with the predictions made by the video perception quality model. The judges were able to make these distinctions solely on the basis of 20-second long video clips. These results are consistent with previous studies suggesting that

the thin-slice methodology is more valuable for evaluating interpersonal skills than non-interpersonal, task related skills [77][78][79]. This may be due to the greater observability of information relating to interpersonal skills through nonverbal channels, as compared to task-related skills.

To evaluate video confidence measure, a different set of videos were used and the results show a match of 96.36% and 98.11% correlation based on the neural network and decision tree models as shown in Figure 17. Prediction performance defined as the association between a witness's thin-slice evaluation of a concept (e.g., extraversion) and a criterion metric computed by a group of stimulus people (targets).

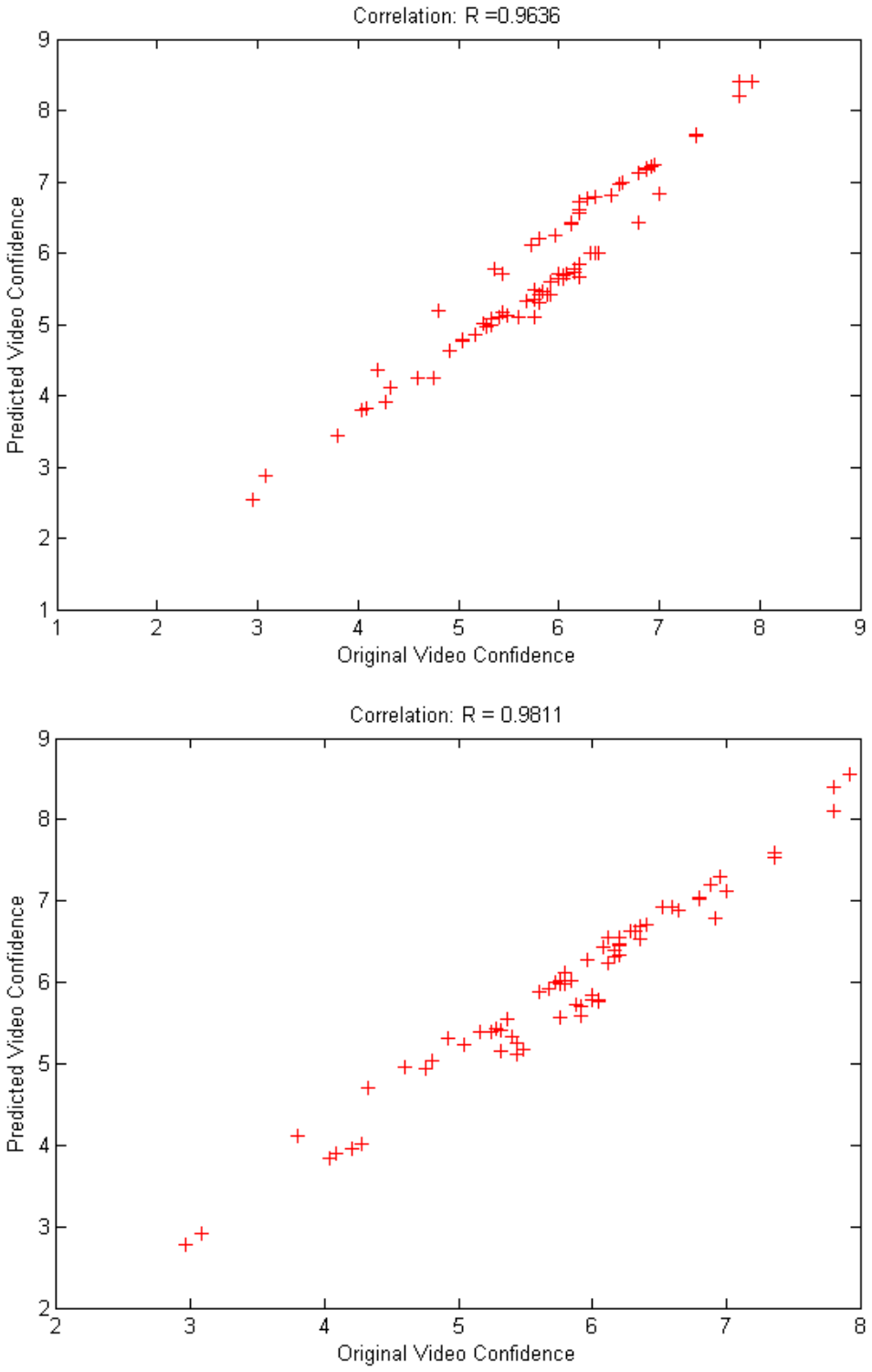

Figure 17: Prediction performance (a): Neural network method, (b): Decision tree method

CHAPTER 5. CONCLUSION AND FUTURE WORK

The wireless network conditions impose strict requirements on coding technologies and there is a much need to support an acceptable perceptual video quality communication over wireless networks. With the rapid increasing demands of video content, it is becoming possible to deploy video services to end users. Given the growing interest in multimedia services delivery, the prediction of accurate human perceptual quality judgements and improvement in perceptual quality in juxtaposition has become a very challenging topic and active area of research. Moreover, video communications frameworks based on H.264/AVC are affected by transmission errors and there is a need to provide error-resilience by adding coding redundancy in network channel. Therefore, to provide services to network users, it is necessary to provide effective control over channel errors.

In this thesis, we first identified the features that affect perceptual video quality based on different packet loss scenarios and changing bandwidth. The next step was the development of a QoE driven machine learning based (neural network and decision tree based) video quality measure and an error control scheme for video communication in wireless multimedia networks. During the research, we recognized that our perceptual video quality model could also account for the variation in thin-slice judgments in videos that are subject to transmission-related artifacts.

The proposed method uses a set of low-level bitstream features in a machine learning framework to learn a mapping from the features to subjective mean opinion scores. Since, the relationship among bitstream features was not evident in neural network. Therefore, a decision tree general framework was designed to address the problem of minimizing the transmission energy required to provide an acceptable level of video quality.

Our proposed algorithm is simple, powerful, and offers low complexity in contrast with many complicated quality assessment systems. It supplies good perceptual video quality prediction accuracy. Through analysis and simulations, we show that our proposed algorithm offers low transmission network overhead, high transmission efficiency, and less energy consumption in wireless networks under QoE requirements. As an application of the model, we also observed the impact of how observers make thin-slice judgments based on the behaviors in short video sequences. The experiments on a video quality dataset shows that our method has comparable performance with the current state of art.

There are several promising perceptual quality models proposed and implemented based on the knowledge of video and image processing, compression, the human vision system, and psychological effects. A plethora of research is dedicated to measure perceptual video quality to establish mathematical models. Still, there is a wide scope for development and improvement of reliable video quality metrics because some issues regarding several aspects of perceptual quality have not yet been resolved satisfactorily. There are many challenges remaining to be resolved by employing advanced mathematical models or technologies. The future work foreseen for this thesis work are:

- To extend and enhance the accuracy by dynamic adjusting of weights through determining region of interests.
- To combine low computational cost to ensure minimal quality distortion while satisfying human perception quality with the combination of joint source and channel coding.
- To scale the current model for other bit rates and a wide range of practical distortion types and extend the model for other application area such as for mobile streaming services.

REFERENCES

[1] J. Ohm, "Multimedia communication technology: representation, transmission and identification of multimedia signals: signals and communication technology series," Springer, Berlin, Heidelberg, Germany, 2004.

[2] G. Sullivan and T. Wiegand, "Video compression - from concepts to the H.264/AVC standard," Proceedings of IEEE, vol. 93, no. 1, pp. 18-31, Jan. 2005

[3] D. Eckhardt, "An internet-style approach to managing wireless link errors," Ph.D. Thesis, Carnegie Mellon University, Pittsburg, PA, May 2002.

[4] W. Lin, C. Kuo, "Perceptual visual quality metrics: a survey," Journal of Visual Communication and Image Representation, vol. 22, no.4, pp. 297-312, May 2011.

[5] A. Shahriar, "Digital video concepts, methods, and metrics: quality, compression performance, and power trade-off analysis," Apress, 2014.

[6] M. Shahid, A. Rossholm, B. Lövström, and H. Zepernick, "No-reference image and video quality assessment: a classification and review of recent approaches," EURASIP Journal on Image and Video Processing, no.1, 40 pp., 2014.

[7] I. Levin, S. Schneider, and G. Gaeth, "All frames are not created equal: A typology and critical analysis of framing effects," Organizational behavior and human decision processes, vol. 76, no.2, pp. 149-188.

[8] X. Yang, C. Zhu, Z. Li, X. Lin, G. Feng, S. Wu, and N. Ling, "Unequal loss protection for robust transmission of motion compensated video over the internet," Signal Processing: Image Communication, vol. 18, no. 3, pp. 157-167, Mar. 2003.

[9] S. Winkler, "Digital video quality: vision models and metrics," John Wiley and Sons, Ltd. 2005.

[10] M. Igarta, "A study of MPEG-2 and H.264 video coding," Ph.D. Thesis, Purdue University, West Lafayette, IN, Dec. 2004.

[11] I. Richardson, "The H.264 advanced video compression standard, second edition," John Wiley and Sons, 2011.

[12] S. Chikkerur, S. Vijay, R. Martin, and L. Karam, "Objective video quality assessment methods: a classification, review, and performance comparison," IEEE Transactions on Broadcasting, vol. 57, no. 2, pp. 165-182, June 2011.

[13] J. Robinson, P. Shaver, and L. Wrightsman, "Personality and social psychological attitudes," Edited by John P. Robinson and Lawrence S. Wrightsman, vol. 1, Gulf Professional Publishing, 1991.

[14] P. Sander and L. Sanders, "Measuring confidence in academic study: A summary report," University of Wales Institute, Cardiff, UK.

[15] H. Koumaras, A. Kourtis, C. Lin, and C. Shieh, "A theoretical framework for end-to-end video quality prediction of MPEG-based sequences," 3rd International Conference on Networking and Services, pp. 62-65, June 2007.

[16] R. Feghali, F. Speranza, D. Wang, and A. Vincent, "Video quality metric for bitrate control via joint adjustment of quantization and frame rate," IEEE Transactions on Broadcasting, vol. 53, no. 1, pp. 441-446, Mar. 2007.

[17] A. Eden, "No-reference image quality analysis for compressed video sequences," IEEE Transactions on Broadcasting, vol. 54, no. 3, pp. 691-697, Sep. 2008.

[18] Q. Huynh-Thu and M. Ghanbari, "Temporal aspect of perceived quality in mobile video broadcasting," IEEE Transactions on Broadcasting, vol. 54, no. 3, pp. 641-651, Sep. 2008.

[19] G. Zhai, J. Cai, W. Lin, X. Yang, and W. Zhang, "Three dimensional scalable video adaptation via user-end perceptual quality assessment," IEEE Transactions on Broadcasting, Special Issue on Quality Issues in Multimedia Broadcasting, vol. 54, no. 3, pp. 719-727, Sep. 2008.

[20] K. Seshadrinathan and A. Bovik, "Motion tuned spatio-temporal quality assessment of natural videos," IEEE Transactions on Image Processing, vol. 19, no. 2, pp. 335-350, Feb. 2010.

[21] S. Tao, J. Apostolopoulos, and R. Guerin, "Real-time monitoring of video quality in IP networks," IEEE/ACM Transactions on Networks, vol. 16, no.5, pp. 1052-1065, Oct. 2008.

[22] P. Calyam, E. Ekicio, C. Lee, M. Haffner, and N. Howes, "A GAP-model based framework for online VVoIP QoE measurement," Journal of Communication and Networks, vol. 9, no. 4, pp. 446-456, Dec. 2007.

[23] G. Cermak, "Subjective video quality as a function of bit rate, frame rate, packet loss rate, and codec," 1st International Workshop on Quality of Multimedia Experience (QoMEX), pp. 41-46, Jul. 2009.

[24] A. Moorthy, K. Seshadrinathan, R. Soundararajan, and A. Bovik, "Wireless video quality assessment: A study of subjective scores and objective algorithms," IEEE Transactions on Circuits, Systems, and Video Technology, vol. 20, no. 4, pp. 513-516, Apr. 2010.

[25] K. Seshadrinathan, R. Soundararajan, A. Bovik, and L. Cormack, "Study of subjective and objective quality assessment of video," IEEE Transactions on Image Processing, vol. 19, no. 6, pp. 1427-1441, Jun. 2010.

[26] A. Khan, L. Sun, and E. Ifeachor, "QoE prediction model and its application in video quality adaptation over UMTS networks," IEEE Transactions on Multimedia, vol. 14, no. 2, pp. 431-442, Apr. 2012.

[27] Y. Wang, S. Wenger, J. Wen, and A. Katsaggelos, "Error resilient video coding techniques," IEEE Signal Processing Magazine, vol. 17, no. 4, pp. 61-82, 2000.

[28] I. Akyildiz, S. Weilian, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: a survey," Computer Networks, vol. 38, no. 4, pp. 393-422, 2002.

[29] W. Dargie and C. Poellabauer, "Fundamentals of wireless sensor networks: theory and practice," Wiley Publishing, 2010.

[30] Y. Yuan, B. Cockburn, T. Sikora, and M. Mandal. "A GoP based FEC technique for packet based video streaming," 10th WSEAS International Conference on Communications (ICCOM'06), Stevens Point, Wisconsin, USA, pp.187-192, 2006.

[31] R. Immich, E. Cerqueira, and M. Curado, "Adaptive video-aware FEC based mechanism with unequal error protection scheme," 28th Annual ACM Symposium on Applied Computing, pp. 981–988, Mar. 2013.

[32] H. Liu, H. Ma, M. Zarki, and S. Gupta, "Error control schemes for networks: An overview," Mobile Networks and Applications, vol. 2, no. 2, pp: 167-182, 1997.

[33] Z. Zhao, D. Rosario, R. Immich, and M. Curado, "QoE-aware FEC mechanism for intrusion detection in multi-tier wireless multimedia sensor networks," 1st International Workshop on Wireless Multimedia Sensor Networks, 2012.

[34] C. Fujiwara, M. Kasahara, K. Yamashita, and T. Namekawa, "Evaluation of error control techniques in both dependent and independent error channels," IEEE Transactions on Communication, vol. 26, no. 6, pp 785-793, 1978.

[35] S. Reddey and J. Robinson, “A construction for convolutional codes using block codes,” Information and Control, vol.12, no. 1, pp 55-70, 1968.

[36] Q. Li and M. Schaar, “Providing adaptive QoS to layered video over wireless local area networks through real-time retry limit adaptation,” IEEE Transactions on Multimedia, vol. 6, no.2, pp. 278-290.

[37] K. Stuhlmuller, N. Farber, M. Link, and B. Girod, “Analysis of video transmission over lossy channels,” IEEE Journal on Selected Areas in Communications, vol. 18, no. 6, pp. 1012-1032, 2000.

[38] T. Mahmoodi, V. Friderikos, O. Holland, and A. Aghvami, “Optimal design of forward error correction for fairness maximisation among transmission control protocol flavours over wireless networks,” IET Communication, vol. 4, no. 10, pp. 1196-1206, 2010.

[39] Q. Mao, B. Xu, and Y. Qin, “A new scheme to improve the quality of compressed image transmission by turbo unequal error protection codes,” 7th International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), pp. 226-229, 2011.

[40] A. Naghdinezhad, M. Hashemi, and O. Fatemi, “A novel adaptive unequal error protection method for scalable video over wireless networks,” IEEE International Symposium on Consumer Electronics (ISCE 2007), pp. 1-6, June 2007.

[41] W. Wang, D. Peng, H. Wang, H. Sharif, and H. Chen, “Energy-constrained distortion reduction optimization for wavelet-based coded image transmission in wireless sensor networks,” IEEE Transactions on Multimedia, vol. 10, no. 6, pp. 1169-1180, Oct. 2008.

[42] T. Ma, M. Hempel, D. Peng, and H. Sharif, "Rate-switching unequal error protection for wireless electrocardiogram (ECG) transmission," Military Communication Conference (MILCOM2010), pp. 1181-1186, 2010.

[43] T. Ma, M. Hempel, D. Peng, and H. Sharif, "A survey of energy-efficient compression and communication techniques for multimedia in resource constrained systems," IEEE Communications Surveys and Tutorials, vol. 15, no. 3, 2013.

[44] G. Balakrishnan, M. Yang, Y. Jiang, and Y. Kim, "Performance analysis of error control codes for wireless sensor networks," IEEE Fourth International Conference on Information Technology, 2007 (ITNG'07), pp. 876-879, Apr. 2007.

[45] F. Yang and S. Wan, "Bitstream-based quality assessment for networked video: A review," IEEE Communications Magazine, vol. 50, no. 11, pp.203-209, Nov. 2012.

[46] T. Lin, S. Kanumuri, Y. Zhi, D. Poole, P. Cosman, and A. Reibman, "A versatile model for packet loss visibility and its application to packet prioritization," IEEE Transactions on Image Processing, vol. 19, no. 3, pp. 722-735, Mar. 2010.

[47] ITU-T, "Recommendation J.144: Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference," 2004.

[48] C. Keimel, M. Klimpke, J. Habigt, and K. Diepold, "No-reference video quality metric for HDTV based on H.264/AVC bitstream features," IEEE International Conference on Image Processing (ICIP), pp. 3325-3328, Sep. 2011.

[49] ITU-T, "Recommendation G.1070: Opinion model for video-telephony applications," Apr. 2007.

[50] A. Rossholm and B. Lovstroem, “A new low complex reference free video quality predictor,” in IEEE 10th Workshop on Multimedia Signal Processing, pp. 765-768, Oct. 2008.

[51 T. Oelbaum, C. Keimel, and K. Diepold, “Rule-based no-reference video quality evaluation using additionally coded videos,” IEEE Journal of Selected Topics in Signal Processing, vol. 3, no. 2, pp. 294-303, Apr.2009.

[52] T. Kawano, K. Yamagishi, K. Watanabe, and J. Okamoto, “No reference video-quality-assessment model for video streaming services,” IEEE 18th International Packet Video Workshop, pp. 158-164, Dec. 2010.

[53] E. Aguiar, A. Riker, M. Mu, and S. Zeadally, “Real-time QoE prediction for multimedia applications in wireless mesh networks,” IEEE Consumer Communications and Networking Conference (CCNC), pp. 592-596, Jan. 2012.

[54] http://media.xiph.org/video/derf/.

[55] ITU-T, “Recommendation P.910: Subjective video quality assessment methods for multimedia applications,” 2008.

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

[57] R. Hamzaoui, V. Stankovic, Z. Xiong, K. Ramchandran, R. Puri, A. Majumdar, and J. Chou. “Channel protection fundamentals,” pp. 187-228, 2007.

[58] A. Katsenou, L. Kondi, and K. Parsopoulos, “Motion–related resource allocation in dynamic wireless visual sensor network environments,” IEEE Transactions on Image Processing, vol. 23, no. 1, pp. 56-68, 2014.

[59] H. Wu, M. Claypool, and R. Kinicki, "Adjusting forward error correction with temporal scaling for TCP-friendly streaming MPEG," ACM Transactions on Multimedia Computing, Communications, and Applications, vol. 1, no. 4, pp. 315-337, Nov. 2005.

[60] D. Turaga, C. Yingwei, and J. Caviedes, "No reference PSNR estimation for compressed pictures," International Conference on Image Processing, vol.3, no.3, pp.61-64, 2002.

[61] M. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," IEEE transactions on broadcasting, vol. 50, no. 3, pp. 312-322, 2004.

[62] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," vol. 13, no. 4, pp. 600-612, 2004.

[63] Y. Ou, Z. Ma, and Y. Wang, "Perceptual quality assessment of video considering both frame rate and quantization artifacts," IEEE Transactions on Circuits and Systems for Video Technology, vol. 21, no. 3, pp. 286-298, Mar. 2011.

[64] R. Hamzaoui, V. Stankovic, Z. Xiong, K. Ramchandran, R. Puri, A. Majumdar, and J. Chou. "Channel protection fundamentals," pp. 187-228, 2007.

[65] K. Kang and H. Shin, "Reduced data rates for energy-efficient Reed–Solomon FEC on fading channels," IEEE Transactions on Vehicular Technology, vol. 58, no. 1, pp. 176-187, Jan. 2009.

[66] H. Radha and D. Loguinov, "Channel modeling and analysis for the Internet," Multimedia over IP and Wireless Networks: Compression, Networking, and Systems, 229 pp., 2011.

[67] A. Hameed, R. Dai, and B. Balas, "Predicting the perceptual quality of networked video through light-weight bitstream analysis," IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom), pp. 48-52, May 2014.

[68] J. Ong, L. Ang, and K. Seng, “Implementation of (15, 9) Reed Solomon minimal instruction set computing on FPGA using Handel-c,” IEEE International Conference on Computer Applications and Industrial Electronics (ICCAIE), pp. 356-361, 2010.

[69] P. Dayal and R. Patial, “Implementation of Reed-Solomon codec for IEEE 802.16 network using VHDL code,” IEEE International Conference in Optimization, Reliability, and Information Technology (ICROIT), pp. 452-455, 2014.

[70] S. Pudlewski, A. Prasanna, and T. Melodia, “Compressed-sensing enabled video streaming for wireless multimedia sensor networks,” IEEE Transactions on Mobile Computing, vol. 11, no. 6, pp. 1060-1072, 2012.

[71] L. Hanzo, T. Liew, B. Yeap, R. Tee, and S. X. Ng, “Turbo coding, turbo equalisation and space-time coding: EXIT-chart-aided near-capacity designs for wireless channels,” John Wiley and Sons, 2011.

[72] A. Katsenou, L. Kondi, and K. Parsopoulos, “Motion–related resource allocation in dynamic wireless visual sensor network environments,” IEEE Transactions on Image Processing, vol. 23, no. 1, pp. 56-68, 2014.

[73] C. Nastasi and A. Cavallaro, “Wise-MNet: an experimental environment for wireless multimedia sensor networks,” Proceedings of Sensor Signal Processing for Defense (SSPD), 2011.

[74] D. Rosário, Z. Zhao, C. Silva, E. Cerqueira, and T. Braun. “An OMNeT++ framework to evaluate video transmission in mobile wireless multimedia sensor networks,” 6th International ICST Conference on Simulation Tools and Techniques (SimuTools’13), pp. 277-284, 2013.

[75] A. Vagra, "The OMNeT++ discrete event simulation system," Proceedings of the European Simulation Multimedia Conference (ESM2011), 2011.

[76] A. Boulis, "Castalia, a simulator for wireless sensor networks and body area networks, version 2.2," August 2009.

[77] N. Ambady, R. Rosenthal, "Half a minute: predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness," Journal of Personality and Social Psychology, vol. 64, pp. 431-441, 1993.

[78] N. Ambady and H. Gray, "On being sad and mistaken: mood effects on the accuracy of thin-slice judgments," Journal of Personality and Social Psychology, vol. 83, pp. 947-961, 2002.

[79] N. Ambady and M. Krabbenhoft, "The 30-sec scale: using thin-slice judgments to evaluate sales effectiveness," Journal of Consumer Psychology, vol. 16, no. 1, pp. 4-13, 2006.

[80] N. Ambady, M. Hallahan, and B. Conner, "Accuracy of judgments of sexual orientation from thin slices of behavior," Journal of Personality and Social Psychology, vol. 77, pp. 538-547.

[81] N. Rule, N. Ambady, "Brief exposures: Male sexual orientation is accurately perceived at 50ms," Journal of Experimental Social Psychology, vol.44, pp.1100-1105, 2008.

[82] B. Balas, P. Sinha, "Learned prediction affects body perception," Visual Cognition, vol. 17, pp. 679-699, 2009.

[83] P. Sinha, B. Balas, Y. Ostrovsky, and R. Russell, "Face recognition by humans: nineteen results all computer vision researchers should know about," Proceedings of IEEE, vol. 94, no. 11, pp. 1948-1962, Nov. 2006.

[84] C. Bidet-Ildei, E. Kitromilides, J. Orliaguet, M. Pavlova, and E. Gentaz, “Preference for point-light human biological motion in newborns: contribution of translational displacement,” Developmental Psychology, vol. 50, no.1, pp. 113-120, Jan. 2014.

APPENDIX A. PUBLICATIONS

Published

A. Hameed, R. Dai, and B. Balas, "Predicting the Perceptual Quality of Networked Video through Light-Weight Bitstream Analysis", *IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom)*, May 2014.

Manuscript in Preparation

A. Hameed, R. Dai, B. Balas, "A Perceptual Video Quality Prediction Model and Its Application in Forward Error Correction for Wireless Multimedia Communications," submitted for journal publication.

A. Hameed, B. Balas, R. Dai, "Thin-slice vision: Another Application of Video Confidence Measure by Perceptual Quality Model," submitted for journal publication.

APPENDIX B. VITA

Abdul Hameed is a graduate student and pursuing his Ph.D. degree in Electrical and Computer Engineering department at North Dakota State University. He received his B.S. in Computer Science from COMSATS Institute of Information Technology in 2006, and M.S. in Computer Software Engineering from National University of Sciences and Technology in 2008 respectively. Before joining NDSU for Ph.D. studies, he was working as Lecturer in COMSATS. His broad research interests include perceptual video coding, image and video quality assessment, multimedia communications, machine learning algorithms, video compression, encoding, decoding, and processing.